UNIVERSITÉ PARIS OUEST

Nanterre La Défense

UNIVERSITÀ CATTOLICA del Sacro Cuore di MILANO

Thèse de doctorat en Sciences du Langage

ED 139 « Connaissance, langage, modélisation », UMR 7114 - MoDyCo

Dottorato di Ricerca in Scienze linguistiche e letterarie - ciclo XXVIII

L'acquisition de la liaison chez des apprenants italophones. Des atouts d'un corpus de natifs pour l'étude de la liaison en français langue étrangère (FLE)

Giulia BARRECA

sous la direction de :

Bernard LAKS

Enrica GALAZZI

Thèse présentée et soutenue publiquement le 7 décembre 2015 devant le jury composé par :
Tesi discussa pubblicamente il 7 dicembre 2015 davanti alla commissione composta da :

Enrica GALAZZI
Professore ordinario, Università Cattolica - Milano, directrice/*direttrice*
Bernard LAKS
Professeur, Université Paris Ouest - Nanterre, directeur/*direttore*
Jean-Pierre CHEVROT
Professeur, Université Stendhal - Grenoble, rapporteur/*relatore*
Ruggero DRUETTA
Professore associato, Università degli studi di Torino, rapporteur/*relatore*

À mes familles

« Oh ! Je voudrais encore vous dire, je vous / aime »

Sim

Sommaire

Remerciements VII
Sigles et abréviations IX
Résumé XI
Abstract XIII
Abstract XV
Introduction 1

Partie 1. La liaison : origines, classifications, traitements et représentations 7

Chapitre 1. La liaison : un cas particulier de *sandhi* externe 9
Chapitre 2. Variables linguistiques d'étude de la liaison 31
Chapitre 3. Normes de prononciation et approches méthodologiques pour l'étude de la liaison 63
Chapitre 4. Typologies et classifications de la liaison 77
Chapitre 5. Traitement et statut de la consonne de liaison : approches linguistiques 101
Chapitre 6. Représentations de la liaison : approches psycholinguistiques 131

Partie 2. Analyse fréquentielle de la liaison dans un corpus de français parlé 153

Chapitre 1. Le corpus « Phonologie du Français Contemporain » (PFC) : un corpus de français parlé codé en liaison et annoté en morphosyntaxe 155
Chapitre 2. Analyse de la distribution de la fréquence de la liaison variable dans la parole semi-spontanée 171

Partie 3. Enseignement et acquisition de la liaison en français langue seconde et étrangère 207

Chapitre 1. Repères théoriques sur l'acquisition d'une langue seconde et étrangère 209
Chapitre 2. L'acquisition de la liaison en français langue seconde et étrangère 239
Chapitre 3. Normes pédagogiques et matériaux didactiques dans l'enseignement de la liaison en français langue étrangère 271

Partie 4. Une étude longitudinale de l'acquisition de la liaison chez des apprenants italophones de français langue étrangère (FLE) : l'enquête IPFC-italien (Rome) 291

Chapitre 1. L'enquête « Interphonologie du Français Contemporain - italien » (IPFC-italien) (Rome) : méthodologies, corpus et traitement des données 293

Chapitre 2. Résultats de l'enquête IPFC-italien (Rome) 331

Chapitre 3. Discussion et aspects d'un itinéraire d'acquisition de la liaison en français langue étrangère 407

Partie 5. Des atouts d'un corpus de natifs dans l'acquisition de la liaison en francais langue etrangère 427

Chapitre 1. Ressources pour l'acquisition de la liaison en français langue étrangère 429

Conclusions générales 451

Références bibliographiques 459

Index des notions 489

Index des noms 491

Table des illustrations 495

Table des exemples 501

Table des figures 503

Table des matières 507

Annexes 527

Annexe 1 529

Annexe 2 535

Annexe 3 539

Annexe 4 547

Remerciements

Mes remerciements vont en premier lieu à mes directeurs, Bernard Laks et Enrica Galazzi pour avoir dirigé cette thèse et pour m'avoir permis de la réaliser dans les meilleures conditions. Je tiens particulièrement à les remercier d'avoir été attentifs et disponibles en dépit de leurs nombreuses charges. Leurs compétences et leur rigueur scientifique m'ont beaucoup appris.

Je voudrais également exprimer ma gratitude aux membres du jury, Jean-Pierre Chevrot et Ruggero Druetta pour l'intérêt qu'ils ont porté à ce travail en s'engageant à être rapporteurs.

Toute ma reconnaissance va également à tous les professeurs et doctorants que j'ai eu l'occasion de rencontrer au sein des projets « Phonologie du Français Contemporain » et « InterPhonologie du Français Contemporain ». Je souhaite notamment remercier Jacques Durand, Chantal Lyche, Marie-Hélène Côté, Sylvain Detey, Isabelle Racine et Elissa Pustka pour m'avoir toujours encouragé dans mon travail. Leurs remarques et suggestions ainsi que leur accueil enthousiaste ont contribué à améliorer la qualité de cette thèse, et je leur en suis très reconnaissante.

Je souhaiterais aussi adresser ma gratitude à Jeff Tennant pour ses patientes relectures et ses conseils. Un remerciement particulier va à mon collègue et ami George Christodoulides (« Christophe » pour les amis). Merci pour ton aide ainsi que pour ta disponibilité et ta patience qui ont marqué toute notre collaboration.

Je voudrais adresser mes remerciements à tous les membres de l'équipe italienne du projet « InterPhonologie du Français Contemporain » ainsi qu'aux informateurs et aux membres de l'Université Sapienza de Rome sans lesquels je n'aurais jamais pu réaliser ce travail. Je tiens en particulier à témoigner toute ma reconnaissance à Oreste Floquet qui a soutenu mes recherches bien avant ma thèse.

J'adresse également mes remerciements à Jean-Luc Minel, Caroline Bogliotti et tous les membres du laboratoire MoDyCo, ainsi que l'École Doctorale 139 de l'Université Paris Ouest pour m'avoir accueilli chaleureusement aux cours des deux dernières années de thèse. Toute ma reconnaissance va également à Christophe Parisse pour m'avoir donnée la possibilité de collaborer au projet Ortolang au début de ma deuxième année de thèse. Je remercie également Myriam Djedi, Alexandre Sourie, Xavier Monnin et Jasmine Debois pour leur professionalisme

ainsi que pour leur gentillesse. Merci bien évidemment à tous les doctorants et doctorantes du laboratoire. Un grand merci va à Julie Beliao, Marine Damiani, Kaja Dolar, Clémentine Gross, Myriam Majdoub, Anaïs Moreno, Laure Martin,Trang Luong, Julie Peuvergne et Ilaine Wang. Les réunions, les diners ainsi que les « trajets » à la bibliothèque et en RER m'ont beaucoup appris.

Mes remerciements vont également à l'école doctorale de l'Università Cattolica de Milan et à Anne Guillermou pour ses précieuses relectures et à Jean-Pierre Chevrot et Mylène Harnois-Delpiano pour la disponibilité et l'attention qu'ils ont su accorder à mon travail pendant ma première année passée à Grenoble.

Je remercie la Région de Île de France, l'Université Franco-Italienne et la Fondazione Primoli qui ont financé cette thèse et sans lesquels elle n'aurait pu être menée à bien.

Finalement, merci à ma famille, ma belle-famille, à mes amis et à mon ancien employeur et ami que j'ai toujours eus à mes côtés malgré les distances plus ou moins importantes !

À Arnaud pour avoir cru en moi davantage que moi. Merci pour ton aide et ton soutien constants qui ont accompagné tout ce parcours dès les tous premiers jours.

Sigles et abréviations

AS	Apprenant ayant réalisé des séjours dans un pays francophone (à savoir « avec séjour »)
CL	Consonne de liaison
Enchaînement	Liaison réalisée enchaînée
IPFC	Projet « Interphonologie du Français Contemporain »
LR	Liaison réalisée
% LR	Taux de réalisation de la liaison
LNR	Liaison non réalisée
M1	Premier mot appartenant à la séquence de liaison (ex. « les » dans la séquence de liaison « les arbres »)
M2	Deuxième mot appartenant à la séquence de liaison (ex. « arbres » dans la séquence de liaison « les arbres »)
Non-enchaînement	Liaison réalisée non-enchaînée
Non-réalisation	Liaison non réalisée
Orthographique	Réalisation orthographique de la consonne de liaison
Pause-hésitation	Consonne de liaison réalisée après une pause temporelle ou intonationnelle
POS	Partie de discours (*Part of speech*)
POS tag	Étiquette morphosyntaxique de l'annotation en partie de discours
Pos:min	Étiquette morphosyntaxique des unités lexicales minimales (*POS tag of minimal tokens*)
PFC	Projet « Phonologie du Français Contemporain »
Substitution	Réalisation d'une consonne autre que la consonne de liaison cible
Tok:min	Token minimal de l'annotation morphosyntaxique (à savoir *Minimal Tokens*) ;
Tok:mwu	Unité poly-lexicale (tok:mwu à savoir *Multi-word units*) ;
Token de liaison	Occurrence de liaison, à savoir la séquence des mots de liaison (ex. « les arbres »)

Type de liaison	Contexte syntaxique de réalisation d'une liaison potentielle, défini par les catégories morphosyntaxiques d'appartenance des mots de liaison (ex. « DET-NOM » dans le cas d'une liaison entre un déterminant « DET » et le nom suivant « NOM »)
SS	Apprenant n'ayant pas réalisé de séjour dans un pays francophone (à savoir « sans séjour »)

Résumé

Dans le cadre du projet « InterPhonologie du Français Contemporain » (IPFC) (Detey et Kawaguchi 2008; Racine, Detey et Kawaguchi 2012), ce travail se propose d'examiner les stratégies d'acquisition de la liaison en L2. Si des modèles développementaux ont été proposés pour l'acquisition de la liaison en L1, aucune hypothèse pour rendre compte de l'apprentissage en L2 n'a reçu à ce jour un appui empirique convaincant (Wauquier 2009). La faible représentativité de sujets ayant des langues sources différentes de l'anglais tout comme les difficultés de comparabilité qui caractérisent les travaux précédents empêchent de formuler des généralisations sur les processus d'acquisition de la liaison en L2.

C'est dans ce contexte que s'inscrit la présente étude longitudinale menée auprès d'étudiants italophones de français langue étrangère (FLE) (niveau A2-B1) dont les résultats ont été enrichis par une comparaison avec les autres travaux multilingues (plusieurs langues sources) issus du projet international IPFC. Cette approche a permis de mettre en évidence la présence de tendances et d'erreurs communes aux différentes populations d'apprenants qui suggèrent que, malgré le recours à des stratégies lexicales (Wauquier 2009), les apprenants sont en mesure de développer des généralisations et des représentations phonologiques de la liaison. Celles-ci sont pourtant susceptibles d'être influencées par les représentations orthographiques (Grainger et Ferrand 1996), comme l'indiquent les réalisations non conformes de la consonne de liaison (ex. « mes amis » réalisé [mesami] au lieu de [mezami]) attestées non seulement dans des tâches de lecture mais également dans celles de parole semi-spontanée et dans un test de jugement. De plus, compte-tenu des difficultés que l'hétérogénéité de la liaison pose dans l'enseignement et l'acquisition en français L2 (Racine et Detey sous presse), les apprenants montrent des faiblesses tant au niveau de la production de la liaison qu'au niveau des connaissances épilinguistiques (Gombert 1990) de la variation de la liaison. Ces résultats nous ont amené à exploiter les données extraites de l'analyse fréquentielle (Barreca et Christodoulides 2015) du corpus des natifs « Phonologie du Français Contemporain » (PFC) (Durand, Laks et Lyche 2009) afin de proposer des ressources pédagogiques dont nous espérons qu'elles seront en mesure, dans le cadre d'une démarche *data-driven learning,* de contribuer au renouvèlement de l′enseignement de la liaison en français langue étrangère (Detey et Racine sous presse).

Mots clés : acquisition de la liaison, français langue étrangère, stratégies d'acquisition, usages des natifs, traitement classificatoire de la liaison, apprentissage sur corpus.

Abstract

Nata nell'ambito del progetto *InterPhonologie du Français Contemporain* (IPFC) (Detey et Kawaguchi 2008; Racine, Detey et Kawaguchi 2012), questa tesi propone un'analisi delle strategie di acquisizione della *liaison* in L2. Sebbene dei modelli di acquisizione della *liaison* in L1 siano già stati proposti, nessuna ipotesi è stata confermata da studi empirici (Wauquier 2009). La scarsa rappresentatività dei locutori di lingue madre diverse dall'inglese, e le difficoltà di comparabilità che caratterizzano gli studi precedenti, rappresentano infatti un ostacolo per la comprensione dei processi che guidano l'acquisizione della *liaison* in L2.

È in questo contesto che si situa il nostro studio longitudinale realizzato su degli studenti universitari italofoni di francese lingua straniera (FLS) (livello A2-B1); un'osservazione i cui risultati sono stati confrontati con altre ricerche multilingui (diverse lingue madre) facenti parte del progetto internazionale IPFC. Questo approccio ha permesso di far emergere la presenza di tendenze e errori comuni alle diverse popolazioni di apprendenti; una similitudine che suggerisce come gli apprendenti, malgrado ricorrano a diverse strategie lessicali (Wauquier 2009), sono in grado di sviluppare delle generalizzazioni e delle rappresentazioni fonologiche della *liaison*. Tuttavia queste ultime possono essere influenzate dalle rappresentazioni grafiche (Grainger et Ferrand 1996), come lo testimoniano le realizzazioni ortografiche delle consonanti di *liaison* (ex. *mes amis* [mesami] invece di [mezami]) registrate non solo nelle attività di lettura ma anche nell'orale semi-spontaneo e in un test di accettabilità. Inoltre, date le difficoltà che la variabilità della *liaison* pone tanto nella didattica quanto nell'acquisizione (Racine et Detey in stampa), gli apprendenti sembrano possedere una scarsa competenza della variazione della *liaison* non solo sul piano della produzione ma anche su quello delle conoscenze epilinguistiche (Gombert 1990).

Questi risultati ci hanno spinto quindi ad utilizzare i dati dell'analisi della frequenza della *liaison* (Barreca e Cristodoulides 2015) condotta sul corpus di nativi *Phonologie du Français Contemporain* (PFC) (Durand, Laks et Lyche 2009) al fine di proporre delle risorse pedagogiche per l'acquisizione della *liaison*. Nell'ambito di un approccio *data-driven learning*, questi strumenti potrebbero contribuire al rinnovamento dell'insegnamento della *liaison* nella didattica del francese lingua straniera (Detey et Racine in stampa).

Parole chiave : acquisizione della *liaison*, francese lingua straniera, strategie di acquisizione, usi dei nativi, classificazioni della *liaison*, *data-driven learning*.

Abstract

As part of the international project *InterPhonologie du Français Contemporain* (IPFC) (Detey Kawaguchi and 2008; Racine, Detey and Kawaguchi 2012), this study aims to examine acquisition strategies of *liaison* in French as a foreign language. While developmental models have been proposed for the acquisition of *liaison* in L1, to date, no hypothesis accounting for L2 *liaison* learning has received convincing empirical support (Wauquier 2009). The low representation of subjects with source languages other than English and the problems of comparability between previous studies preclude generalisations about the processes guiding the acquisition of *liaison* in L2 French.

It is in this context that the present longitudinal study of Italian-speaking students of French as a foreign language (FFL) (level A2-B1) can be situated. The results have been enriched by a comparison with other multilingual studies (several source languages) from the international project IPFC.

This approach shows the presence of common trends and errors in different populations of learners that suggest that, despite the use of lexical strategies (Wauquier 2009), learners are able to develop generalisations and phonological representations of *liaison*. These are however influenced by orthographic representations (Grainger and Ferrand 1996) as indicated by erroneous productions of the *liaison* consonant (eg. *mes amis* realised [mesami] instead of [mezami]) attested not only in reading tasks but also in semi-spontaneous speech and in a judgment test. Taking into account the difficulties that the heterogeneity of *liaison* presents for second language teaching and acquisition (Racine and Detey in press), learners show weaknesses in both production of *liaison* and in epilinguistic knowledge (Gombert 1990) of the variation of *liaison*.

These results led us to use data from a frequency analysis (Barreca and Christodoulides 2015) of a corpus of native speakers *Phonologie du Français Contemporain* (PFC) (Durand, Laks and Lyche 2009) to provide learning resources, based on a data-driven learning approach, to contribute to the renewal of the teaching of *liaison* in French as a foreign language (Detey and Racine in press).

Keywords: acquisition of *liaison*, French as a foreign language, acquisition strategies, native speaker usage, classificatory treatment of *liaison*, data-driven learning.

Introduction

La syllabation ouverte qui caractérise tout le groupe des langues romanes (Martinet 1964 ; Delattre 1965 ; Laks 2005b) est à l'origine d'une série de changements qui affecteront le français du XVIème siècle (Martinet 1964), parmi lesquels la simplification des géminées, l'affaiblissement suivi de la chute du [s] implosif ainsi que la disparition des obstruantes finales et des nasales finales qui affectera les voyelles précédentes (Martinet 1964). Il s'agit en effet d'une tendance très ancienne qui relève d'une dynamique qui favorise la création d'une suite consonne-voyelle (Martinet 1964) et qui se réalise par la chute complète des consonnes finales ou par leur maintien à l'initiale du mot suivant. Ces changements se traduisent par un enchaînement généralisé qui joue un rôle crucial dans l'organisation prosodique de l'énoncé français. En tant que phénomène phonologique se produisant à la frontière entre deux mots, au contact entre la phonologie et d'autres composants de la grammaire, l'enchaînement est décrit comme une forme de *sandhi*[1] externe (Burov 2012). Comme le souligne Laks (2005a), si l'enchaînement est répandu dans plusieurs langues, le français connaît en revanche un autre type de *sandhi* avec une nature tout à fait spéciale, à savoir la liaison. Celle-ci correspond à la prononciation d'une consonne à la frontière de deux mots dont la réalisation est soumise à un conditionnement syntaxique (fort degré de cohésion syntaxique entre les deux mots) et phonologique (le deuxième mot commençant par une voyelle). En dehors de ces conditions la liaison n'a généralement pas lieu. Bien que les deux phénomènes de liaison et enchaînement soient étroitement liés, leur rapport ne saurait être défini en termes d'implication (Burov 2012). Si l'enchaînement consonantique produit une resyllabation de tout type de consonne finale prononcée, le processus de *sandhi* de liaison ne concerne que des consonnes qui, en vertu de leur statut particulier (consonnes de liaison), ne sont pas susceptibles d'être prononcées en dehors du contexte syntaxique et phonologique décrit ci-dessus. La spécificité[2] ainsi que la variabilité de ce phénomène ont fait de la liaison un sujet de recherche qui continue encore aujourd'hui à faire couler beaucoup d'encre.

De nombreuses recherches ont mis en évidence l'influence de divers paramètres linguistiques sur la réalisation des liaisons. On sait que la réalisation de la liaison est extrêmement sensible à la cohésion syntaxique (Damourette et Pichon 1911-1927 ; Delattre 1966), à la cohésion sémantique et rythmique (Grammont 1914), aux facteurs lexicaux

[1] Terme emprunté au sanscrit pour décrire en linguistique moderne plusieurs phénomènes relevant de la phonétique combinatoire.

[2] Non seulement syntaxique et phonologique, mais également phonétique (cf. Partie 1.Chapitre 1.4).

(longueur des mots; fréquence lexicale et fréquence de cooccurrence des mots) (Ågren 1973 ; De Jong 1991 ; Bybee 2005) et sociolinguistiques (Delattre 1966 ; Ågren 1973 ; Anderson 1975) (niveau stylistique du discours, situation de communication, niveau socio-culturel du locuteur et débit de parole).

De plus, on peut dire non seulement que la liaison relève de plusieurs niveaux d'analyse mais aussi qu'elle ne constitue pas un phénomène unique et homogène (Laks 2005a). Elle est au contraire profondément hétérogène et variable. Plusieurs études de corpus (parmi d'autres Mallet 2008 ; Durand et al. 2011 ; Côté 2013) soulignent la forte diversité de comportements qui caractérisent les réalisations à l'intérieur même d'une seule catégorie morphosyntaxique (ex. adverbes monosyllabiques) et du même contexte segmental[3].

Multi-dimensionnalité et hétérogénéité rendent donc difficile un traitement classificatoire de la liaison sous forme de règles formelles ainsi que la prédiction de la réalisation de la liaison (Boula de Mareüil et al. 2003 ; Fougeron et al. 2001).

Il est donc évident que la variabilité qui caractérise ce phénomène de *sandhi* externe est une source de complexité pour l'enseignement et l'apprentissage de la liaison en français. Racine et Detey (sous presse) ont souligné que les difficultés liées à l'acquisition de la liaison touchent d'une part le « niveau de la macro-planification » (Racine 2014) impliquant la maîtrise de la variation sociolinguistique de la liaison et d'autre part, un « niveau de micro-planification » (Racine 2014) qui correspond aux aspects phonético-phonologiques et prosodiques de réalisation de la liaison.

De nombreux chercheurs se sont penchés sur l'acquisition de la liaison en français langue seconde mais la plupart de leurs études ont été menées auprès d'apprenants dont la langue source appartenait au groupe des langues germaniques (anglais et suédois) (Dejean de la Bâtie 1993 ; Mastromonaco 1999 ; Howard 2005 ; Thomas 1998 ; Thomas 2002 ; Thomas 2004; Stridfeldt 2005 ; De Moras 2011)[4]. De plus, les résultats de ces travaux s'avèrent dans l'ensemble difficilement comparables entre eux à cause de la diversité des niveaux des apprenants, des méthodologies utilisées et des différentes classifications de la liaison auxquelles ils font référence. En outre, la plupart des recherches précédentes ont évalué les productions des apprenants en faisant appel à des descriptions prescriptives de la liaison parfois éloignées des usages réels des locuteurs natifs (Racine et Detey sous presse). Une telle approche d'analyse

[3] Par exemple le mot « temps » lie de façon catégorique dans les constructions figées (ex. « de temps en temps »). La liaison est en revanche variable lorsqu'elle marque le nombre (ex. « les temps obscurs ») et ne lie pas lorsqu'elle est finale de morphème au singulier (ex. « le temps obscur ») (Durand et al. 2011).

[4] Sauf l'étude menée auprès d'apprenants coréens (Harnois-Delpiano, Cavalla et Chevrot 2012).

des données risque d'accentuer la distance qui sépare la maîtrise de la variation sociolinguistique chez l'apprenant des usages réels en vigueur chez les natifs.

Quant aux modèles développementaux proposés pour l'acquisition de la liaison en français langue seconde, si la plupart des auteurs s'accordent pour formuler l'hypothèse selon laquelle les apprenants utiliseraient des stratégies d'acquisition de type lexical, les positions divèrgent quant à la capacité chez les apprenants de français langue seconde à développer des généralisations.

D'après Wauquier (2009), les apprenants traiteraient la liaison « au cas par cas » en s'appuyant sur une représentation orthographique de surface, et ils ne seraient pas en mesure, contrairement aux enfants francophones, d'élaborer des généralisations morpho-phonologiques.

En revanche Harnois-Delpiano et al. (2012) n'excluent pas que les apprenants puissent opérer des généralisations indépendantes du mot en s'appuyant sur le système général des correspondances phonographiques. Compte tenu de ces considérations, il est évident que l'étude de l'acquisition de la liaison en langue seconde appelle à davantage de recherches.

C'est dans ce cadre que se situent les travaux issus du projet « InterPhonologie du Français Contemporain » (IPFC) (Detey et Kawaguchi 2008; Racine et al. 2010; Racine, Detey et Kawaguchi 2012), qui a donné lieu à la première ressource de français L2 à visée phonologique et qui compte à ce jour quinze L1 différentes. La représentativité linguistique ainsi que les possibilités de comparabilité entre les différentes études au sein du projet (assurées par le recours à un protocole commun) semblent pouvoir offrir des réponses adéquates aux inconvénients posés par les études précédentes. De plus la proximité avec le protocole d'étude adopté par le projet « Phonologie du Français Contemporain » (PFC) (Durand, Laks et Lyche 2009) (dont le projet IPFC représente le volet non-natif) offre un modèle de référence pour l'étude des productions des apprenants de français langue étrangère qui émerge d'une vaste étude de corpus.

C'est au sein de ce projet international que s'inscrit la présente recherche dont l'objectif principal est de suivre, par le biais d'une étude longitudinale sur une année, l'acquisition de la liaison chez 17 italophones de niveau intermédiaire, tous étudiants inscrits en Licence de Langues et Littératures Étrangères à l'Université Sapienza de Rome.

Dans ce travail, nous proposons d'approfondir la question de l'acquisition de la liaison en langue étrangère sous plusieurs angles, celui de la production (lecture et parole semi-spontanée) et celui des compétences épilinguistiques. L'approche longitudinale a été retenue afin d'essayer

d'esquisser un itinéraire acquisitionnel (Bartning et Schlyter 2004) qui à ce jour n'a pas encore été proposé pour décrire l'apprentissage de la liaison en langue seconde.

Dans une première partie, le premier chapitre (cf. Partie 1.Chapitre 1) entend offrir un cadre théorique sur les phénomènes de *sandhi* (Burov 2012) ainsi qu'un bref parcours historique de l'évolution des consonnes finales en français afin de mieux saisir les particularités de la liaison.

Le deuxième chapitre (cf. Partie 1.Chapitre 2) s'attache à donner un aperçu de l'ensemble des facteurs linguistiques qui font de la liaison un phénomène multidimensionnel, extrêmement hétérogène et variable.

Cela nous amènera dans le chapitre suivant (cf. Partie 1.Chapitre 3) à examiner la question des modèles de prononciation de référence et des approches méthodologiques à adopter dans l'étude de la liaison. Un bref parcours des modèles de prononciation du français contemporain proposés au cours du XXème siècle sera ici proposé afin de tenter de décrire les caractéristiques phonético-phonologiques (y compris la liaison) du français de référence.

La même démarche sera retenue dans le quatrième chapitre (cf. Partie 1.Chapitre 4), entièrement consacré à l'évolution des traitements classificatoires de la liaison et des révisions auxquelles ceux-ci ont été soumis en fonction des sources et des approches méthodologiques. Pour conclure cette première partie, nous présenterons dans les cinquième et sixième chapitres (cf. Partie 1.Chapitre 5 ; Partie 1.Chapitre 6) une vue d'ensemble des modèles phonologiques et psycholinguistiques élaborés afin de rendre compte du statut particulier ainsi que des représentations cognitives des consonnes de liaison.

La deuxième section de ce travail correspond à une première analyse expérimentale de cette recherche. Par le biais d'une étude fréquentielle de la liaison variable menée dans le corpus PFC (Durand, Laks et Lyche 2002), nous nous attacherons dans cette partie à parvenir à une description plus fine de la variation de la liaison pouvant servir de modèle de référence pour l'étude des productions des apprenants de français langue étrangère.

Le premier chapitre (cf. Partie 2.Chapitre 1) sera ainsi l'occasion de présenter les objectifs, les caractéristiques principales du corpus PFC (Durand, Laks et Lyche 2002) ainsi que l'important travail de réannotation morphosyntaxique (Barreca et Christodoulides 2015) que nous avons conduit sur ce corpus.

Dans le chapitre suivant (cf. Partie 2.Chapitre 2), en ligne avec les travaux précédents menés dans le cadre du corpus PFC (Durand et al. 2011; Laks, Calderone et Celata 2014), nous présenterons les résultats de l'analyse fréquentielle de la liaison variable.

Au terme de cette analyse des usages des natifs qui nous serviront de modèles de référence pour l'étude des productions des apprenants, nous nous pencherons sur les questions liées à proprement parler à l'acquisition de la liaison en langue étrangère.

Dans la troisième partie du travail, après avoir proposé un bref aperçu des repères théoriques dans l'acquisition d'une langue étrangère (cf. Partie 3), nous passerons en revue dans le deuxième chapitre (cf. Partie 3.Chapitre 2) les études précédentes menées sur la liaison (y compris les travaux issus du projet IPFC).

Le dernier chapitre de cette section (cf. Partie 3.Chapitre 3) nous permettra d'aborder à nouveau la question du modèle de prononciation de référence, mais cette fois d'une perspective pédagogique en vue d'une sensibilisation des apprenants à la variation sociolinguistique. La deuxième partie de ce chapitre sera en effet l'occasion de présenter un bref survol des matériaux didactiques proposés pour l'enseignement de la liaison.

Dans la quatrième partie de cette étude, le premier chapitre (cf. Partie 4.Chapitre 1) s'attachera à présenter les objectifs et les méthodologies du projet IPFC. Au sein de ce projet international, nous nous attarderons sur les travaux de l'équipe IPFC-italien dirigée par Enrica Galazzi avant de détailler le protocole, les méthodologies, les objectifs ainsi que les hypothèses de la présente étude qui a abouti à la constitution du corpus IPFC-italien (Rome) (environ 12.000 sites de liaison codés) (Barreca et Floquet 2015). Dans cette recherche, nous nous proposons de suivre l'évolution de l'acquisition de la liaison en langue étrangère et de vérifier les capacités des apprenants à établir des généralisations.

Nous nous attacherons dans un deuxième temps à analyser l'influence exercée par la L1 sur la capacité de resyllabation de la CL ainsi que celle de l'input écrit sur la production de la liaison (notamment sur les réalisations des formes orthographiques de la CL dans les différentes tâches prévues par le protocole). Il s'agira par la suite d'étudier dans quelle mesure l'apprenant de français langue étrangère possède des compétences en production et des connaissances épilinguistiques de la variation de la liaison.

Le deuxième et le troisième chapitre (cf. Partie 4.Chapitre 1.42 ; Partie 4.Chapitre 3) seront respectivement consacrés à la présentation des résultats et des discussions de cette étude tout en proposant des comparaisons avec les données issues des autres travaux menés auprès de différentes populations d'apprenants dans le cadre du projet IPFC.

Pour conclure, dans la dernière partie de ce travail (cf. Partie 5), nous nous interrogerons sur le traitement des CL et de la variabilité de liaison proposé par les matériaux didactiques dans l'enseignement de la liaison en FLE.

Nous nous appuierons ensuite sur l'analyse fréquentielle de la liaison menée à partir des données des natifs (corpus PFC) pour développer des ressources pédagogiques (notamment une classification et un concordancier multi-niveaux et multimédia) pouvant servir de supports complémentaires pour l'acquisition de la liaison, dans une démarche d'apprentissage sur corpus.

Partie 1. LA LIAISON : ORIGINES, CLASSIFICATIONS, TRAITEMENTS ET REPRESENTATIONS

Chapitre 1. La liaison : un cas particulier de *sandhi* externe

Le terme *sandhi*, a été emprunté au sanskrit et signifie « mise ensemble », « jonction ». Renou (1975) le définit notamment comme une « jonction » dans laquelle « la phrase est sentie comme une unité phonique et les finales de mot sont mises en harmonie avec les initiales selon des procédés minutieux que la graphie reproduit » (Renou 1975, p. 31). Dans la grammaire sanscrite, ce terme désignait l'ensemble des modifications que subissent les fins et les débuts de mot à l'intérieur de la phrase. Aujourd'hui, en linguistique moderne le mot *sandhi* regroupe, selon les auteurs, plusieurs phénomènes relevant de la phonétique combinatoire qui présentent une nature différente les uns des autres (notamment des phénomènes morphophonologiques, des phénomènes de coarticulation etc.). Parmi ceux-ci, des exemples de *sandhi* sont attestés dans plusieurs langues romanes : le *raddoppiamento fonosintattico* en italien, les mutations consonantiques en occitan et gallois, l'enchaînement et la liaison en français qui sera l'objet de ce travail. Dans ce premier chapitre nous ferons référence au travail de Burov (2012) afin de dégager les caractéristiques principales du *sandhi* qui nous permettront de situer le phénomène de la liaison. Cette forme particulière de *sandhi* externe, spécifique au français, consiste à prononcer une consonne à la frontière de deux mots dont la réalisation est soumise à un conditionnement syntaxique et phonologique (cf. Paragraphe 2.3). Nous nous attacherons ensuite à tracer un bref parcours diachronique de la liaison, qui nous amènera au moins partiellement à définir les origines des typologies de ce phénomène (cf. Paragraphe 5).

1. Le phénomène de sandhi

Comme nous l'avons souligné dans le paragraphe précédent, le mot *sandhi* a été adopté par la linguistique moderne afin de rendre compte d'un ensemble de phénomènes phonotactiques différents qui se produisent généralement au contact de la phonologie avec les autres composants de la grammaire (morphologie et syntaxe) (Burov 2012). Les unités linguistiques, telles que les mots et les morphèmes, subissent des modifications phonétiques selon les contextes qu'ils occupent. En sanscrit, les linguistes historiques, notamment Whitney (1889), avaient déjà opéré une différence entre deux typologies de phénomènes de *sandhi*. D'une part, ceux qui se produisent à la frontière morphémique résultant de la juxtaposition des morphèmes à l'intérieur du mot (*sandhi* interne) et donc régis par des règles de combinaison interne. D'autre part, ceux qui intéressent le début (*sandhi* initial) ou la fin

(*sandhi* externe) des mots, soumis à des règles de combinaison externe et qui donnent lieu à l'enchaînement des unités dans la phrase. Ces observations seront ensuite reprises et réinterprétées par des recherches suivantes sur le sanscrit comme celle de Allen (1962 cité par Burov 2012) qui distingue les phénomènes morphonologiques proprement dits de *sandhi* (conditionnés par le contexte phonétique) de ceux qui ont une nature morpholexicale et qui relèvent d'un processus de supplétion. Dans ce dernier cas, selon l'auteur, les alternances représentent des variantes suppletives d'un seul morphème. Il en résulte que seules les premières constituent des formes de *sandhi* et se différencient entre elles par des degrés différents de productivité. En effet, les modifications morphophonologiques peuvent être conditionnées par le seul contexte phonétique (dans ce cas elles sont dites « automatiques ») ou être déclenchées par des facteurs supplémentaires d'ordre lexical et/ou morphologique.

Comme nous l'avons déjà souligné, dans le cadre de la linguistique contemporaine, les définitions et les domaines d'application de la notion de *sandhi* varient souvent d'une étude à l'autre. En effet, ce terme regroupe souvent des phénomènes qui se situent à l'interface entre la phonologie, la morphologie et la syntaxe dont les conditionnements sont parfois difficiles à cerner. Certains auteurs issus de la phonologie générative naturelle (Hooper 1976 cité par Burov 2012), préfèrent assigner aux phénomènes de *sandhi* un statut de « classe intermédiaire » entre les règles phonétiques et les règles morphophonémiques en excluant les alternances internes au mot de la définition de *sandhi*. Cette définition du *sandhi* comme catégorie intermédiaire propre à la phonologie générative, sera critiquée par d'autres auteurs, parmi lesquels Clayton (1981 cité par Burov 2012) qui considère les phénomènes de *sandhi* comme soumis à des conditionnements phonétiques qui ne demandent pas d'information morphosyntaxique supplémentaire. Selon Clayton (1981), les règles de *sandhi* peuvent donc être de nature phonétique (automatiques et conditionnées seulement par le contexte phonétique) ou morphophonémique (non automatiques) selon qu'elles requièrent ou non une information non phonologique. C'est le cas, par exemple du phénomène de *raddoppiamento fonosintattico* irrégulier (Loporcaro 1997), phénomène de *sandhi* externe présent dans certaines variétés d'italien qui consiste en un redoublement de la consonne initiale après un nombre restreint de mots-outils capables de déclencher le processus (ex. « tra poco » [trappɔko]). Ce type de *sandhi* présuppose un conditionnement lexical qui, d'après l'auteur, attribue à la règle de *sandhi* un caractère morphophonémique. Crystal (1980 cité par Burov 2012) semble également assigner au terme de *sandhi* un sens large qui désigne de façon générale les changements phonologiques auxquels sont soumises les formes grammaticales juxtaposées. Cette définition intègre donc également des modifications phonétiques telles que les assimilations.

D'autres modèles (Ternes 1986 ; Selkirk 1978 ; Nespor et Vogel 1986 cités par Burov 2012) ont en revanche limité l'extension du domaine d'application du phénomène de *sandhi*. Selon Ternes (1986) les phénomènes de jonction ne constituent pas tous des phénomènes de *sandhi*. L'auteur établit une différenciation à partir de la nature et de la place hiérarchique des éléments (phonèmes, syllabes, morphèmes et mots) comme étant l'objet de la jonction. Les phénomènes de coarticulation, relevant de la superposition des gestes articulatoires au moment de la production de la parole, représentent des variantes allophoniques d'un seul phonème (Troubetzkoy 1949) qui ne peuvent pas être traitées en tant que phénomènes de *sandhi.* Il en va de même pour les contraintes phonotactiques qui gouvernent la distribution des phonèmes à l'intérieur des morphèmes qu'il considère comme des phénomènes d'allomorphie. Il ne réserve le terme de *sandhi* qu'à des changements d'ordre purement phonétiques qui se produisent de manière automatique à la frontière de mot. Celle-ci, selon l'auteur, constitue la seule information non phonétique capable de déclencher le phénomène de *sandhi* ce qui revient à nier (comme d'autres l'avaient déjà fait) la notion de *sandhi* interne. Tout changement morphophonémique ou morpholexical tel que le *raddoppiamento* de l'italien, la liaison du français et les mutations consonantiques des langues celtiques relevant partiellement d'un conditionnement lexical ou impliquant une information morphosyntaxique se trouvent exclus de la définition de *sandhi.*

D'autres auteurs, parmi lesquels Selkirk (1978 cité par Burov 2012), et Nespor et Vogel (1986 cité par Burov 2012), appartenant au courant de la phonologie prosodique, ont en revanche essayé de définir les limites d'application du *sandhi* dans plusieurs langues (*raddoppiamento fonosintattico* en italien, liaison et enchaînement en français et mutations consonantiques en gallois) en intégrant le niveau d'analyse prosodique à l'étude de ces phénomènes. En particulier, ils établissent une différence entre les altérations qui s'appliquent à la jonction entre deux unités prosodiques de même type (*domain-juncture rules*) (et qui correspondent à ce que les auteurs identifient comme des *sandhi*) et les autres changements qui se produisent à l'extrémité (gauche ou droite) d'un constituant prosodique (*domain-limit rules*) qui relèvent d'un processus différent (ex. renforcement du /r/ initial en espagnol). Andersen (1986 cité par Burov 2012) se situant dans le même cadre théorique considère que le *sandhi* constitue la marque phonétique de la cohésion sémantique à l'intérieur de l'énoncé. En effet, contrairement aux règles du type *domain-limit rules* (ayant une simple « fonction démarcative » des unités prosodiques), les règles *domain-juncture rules* (qui gouvernent les phénomènes de

sandhi) marquent l'enchaînement entre deux constituants prosodiques du même type à l'intérieur d'une unité supérieure (« fonction concaténative »)[5].

Malgré la pluralité des définitions et des champs d'analyse du *sandhi* que nous avons brièvement illustrée, l'idée de cohésion semble être évoquée de façon plus au moins explicite dans la plupart des approches. L'altération phonétique réalisée dans le *sandhi* est en effet décrite comme le produit des facteurs (lexicaux, morphologiques, syntaxiques et prosodiques) qui présupposent une notion de cohésion syntagmatique.

2. Les phénomènes de sandhi externe en français : l'enchaînement et la liaison

Le français connaît deux typologies de *sandhi* externe : l'enchaînement (vocalique et consonantique) (cf. Paragraphe 2.1) et la liaison qui représente un cas particulier d'enchaînement consonantique (cf. Paragraphe 2.3). Dans ce bref paragraphe, nous allons essayer de définir le rapport existant entre ces deux phénomènes qui caractérisent la langue française ainsi que son organisation prosodique (cf. Paragraphe 2.2).

2.1. L'enchaînement

En ce qui concerne l'enchaînement, il s'agit d'un phénomène qui consiste à rattacher tout segment vocalique et consonantique prononcé en finale de mot à la syllabe initiale du mot qui suit. Ce phénomène est le résultat en français d'une dynamique de syllabation ouverte très répandue dans tout le groupe des langues romanes (Martinet 1964 ; Delattre 1965 ; Laks 2005b). Delattre (1965) a montré à travers une étude statistique menée sur 8000 syllabes appartenant aux langues germaniques (anglais et allemand) et romanes (espagnol et français) que ces dernières présentent majoritairement une structure syllabique ouverte contrairement à l'anglais et à l'allemand.

> « German strongly favors the CVC type; English is more equally divided between CVC and CV, nevertheless favoring a consonant rather than a vowel terminal; Spanish and French clearly prefer the CV type. The predominance of closed syllables in the two Germanic languages is almost as marked as the predominance of the open syllables in the two Latin languages ». (Delattre 1965, p. 41-42)

[5] Selkirk (1978 cité par Burov 2012) et Nespor et Vogel (1986 cité par Burov 2012) considèrent que les phénomènes de *sandhi* tels que le *raddoppiamento fonosintattico* et la liaison se réalisent à la jonction de deux mots phonologiques qui appartiennent au même syntagme phonologique.

Il s'agit d'une dynamique très ancienne qui remonte au latin classique (Niedermann 1953) perdurant au cours des siècles, du latin vulgaire (Vâânânen 1962) à l'ancien français (Chaurand 1998) jusqu'au français moderne (Martinet 1964 ; Pulgram 1970). Comme nous le verrons dans les pages qui suivent, cette tendance pourrait être à l'origine d'une série de changements qui affecteront le français du XVIème siècle (Martinet 1964), parmi lesquels la simplification des géminées, l'affaiblissement puis la chute du « s » implosif ainsi que la disparition des consonnes finales et des nasales finales qui affectera les voyelles précédentes (Martinet 1964). Il s'agit en effet d'une tendance très ancienne qui relève d'une dynamique universelle qui favorise la stabilisation de la syllabe et donc la création d'une suite CV (Martinet 1964, p. 328) et qui se réalise soit par la chute complète des consonnes finales soit par leur resyllabation à l'initiale du mot suivant. Ce phénomène pourrait être à l'origine de l'enchaînement généralisé aussi bien consonantique[6] (qui déclenche des resyllabations donnant lieu à des structures syllabiques du type CC ou CV), que vocalique (qui produit des suites syllabiques VV ou VC). Fouché (1959) distingue quatre contextes possibles d'enchaînement qui produisent des typologies différentes d'enchaînement vocalique et consonantique : (a) voyelle finale de mot + voyelle initiale du mot suivant ; (b) voyelle finale de mot + groupe consonantique initial du mot suivant ; (c) consonne finale de mot + voyelle initiale du mot suivant ; (d) consonne finale de mot + consonne initiale du mot suivant.

Le premier cas (a) détermine le contact entre deux voyelles sans que la voix s'interrompe entre elles. En effet, les resyllabations résultant des enchaînements vocaliques ne déterminent pas la création de *hiatus* (Grammont 1933, p. 416; Fouché 1959, p. XLIII) et donc d'occlusion glottale transitoire entre les voyelles.

> « Il n'y a pas de *cassure* entre les deux voyelles. Elles sont au contraire liées. Ce qui fait qu'en plus de la liaison *consonantique*, qui est celle qu'on entend lorsqu'on emploie le terme de *liaison,* le français en connaît une autre, qu'on pourrait appeler *vocalique* ». (Fouché 1959, p. XLIII)

Grammont (1933) explique que originairement le mot *hiatus* signifie que la bouche reste ouverte d'une voyelle à l'autre, parce qu'il n'y a pas de consonnes entre les deux qui l'obligent à la fermer plus au moins complètement. Cependant, Grammont (1933, p. 416) précise que ce terme s'applique selon les langues à des phénomènes différents. Contrairement à d'autres langues où les deux voyelles en contact sont séparées par un coup de glotte, en français la glotte ne cesse pas de vibrer et une « liaison vocalique » se produit entre les deux voyelles. Si les deux voyelles ont un timbre différent, une sorte de modulation se produit alors que dans le cas

[6] Dont la liaison est un cas particulier.

contraire on enregistre un léger fléchissement d'intensité vibratoire qui correspond au cas où la bouche reste sensiblement ouverte et donc à la définition même de *hiatus* qui se traduit en français par un enchaînement vocalique.

Dans le deuxième contexte d'enchaînement (b), une voyelle finale de mot rencontre avec un groupe consonantique initial du mot suivant. Si ce dernier contient comme deuxième élément une consonne autre qu'une liquide (« r », « l »)[7], Fouché (1959) souligne qu'il faut éviter de le dissocier et de faire de la première consonne l'élément final de la syllabe précédente (ex. « il est svelte » réalisé [ilɛ.svɛlt] et non [ilɛs.vɛlt]).

Dans le contexte d'enchaînement (c), Fouché (1959) opère une différence entre deux cas, selon que le premier mot se termine par une seule consonne prononcée ou par un groupe consonantique prononcé. Dans le premier cas, la consonne finale du premier mot est resyllabée avec la voyelle initiale du mot suivant (ex. « il entend » [i.lɑ̃.tɑ̃] et non [il.ɑ̃tɑ̃]). Dans le second cas, le groupe consonantique est séparé et la première consonne appartient à la syllabe finale du premier mot alors que la deuxième est rattachée au deuxième mot (ex. « une fresque admirable » [yn.fʀɛs.kad.mi.ʀa.bl][8] ; « ils oublient » [il.zu.bli]). Ce dernier exemple correspond à la liaison qui, comme nous le verrons dans les paragraphes suivants, constitue en cas particulier d'enchaînement consonantique. Le dernier contexte (d) est représenté par le lien entre une consonne finale et la consonne initiale du mot suivant. Plusieurs combinaisons sont possibles dans ce cas selon qu'il s'agit de consonnes simples ou de groupes consonantiques. Le contact entre la consonne finale du premier mot et la consonne initiale du deuxième mot détermine des changements de nature différente suivant les caractéristiques phonétiques articulatoires des deux consonnes. Si les deux consonnes sont identiques elles donnent lieu à une gémination (ex. « chaque quartier »). En revanche si elles ont le même mode et lieu d'articulation, mais une sonorité différente, la première consonne s'assimile à la seconde du point de vue de la sonorité (ex. « une nappe blanche »). Dans le cas contraire (même sonorité et mode et lieu d'articulation différents), la consonne finale ne subit aucune modification (ex. « une étape très longue »). Enfin, si aussi bien la sonorité, que le mode et le lieu d'articulation sont différents, une assimilation de sonorité du même type que le précédent se produit entre les consonnes.

[7] En revanche, Passy (1890) considère que « souvent même deux consones sont réunies à la voyelle qui suit : *tableau* (ta-blo) » (Passy 1980, 37).

[8] En revanche, d'après Delattre (1966) les deux consonnes appartenant au groupe consonantique sont rattachées à la syllabe qui suit (ex. [yn.fʀɛ.skad.mi.ʀa.bl]).

2.2. *Le français langue cursus*

L'enchaînement généralisé joue un rôle crucial dans l'organisation de l'énoncé français dans des unités prosodiques minimales portant un seul accent final ou primaire sur la dernière syllabe. Cette unité bornée à sa droite par l'accent a reçu de nombreuses autres appellations dans la littérature telles que « groupe de souffle » (Passy 1890)[9], « élément rythmique », « groupe d'idée » (Grammont 1914), « mot phonologique » (Selkirk 1972), « groupe rythmique » (Delais-Roussarie 1995), « unité rythmique » (Di Cristo et Hirst 1993), « groupe accentuel » (Lacheret-Dujour 1999). Les critères évoqués pour définir les frontières de ces groupes sont d'ordre prosodique, rythmique, syntaxique[10] et sémantique[11].

Au sein de ces regroupements prosodiques, les mots (sujets à désaccentuation)[12] perdent leur identité phonétique, formant une chaîne de morphèmes (ou mots phonologiques) unique. Selon certains auteurs (Laks 2005b)[13], cela revient à nier la notion même de mot au niveau phonétique. Le débat autour de la démarcation lexicale en français remonte aux observations formulées par Palsgrave (1852 cité par Laks 2005b) qui soulignait déjà la tendance du français oral à former des groupes rythmiques de cinq ou six mots traités comme une seule unité. Sur la base des observations, Pulgram (1970) établit une distinction entre deux typologies de langue. D'une part les langues *nexus* (ex. l'anglais et en général les langues germaniques) où le mot à l'intérieur de l'énoncé conserve son autonomie phonétique. D'autre part les langues *cursus* (« groupe de souffle »), comme le français, où les éléments regroupés dans une seule unité principale du discours perdent leur identité phonétique et suprasegmentale constituant des groupes solidaires pour le sens, pour la syntaxe et pour l'accent.

> « An analysis of French discloses that each pause group is in fact a 'cursus' (called breath group in most French grammars), in which every component word loses both its boundaries

[9] « Il ne faudrait pas croire que les groupes d'acsantuation soient séparés les uns des autres par un aret de la parole ; il y a diminuition de la forse d'expiration rien de plus. Des arets se produisent cependant, de tems autres, lorsque l'on parle ; ils partajent la parole en *groupes de soufle,* seule division toute à fait inataquable du langaje ». (Passy 1890, p. 41)

[10] Ces modèles s'appuient principalement sur des contraintes d'alignement rythmico-syntaxique (qui conduisent à faire coïncider la frontière droite d'un groupe accentuel avec celle d'un groupe syntaxique) et sur des contraintes rythmiques, à savoir sur des principes d'équilibre syllabique (symétrie, récurrence et eurythmie) qui régissent et limitent la taille des groupes accentuels. Voir Lacheret-Dujour (1999) pour une étude approfondie des modèles intonatifs de la phrase française.

[11] Grammont (1914) fait correspondre à ces groupes la notion de « unités de sens » ou « groupe d'idée».

« Toute suite de mots qui exprime une idée simple et unique constitue un seul élément rythmique et n'a d'accent que sur sa dernière syllabe » (Grammont 1914, p.105).

[12] Au sein des unités prosodiques minimales, les éléments peuvent être sujets à désaccentuation tout en gardant un accent final sur la dernière syllabe du groupe (Delais-Roussarie 1993).

[13] D'après Laks (2005) le mot ne serait qu'une projection de la norme orthographique. En revanche, certains auteurs parmi lesquels Adda-Decker, Nemoto et Durand (2009) défendent l'existence d'indices prosodiques et phonétiques de démarcation du mot en français.

> and its suprasegmental properties. The French 'cursus' is therefore in both respects, segmental and suprasegmental, a single word, hence a true phonological word ». (Pulgram 1970, p. 35)

C'est généralement à l'intérieur de ces groupes de souffle ou groupes rythmiques, où les frontières lexicales sont effacées, que l'enchaînement et la liaison se réalisent.

> « On lie dans l'intérieur d'un élément rythmique, on ne lie pas d'un élément rythmique au suivant, en d'autres termes : on lie d'une syllabe inaccentuée sur la suivante, on ne lie pas d'une syllabe accentuée ». (Grammont 1914, p. 129)[14]

2.3. La liaison : un cas particulier de sandhi externe

Comme nous l'avons souligné, la liaison constitue un phénomène de *sandhi* externe de la langue française qui correspond à la prononciation d'une consonne à la frontière de deux mots dont la réalisation est soumise à un conditionnement syntaxique (fort degré de cohésion syntaxique entre les deux mots) et phonologique (le deuxième mot commençant par une voyelle)[15]. En effet, la liaison est réalisée entre deux mots lorsque : (a) le deuxième mot commence par une voyelle (conditionnement phonologique) ; (b) il existe un lien syntaxique étroit entre les deux mots (conditionnement syntaxique) à savoir le premier mot (dorénavant M1) et le deuxième mot (dorénavant M2). Le degré de cohésion syntagmatique de la séquence constitue donc une condition nécessaire pour la réalisation de la liaison sans laquelle la consonne de liaison ne pourrait pas être réalisée. La spécificité du phénomène de liaison réside précisément dans ce conditionnement à la fois phonologique et syntaxique qui rend ce type de *sandhi* externe unique.

Bien que les deux phénomènes de liaison et d'enchaînement soient étroitement liés, leur rapport ne peut pas être défini en terme d'implication (Burov 2012). Si l'enchaînement consonantique produit une resyllabation de tout type de consonne finale prononcée, le processus de *sandhi* de liaison, comme nous le verrons dans le paragraphe suivant, ne concerne que certaines consonnes particulières. Ces consonnes finales dont la réalisation est soumise à un conditionnement syntaxique et phonologique sont appelées consonnes de liaison (dorénavant CL)[16] et possèdent donc un statut particulier qui les distingue des autres consonnes, dites fixes qui sont par définition toujours réalisées.

[14] Bien que la liaison se réalise le plus souvent à l'intérieur des groupes accentuels, comme nous le verrons dans les pages suivantes, certains auteurs (Encrevé 1988 ; Morin et Kaye 1982) ont démontré que la liaison peut être réalisée aussi bien après une pause (temporelle ou intonationnelle).

[15] Étant donné que l'attachement lexical ou phonologique des consonnes de liaison (cf. Partie 1.Chapitre 5.2) ne fait pas encore à ce jour l'objet d'un consensus, nous préférons à la définition traditionnelle proposée par Fouché (1959) la description proposée par Côté (2015) d'après laquelle « la liaison correspond à la prononciation d'une consonne à la frontière de deux mots (M1, M2), le M2 commençant par voyelle » (Côté 2015).

[16] Plusieurs définitions de cette consonne ont été proposées selon les approches que nous nous attacherons à présenter dans les prochains chapitres (cf. Partie 1.Chapitre 5).

Si nous revenons sur le rapport existant entre enchaînement et liaison, nous pouvons définir l'enchaînement comme un phénomène d'ordre général qui intéresse tous les segments finals se trouvant à la frontière lexicale et détermine leur resyllabation dans toutes les langues *cursus* (cf. Paragraphe 2.2). Quant à la liaison, elle constitue un phénomène unique de la langue française qui n'affecte que les CL qui peuvent ou non être l'objet d'un enchaînement (cf. Partie 1.Chapitre 2.2.1.1)[17].

3. Consonnes finales : consonnes fixes et consonnes de liaison

Comme nous l'avons évoqué dans le paragraphe précédent, on considère traditionnellement que, contrairement aux CL, les consonnes fixes ne sont pas sujettes à chute ou parce qu'elles sont suivies d'un *schwa* [ə] (comme dans l'adjectif « grande ») ou parce qu'elles constituent des exceptions lexicales (Côté 2005). En particulier, Fouché (1959, p. 376-409) et Léon (1978, p. 74-116) développent une étude détaillée du traitement des consonnes finales en fonction des cas où elles sont prononcées ou « muettes ».

D'autres arguments que ceux purement lexicaux ou phonétiques ont été évoqués afin de distinguer les CL des consonnes finales fixes, comme par exemple le degré de cohésion lexico-phonologique et morphosyntaxique entre la CL et le M2. Ce lien serait plus fort que celui qui unit les consonnes finales fixes au mot suivant. Cela s'explique par le fait que si les consonnes fixes sont dans la plupart des cas indépendantes du mot suivant, une des conditions de réalisation de la liaison est la présence d'un M2 étroitement lié au M1 auquel appartient graphiquement la CL (Côté 2010, p. 1285).

Cependant ces critères ne clarifient qu'en partie le rapport ambigu existant entre les CL et les consonnes finales fixes. En effet le débat autour de cette opposition n'est pas purement terminologique puisque la définition des frontières entre ces deux types de consonnes oblige à une prise de position quant au traitement et au statut lexical des CL. Les analyses traditionnelles de la liaison (Fouché 1959 ; Delattre 1966) définissent la CL comme une consonne finale (donc rattachée au M1) se prononçant devant des M2 à initiale vocalique. Si d'un point de vue diachronique les CL peuvent être traitées comme des consonnes finales du M1 (cf. Paragraphe 5), certaines études envisagent d'autres possibilités de rattachement (notamment une affiliation

[17] À la suite de la présentation de la dimension phonologique d'analyse de la liaison, dans le chapitre suivant (cf. Partie 1.Chapitre 2.2.1.1), nous nous attacherons à décrire le cas de liaisons non-enchaînées.

lexicale au M2 ou une autonomie de la CL par rapport au M1 et M2) que nous nous attacherons à décrire dans les prochains chapitres (cf. Partie 1.Chapitre 5.2).

4. Typologies des CL

En ce qui concerne la typologie des CL, la littérature établit une liste totale de sept CL : [z], [t], [n], [ʀ], [p], [g] et [k] dont les trois premières sont les plus fréquentes. Néanmoins les positions des linguistes divergent quant à l'ordre de fréquence de ces CL. Si Malécot (1975) suggère que les liaisons en [n] sont plus fréquentes que les liaisons en [z] et [t], Léon (1992) propose un ordre différent. En effet d'après Léon (1992) 50% environ des cas possibles de liaisons sont en [z] et que [t] et [n] se partagent la presque totalité des 50% restants. Les liaisons avec [ʀ] et [p] représentent moins de 1%. Malécot (1975) et Léon (1992) avaient observé des taux de fréquence relative différents des CL qui correspondent respectivement aux ordres /n/ > /z/ > /t/ (Malécot 1975) et /z/ > /t/ > /n/ (Léon 1992).

En revanche, les fréquences relatives des CL issues du corpus de français parlé « Phonologie du Français Contemporain » (Durand, Laks et Lyche 2009; Mallet (2008) (cf. Partie 2.Chapitre 1) montrent que, contrairement aux taux relevés par Malécot (1975)[18] la liaison en [z] est plus fréquente que la liaison en [n] qui possède un taux plus important de réalisation par rapport à la liaison en [t].

En ce qui concerne les graphèmes correspondant à ces CL, en tant que phénomène de *sandhi* la liaison détermine dans certains contextes la production de changements phonétiques qui peuvent affecter aussi bien les consonnes orthographiques finales du M1 (dévoisement ou voisement consonantique) que les voyelles nasales du M1 (dénasalisation) (Fouché 1959 ; Delattre 1966).

La liaison en [z] peut correspondre à plusieurs graphies, parmi lesquelles on compte également les graphèmes qui caractérisent le pluriel en français à savoir < s > et < x >:

- < s > : comme dans « les amis » [lezami]. Dans ce contexte la consonne < s > est soumise à un processus de voisement lors de la réalisation de la liaison ;
- < z > : comme dans « chez elle » [ʃezɛl] ;
- < x > : comme dans « deux autres » [døzotʀ].

[18] Il est possible que cette différence relève d'une variation diachronique (cf. Partie 1.Chapitre 2.2.6.1.2.).

La liaison en [z] représente donc une marque du pluriel. Cette spécificité a assuré à ce type de liaison une certaine stabilité au cours de l'évolution diachronique du phénomène (cf. Paragraphe 5).

À la liaison en [t] correspondent deux différents graphèmes :

- < t > : comme dans « voit-elle » [vwatɛl] ;
- < d > : comme dans « quand il pleut » [kɑ̃tilplø]. L'occlusive dentale sonore /d/ est soumise dans ce cas à un changement phonétique qui détermine son dévoisement.

La valeur de cette liaison joue dans certains contextes un rôle crucial et distinctif permettant de faire la différence entre le masculin (où la liaison peut être réalisée) et le féminin (pour lequel un simple enchaînement se produit et donc la consonne enchaînée demeure /d/).

En ce qui concerne la consonne de liaison [n], (ex. « bon ami » [bɔnami]), elle est le résultat d'un changement phonétique, à l'occurrence une dénasalisation d'une voyelle nasale. Cette modification entraîne dans certains cas une correspondance entre la forme phonique des adjectifs masculins devant un mot à initiale vocalique et celle des adjectifs féminins (ex. « bon » et « bonne »). Dans les pages qui suivent (cf. Partie 2.Chapitre 2.2.2.3.1), nous verrons que cet aspect amènera plusieurs auteurs (parmi lesquels Steriade 1999) à traiter ces formes de liaison comme dérivant du paradigme du féminin du M1.

Quant aux consonnes de liaison [ʀ] et [p], elles correspondent respectivement aux graphèmes < r > comme dans « premier étage » [pʀəmjeʀetaʒ] et < p > comme dans « beaucoup aimé » [bokupeme].

Pour terminer l'inventaire des CL, les consonnes [g] et [k] ne sont prononcées que dans un nombre très limité de séquences. Par exemple [g] et [k][19] correspondant au graphème < g > ne se trouvent que dans des expressions telles que « long été », « long hiver », « croc-en-jambe » et « Bourg-en-Bresse » etc.

Comme nous l'avons vu dans les paragraphes précédents, les consonnes orthographiques < d > et < s > correspondent aux CL [t] et [z]. La deuxième typologie de changements phonétiques, en revanche, intéressent les voyelles nasales finales du M1. Contrairement au cas précédent, les modifications phonétiques sont moins systématiques et la voyelle nasale conserve sa nasalité suivant le contexte et la typologie de la voyelle nasale[20]. Fouché (1959) souligne que la voyelle [ɛ̃] dans les mots se terminant en « -ain » et « -ein » devient orale

[19] Néanmoins cette CL n'est aujourd'hui plus réalisée même dans la Marseillaise (ex. « sang impur »).

[20] Comme nous le verrons dans les prochains chapitres (cf. Partie 1.Chapitre 5.2.1.2) certains auteurs rejettent l'hypothèse de la dénasalisation des voyelles nasales en faveur d'une analyse supplétive de la liaison en [n].

comme dans « certain auteur » [sɛʀtɛnotœʀ]. Il en est de même pour les terminaisons en « -ien » et « -yen » dans « ancien », « moyen » et « mien », « tien » et « sien » comme dans « un ancien état ». Toutefois, dans « bien » et « rien », la voyelle conserve sa nasalité comme dans « rien à faire » [ʀjɛ̃nafɛʀ]. Parmi les adjectifs en « -in », Fouché (1959) distingue le cas de « malin » et celui de « divin »[21].

Quant à la voyelle [ɑ̃] la préposition « en » en liaison conserve la voyelle nasale (ex. « en effet » [ɑ̃nefɛ] ainsi que la voyelle [ɔ̃] du pronom « on » (ex. « on a dit » [ɔ̃nadi]). Cependant, celle-ci devient souvent orale dans le cas des adjectifs antéposés (ex. « bon ami » [bɔnami]). Fouché (1959) signale également la présence d'une alternance entre voyelle orale et voyelle nasale dans les adjectifs possessifs « mon », « ton » et « son » bien qu'il précise que « la prononciation avec voyelle nasale est considérée pour le moment comme la meilleure ». Pour terminer le traitement des voyelles nasales en contexte de liaison, la voyelle [œ̃] dans les mots « aucun », « commun » et « un » en liaison conserve la voyelle nasale (ex « un enfant » [œ̃nɑ̃fɑ̃]).

Pour résumer, les CL ([z], [t], [n], [ʀ], [p], [g] et [k]) sont des consonnes qui ne se prononcent pas en position isolée et sont susceptibles d'être réalisées devant un mot à initiale vocalique lorsqu'il existe un lien syntaxique étroit entre les deux mots (M1 et M2).

Les CL, nous le rappelons, possèdent un statut particulier[22] qui le distingue des autres consonnes, dites fixes (qui sont par définition toujours réalisées ex. « net » [nɛt]). Cependant, d'un point de vue diachronique ces consonnes n'ont pas toujours été « muettes ». En effet, comme le souligne Delattre (1947, p. 148), elles représentent une trace de la survivance de l'enchaînement des consonnes finales de l'ancien français lorsque « toutes les consonnes finales que l'on écrivait étaient prononcées (ah ! le bel âge !) » (Delattre 1947, p.148).

[21] « Malin » conserve la nasalité de la voyelle dans l'expression « le malin esprit » alors que la voyelle devient orale dans des formes telles que « divin Enfant » et « le divin amour ».

[22] Nous évitons ici d'avoir recours aux notions de « consonne latente » et « consonne muette ». Nous nous attacherons dans les prochains chapitres (cf. Partie 1.Chapitre 5.2) à analyser les différentes approches proposées pour décrire le statut des CL.

5. Évolution diachronique des consonnes finales en français

Les origines du phénomène de la liaison, comme nous l'avons évoqué dans le paragraphe précédent, sont à rechercher dans l'histoire du traitement des consonnes finales en ancien français. En effet avant cette époque, les consonnes finales étaient prononcées dans tous les contextes. À partir du XIIème au XIIIème siècle, les consonnes finales ont commencé à disparaître, ne se maintenant d'abord que devant des mots à initiale vocalique et devant une pause et chutent ensuite également dans le mot isolé (Zink 2006 ; Wartburg 1988).

La préservation des consonnes finales dans le contexte de liaison peut s'expliquer par la tendance à la syllabation ouverte très répandue dans tout le groupe des langues romanes (Martinet 1964) qui, comme nous l'avons vu, remonte au latin classique (Niedermann 1953), perdurant au cours des siècles du latin vulgaire (Vâânânen 1962), à l'ancien français jusqu'au français moderne (Pulgram 1970). Il s'agit en effet d'une tendance très ancienne qui se réalise aussi bien par la chute complète des consonnes finales que par leur maintien grâce à la resyllabation à l'initiale vocalique du mot suivant. Selon Martinet (1964, 328) elle relève d'une dynamique universelle qui favorise la stabilisation de la syllabe et donc la création d'une suite CV.

En ce qui concerne le latin classique, la structure prosodique descendante (fort-faible) du latin (Niedermann 1953) favorise une syllabation ouverte. Niedermann (1953) atteste qu'en latin classique les consonnes finales pouvaient être resyllabées devant un nom commençant par une voyelle donnant lieu à une syllabe ouverte finale du type CV où la voyelle devenait brève. C'est le cas par exemple d'une séquence telle que *bonus ēventus* qui devant un mot à initiale vocalique était prononcé *bonu sēventus*.

> « La dernière syllabe de mots tels que *bonus, canis, flūmen, procul, tegit* s'ils étaient placés devant un mot commençant par une voyelle, […] leur consonne finale [s'y] rattachait par liaison dans la chaîne du discours. *Bonus ēventus, procul abest,* étant prononcés *bonu sēventus, procu labest* ». (Niedermann 1953, p. 173)

De plus, l'auteur précise que ces resyllabations pouvaient être bloquées dans les cas où ces mêmes syllabes finales étaient suivies d'un mot à initiale consonantique ou devant une pause, « où la consonne finale ne pouvait pas se détacher par liaison, de sorte que la syllabe dont elle faisait partie demeurait fermée ».

Cette tendance est également présente en latin vulgaire où la disparition de l'opposition de longueur vocalique supplantée par l'opposition de hauteur détermine l'affaiblissent des syllabes post-toniques et donc leur progressive chute. On assiste donc déjà en latin vulgaire à un affaiblissement des consonnes finales qui subissent des changements selon qu'elles se trouvent en position isolée ou qu'elles appartiennent à un mot plus ou moins lié au mot suivant.

> « Les consonnes finales ont une articulation relativement faible et subissent, par conséquent, des modifications variées selon qu'elles se trouvent à la pause ou qu'elles appartiennent à un mot plus ou moins lié au mot suivant. Dans ce dernier cas, leur traitement dépend de la nature du mot qui suit ». (Vâânânen 1962, p. 66)

Cependant, cette tendance à l'assourdissement total des consonnes finales s'oppose à l'exigence de préserver l'information morphosyntaxique véhiculée par les consonnes à valeur fonctionnelle qui préfigure le marquage morphologique du nombre du français moderne. En effet, Vâânânen (1962) souligne que les consonnes finales < t > et < s > se sont mieux conservées en gallo-roman que dans d'autres langues néolatines grâce à leur fonction désinentielle.

> « […] l'effacement total des consonnes finales a été contrecarré par le souci de netteté, surtout à l'égard des consonnes remplissant des fonctions flexionnelles; à comparer en français moderne, les deux (et même trois) prononciations de 'plus' et 'tous'». (Vâânânen 1962, p. 66)

En particulier, l'auteur fait remarquer que la consonne finale < s > a pu résister en latin vulgaire grâce sans doute au rôle crucial que cette consonne joue sur le plan syntaxique et morphologique. Plusieurs matériaux épigraphiques provenant de Pompéi témoignent de la stabilité du < s > final. La consonne finale < m > est également soumise à un affaiblissement surtout devant une voyelle initiale mais le gallo-roman la conserve dans quelques monosyllabes : *revĕm* > *rien*, les possessifs masculins atones vulgaires **muŭn, *tuŭm, *suŭm* > *mon, ton, son* et les dissyllabes toniques correspondants *méum, tŭum, sŭum* > *mien, tuen, seun* mais non dans les possessifs féminins vulgaires **mam, *tam, *sam* > *ma, ta, sa*. Il en va de même pour < n > final qui commence à tomber déjà en latin pré-littéraire après une voyelle longue (**ration* > **ratĭo*), s'efface en latin vulgaire après une voyelle brève dans les plurisyllabes *flumen* > *flum* et ne se maintient que dans des monosyllabes *ĭn* > *en*, *non* > *non* (Zink 1986, p. 75).

Pendant la période « proto-français » (Zink 1986), les consonnes finales semblent connaître une période de stabilité qui durera jusqu'au IXème siècle où après une série d'assourdissements on assiste à deux différentes périodes d'effacement, l'une qui aura lieu entre le IXème et le XIème siècle et qui intéressera les deux consonnes dentales, et l'autre entre le XIIIème et le XVIème siècle qui concernera toutes les autres consonnes finales. Ici, la

dynamique de syllabation ouverte fait de nouveau surface et se traduit par une tendance à l'effacement des consonnes finales de l'ancien français.

Comme nous venons de le dire ces tendances à l'effacement seront précédées par des changements phonétiques des consonnes finales à savoir des assourdissements qui auront lieu au VIIème siècle (*plŭmbu(m) > plo'mbo > plomp*; *grande(m) > grande > grant* ; *longu(m) > longo > lonk* ; *clasu(m) > klozo > klos* ; *voce(m) > voydze > voyts*. La consonne finale <m> tendra à se dentaliser *rĕm > rien ; *mum > mon.*

Ensuite, entre le IXème et le XIème siècle, on assiste à l'affaiblissement des deux dentales /t/ et /θ/ après une voyelle simple. La première se spirantise et l'interdentale disparaît. De plus, la précocité ou non de la chute de la consonne < t > est conditionnée par sa place par rapport à l'accent et par le phonème qui suit. L'amuïssement se produit d'abord après une voyelle atone, à distance de l'accent, et ensuite après une voyelle tonique. Après une consonne les consonnes dentales demeurent jusqu'au XIIIème siècle, que l'appui soit étymologique comme dans *grant* ou secondaire comme dans *deivt (débet)* ou qu'il vienne d'une semi-consonne non encore vocalisée, comme dans *fayt* (*factum*) (Zink 1996, p. 210).

À la fin du XIIIème siècle un nouveau courant d'effacement affecte l'ensemble des consonnes finales. Ce courant aura lieu progressivement entre le XIIIème et le XVIème siècle. Comme le souligne Zink (1986), l'évolution de ce phénomène est conditionnée par la phonétique syntactique.

> « Etant donné la place de la consonne, le déroulement du processus relève de la phonétique syntactique qui l'accélère devant une initiale consonantique et le retarde devant une voyelle où la consonne fait office de liaison. Sur ce point, la sifflante *-s* semble avoir bénéficié d'un traitement particulièrement net qui la désigne comme exemple. Elle reste sourde à la pause, se sonorise devant voyelle, s'efface devant consonne. […] Les enchaînements en liaison étroite sonorisent encore aujourd'hui en *-s* final : *les* (*des*, *deux*, *de grands*) *hommes*, mais assourdissent les occlusives ; *un grand homme*, *vend-il* ? […] et l'usage reste flottant pour *-f* ». (Zink 1986, p. 78)

En effet, comme nous l'avons déjà évoqué, à partir de l'ancien français les consonnes finales ont disparu ne se maintenant d'abord que devant des mots à initiale vocalique (donc en contexte de liaison) et devant pause et ensuite chutant également dans le mot isolé.

La forte tendance à la syllabation ouverte qui caractérise la prononciation de l'ancien français déterminera l'enchaînement et donc la création d'un lien syllabiquement étroit entre les consonnes finales du M1 et la voyelle initiale du M2. Ce lien (à la fois phonologique et syntaxique) évite leur amuïssement.

En ce qui concerne la consonne < s >, elle s'est effacée, d'abord devant une consonne commençant le mot suivant tandis qu'elle s'est conservé comme sonore devant une initiale vocalique et comme sourde à la pause. L'origine de la liaison en [z] se trouve donc dans ce changement qui a intéressé la consonne finale < s > entre le XIIIème et le XVIème siècle.

> « […] depuis le *s* est resté sensible (avec la valeur *z*) dans les cas de liaison étroite devant la voyelle […] Mais, à la pause, il ne se prononce plus que dans les noms de nombre *six, dix* où on l'écrit *x* d'après le lat. *sex*, dans les adjectifs *tous*, dans l'adverbe *sus* ». (Bourciez 1958, p. 158)

Cette dynamique d'effacement réglée par une loi de phonétique syntaxique perdurera également pendant le XVIème siècle.

> « La prononciation des consonnes finales est régie au XVIème siècle par une loi de phonétique syntaxique : les consonnes finales se prononcent devant un mot commençant par une voyelle (c'est notre « liaison ») et à la pause ; elles ne se prononcent pas devant une consonne […]. La liaison était la conséquence pure et simple de la loi de prononciation des consonnes finales […]. Elle n'était pas restreinte comme aujourd'hui, à l'intérieur des groupes de mots étroitement unis par le sens. Elle se faisait automatiquement du mot à finale consonantique au mot à initiale vocalique ». (Gougenheim 1951, p. 31)

Cependant, à la dynamique d'effacement, que nous venons de décrire, s'oppose une tendance antagoniste de préservation des consonnes finales structurales conditionnée par les mêmes « exigences morphosyntaxiques » qui avaient également assuré la conservation des consonnes à valeur morphosyntaxique en latin vulgaire.

> « La position finale est, par définition affaiblissante, mais elle n'en est pas moins la place obligée des marques, tant casuelles du nom et de l'adjectif que personnelles du verbe, qui se doivent de rester claires. D'où des conflits entre les courants phonétiques qui portent à l'érosion et les exigences morpho-syntaxiques qui portent à la conservation. Suivant les époques, les parlers, les niveaux de langue on verra se prédominer l'une ou l'autre tendance ». (Zink 1986, p. 75)

Bien que la langue savante semble respecter ces règles d'effacement, à partir du XVIème siècle la langue « populaire » tend à unifier la prononciation en généralisant l'effacement à tous les contextes syntaxiques. Au même moment, les grammairiens du XVIème siècle commencent à mettre en place un important travail de rétablissement des consonnes finales qui avaient disparu. En particulier, la consonne [r] semble garder les traces de cette intervention conservatrice. Les grammairiens du XVIème siècle rétablissent temporairement le [r] dans la plupart des terminaisons des morphèmes désinentiels ou suffixaux : « -er », « ier », « eur », « -ir », « -oir » (ex. « premier ») qui avaient été effacés au XVème et au XVIème siècle (sauf dans le cas de liaisons étroites comme aujourd'hui dans « le premier homme »).

À cette époque, deux niveaux de langue sont attestés sous l'effet de deux courants antagonistes : les grammairiens (Meigret 1542) et ceux qui défendent une forme de langue plus proche de l'usage oral (Estienne 1557) qui donc s'oppose au rétablissement massif des consonnes qui avaient disparu dans la graphie ainsi que dans la prononciation (Gougenheim 1951). Les arguments soutenus par les premiers sont principalement de caractère étymologique et grammatical. La réintroduction des consonnes finales permettrait de garder une valeur fonctionnelle et étymologique de l'écriture.

Ce débat mènera à plusieurs solutions du point de vue orthographique telles que la notation diacritique et la distinction entre lettre de signification et lettre de prononciation (Gilles de Vaudelin 1713 cité par Martinet 1974, p. 156)[23]. Dans un premier temps les normes orthographiques et d'usage de la liaison ne sont pas encore fixées et la liaison est encore l'objet d'une profonde variation stylistique et sociale (Bourciez et Bourciez 1967, p. 123).

6. Etablissement d'une norme orthographique et d'usage de la liaison

La « norme » orthographique et l'usage de la liaison, nous l'avons vu dans le paragraphe précédent, n'est pas encore fixée aux XVème et XVIème siècles (Bourciez et Bourciez 1967, p. 123). Certaines liaisons semblent encore suivre un comportement irrégulier, par exemple la liaison en [z] des noms et des adjectifs devant une pause. Thurot (1881, p. 61) atteste également la présence de productions « perturbées » de liaison. En particulier, on relève des cas où la consonne [s] remplace les dentales [t] et [d] devant un mot à initiale vocalique[24]. Parmi ces « perturbations » Domergue (1805) mentionne aussi l'alternance entre les consonnes [s] et [t] entre « point » et « pas » qui donnera lieu à l'anecdote à l'origine du terme «pat-à-qu'est-ce » pour désigner les fautes de liaison[25].

Thurot (1881) précise que ces remplacements avaient pour but d'éviter le *hiatus* et plus généralement des cacophonies, mais qu'elles ne se sont pas établies dans l'usage.

[23] Dans cette perspective les consonnes finales, contrairement aux voyelles qui assurent l'identité du mot, ont une fonction euphonique qui évite la production des *hiatus*.

[24] « un laid [z] homme » Cauchie; « à tort [z] et à travers » Estienne ; « avant [z] hier » Buffet (cités par Thurot 1881, p. 61).

[25] « Un beau diseur étoit au spectacle dans une loge, à côté de deux femmes, dont l'une étoit l'épouse d'un agioteur ci-devant laquais ; l'autre d'un fournisseur, ci-devant savetier… Tout à coup, le jeune homme trouve sous sa main un éventail. 'Madame', dit-il à la première, 'cet éventail est-il à vous ? Il n'est poin-z-à moi. Est-il à vous, en le présentant à l'autre ? Il n'est pa-t-à moi'. Le bon diseur, en riant : 'Il n'est poin-z à vous, il n'est pat-à –vous, je ne sais pa-t-à qu'est-ce'. Cette plaisanterie a couru dans les cercles, et le mot est resté ». (Domergue 1805, p. 464)

En effet, la prononciation et l'orthographe des consonnes finales au XVIIème siècle fait encore l'objet d'une profonde variation.

> « […] il s'est fait un compromis entre l'ancien usage où la consonne finale s'articulait avec la voyelle initiale de n'importe quel mot suivant et les tendances populaires récentes qui allaient à l'amuïssement général de la consonne. La date où ce fait s'est produit peut être fixée au second quart du XVIIe siècle ». (Brunot 1966)

La fixation des règles de prononciation de la liaison sera donc accomplie au cours de la première moitié du XVIIème siècle. À ce moment, des modèles linguistiques basés sur le registre de langue de la Cour sont établis afin de définir le « bon usage » de la langue française (Trudeau 1992). En particulier, l'ouvrage « Traicté de la Grammaire françoise » de Robert Estienne (1557) représente selon Trudeau (1992) la première grammaire ayant comme objectif de distinguer le bon usage de la langue des gens de la cour, à savoir « l'usage approuvé », de « l'usage commun ».

> « Le *Traicté* de Robert Estienne est le premier ouvrage grammatical qui se donne explicitement pour une description de la langue légitime, laquelle est identifiée à la pratique des gens de cour et épurée des constructions douteuses et des dialectalismes […] sa grammaire accorde ainsi les pleins pouvoirs sur la langue à la classe politique ». (Trudeau 1992, p. 86-87)

En effet, contrairement à ses prédécesseurs (Meigret 1542), pour Estienne (1557) l'objectif de la grammaire n'est plus celui de perfectionner la langue des gens cultivés et de conseiller les « courtisans », mais de diffuser les usages d'un groupe particulier en les présentant comme la seule pratique légitime. Le rôle du grammairien n'est donc pas celui de montrer le fonctionnement de la langue, de comparer les dialectes ou d'améliorer l'usage mais plutôt d'enregistrer et transmettre le registre des individus les plus hauts placés dans la société. Cela donne à l'ouvrage un caractère plus descriptif et plus catégorique que les grammaires des prédécesseurs. La légitimité de cet « usage approuvé » décrit par Estienne (1557) sera ensuite renforcée par Vaugelas (1647) qui s'attache à définir les groupes sociaux qui détiennent l'autorité sur la langue[26]. L'ouvrage de Vaugelas (1647), « Remarques sur la langue françoise », représente un manuel voué à la fois à affirmer la supériorité linguistique de l'aristocratie de la Cour et à assurer la poursuite du bon usage chez les courtisans. Toute motivation de l'usage est

[26] Comme le rappelle Le Prieult (2006) la norme d'usage est une notion qui se définit et change selon les époques en fonction des modèles grammaticaux et souvent socioculturels de référence.

> « […] il s'agit d'une notion qui peine, à travers les différentes acceptions que l'histoire a pu lui donner (usage commun, usage de Cour, usage littéraire, standards divers), à se stabiliser autour d'une définition stricte. Chaque époque semble avoir construit la notion, en réaction ou en prolongement aux coutumes grammaticales qui l'ont nourrie ». (Le Prieult 2006, p. 120)

donc écartée et il suffit qu'« une expression soit employée ou rejetée par les courtisans pour être désignée comme *bon usage* ou *mauvais usage* » (Trudeau 1992, p.194).

Ce modèle exemplaire de langue fixée par les grammairiens normatifs de l'époque (Estienne 1557 ; Vaugelas. 1647) a servi à proscrire les usages différents comme incorrects et par conséquent à établir une norme d'usage de la liaison dont la définition sera assurée par l'enseignement et la fixation de l'orthographe. Celle-ci tend à se stabiliser au cours du XVIIIème siècle sous l'effet de plusieurs facteurs et une « orthographe étymologique » finira par s'imposer (Catach 1968).

> « De phonétique qu'est toujours plus au moins la notation écrite d'une langue à ses débuts, celle du français s'est ensuite détachée de ses origines et des autres langues romanes. Son orthographe […] est devenue savante et étymologique […] ». (Catach 1968, p. 259)

Bourciez et Bourciez (1967) soulignent que les témoignages des grammairiens comme Corneille et Hindret attestent que vers la fin du XVIIème siècle, l'usage de la liaison était déjà sensiblement proche de celui d'aujourd'hui. L'effort des grammairiens de rétablir l'usage des consonnes s'accompagne de la diffusion progressive de l'écriture à travers la mise en place de l'école publique qui déterminera la fixation d'un grand nombre de consonnes. À cette époque, comme le souligne Brunot (1966, VI, p. 2) le renversement entre tradition orale et tradition écrite est accompli, la norme écrite transmise par l'école s'impose sur la langue orale. À ce moment, la norme écrite se diffuse à tout niveau social (Blanche-Benveniste et Chervel 1969), posant les bases pour une adéquation de la tradition orale à la norme écrite.

Laks (2005b) illustre bien ce phénomène selon lequel la fixation d'une grammaire normative et de l'orthographe ainsi que la mise en place de l'école publique conduiront à une imposition de la norme écrite dans la réalisation de la liaison.

> « Le mécanisme normatif de la liaison moderne qui opérationnalise les consonnes graphiques se met alors rapidement en place […] et c'est l'œil désormais qui impose la norme de la liaison. […] L'apprentissage des usages normés de prononciation ne se fait plus par imprégnation/imitation *naturelle*, il s'opère à l'école par l'exercice de la lecture et de la déclamation des vers. Les modalités changent donc et c'est, nous l'avons déjà souligné, par l'œil désormais qu'on apprend la phonie ». (Laks 2005b, p. 119)

Les origines du phénomène de la liaison moderne sont à rechercher dans cette période où, à côté des consonnes à valeur structurale qui avaient résisté aux différentes vagues d'effacement, d'autres consonnes seront rétablies à la suite d'un procès de fixation de l'écriture et d'une forte intervention des grammairiens. Il est donc possible d'identifier deux types de traitements dans l'évolution diachronique des consonnes finales du français qui donneront lieu à des CL de nature différente. D'une part, les consonnes finales ayant une valeur morphologique

(ex. < s > marque du pluriel) qui a assuré leur préservation au fil des siècles des amuïssements qui ont intéressé le latin vulgaire jusqu'aux vagues d'effacements du XVème au XVIème siècle. D'autre part, des consonnes plus faibles n'ayant pas de valeurs structurales qui ont été d'abord effacées et ensuite rétablies par les grammairiens du XVIème siècle dont la réalisation est tributaire de la fixation de la norme écrite.

Cette différentiation dans le développement diachronique des consonnes finales n'est pas anodine et sera à l'origine de deux traitements de la liaison dans le français moderne (Laks 2005b). D'un côté, des liaisons qui à cause de leur caractère fonctionnel sont toujours produites et reçoivent donc un traitement morphologique. De l'autre, des liaisons proprement phonologiques dont la réalisation variable dépend de l'imposition des consonnes écrites rétablies ainsi que de la combinaison de différents facteurs syntaxiques, prosodiques et lexicaux (cf. Partie 1.Chapitre 2).

7. Conclusion

Le contexte conceptuel et historique que nous avons évoqué dans ce premier chapitre a été l'occasion de décrire la liaison en tant que phénomène de *sandhi* externe du français tout en le distinguant de celui de l'enchaînement consonantique qui s'applique à tout type de consonne finale prononcée (CL ou consonne fixe). En effet, le processus de *sandhi* de liaison ne concerne que des consonnes qui en vertu de leur statut particulier sont susceptibles de ne pas être prononcées en dehors de certains conditionnements syntaxiques et phonologiques.

Les origines des CL de la liaison moderne ont été recherchées dans les changements diachroniques qui ont intéressé les consonnes finales dans le français du XVème au XVIème siècle. Ce n'est en fait qu'à la fin du XVIIème siècle que les usages modernes se définissent à la suite d'un procès de fixation de l'écriture et de l'effort des grammairiens. À ce moment, la norme écrite s'impose dans la réalisation de la liaison grâce aussi à la mise en place de l'école publique.

Le parcours historique que nous avons retracé ci-dessus nous a permis d'évoquer la distinction entre d'une part des consonnes de liaison à valeur fonctionnelle qui avaient résisté aux différentes vagues d'effacement et d'autre part des liaisons phonologiques issues d'une prononciation restaurée à partir de l'orthographe. Toutefois, il est évident que cette opposition ne suffit pas à rendre compte du comportement hétérogène de la liaison.

Toute recherche portant sur la liaison demande d'abord une étude de la pluralité des facteurs linguistiques (syntaxiques, phonologiques, phonétiques, lexicaux, prosodiques, sociolinguistiques) qui influencent sa variabilité.

Chapitre 2. Variables linguistiques d'étude de la liaison

Dans le chapitre précédent (cf. Partie 1.Chapitre 1) nous avons retracé les origines historiques de la liaison qui nous permettront, dans les pages qui suivent, de mieux saisir la diversité des comportements qui caractérisent la production de ce phénomène. Nous avons évoqué une différence historique entre des liaisons ayant un caractère fonctionnel (dont la réalisation de la CL s'appuie sur des consonnes à valeur morphologique) et des liaisons de nature phonologique (dont la production se base le plus souvent sur la prononciation de consonnes lexicales).

Dans les pages qui suivent, nous approfondirons dans un premier temps l'opposition entre ces deux différentes types de liaison et nous verrons que celle-ci non seulement ne permet pas d'expliquer les comportements variables de la liaison mais qu'elle ne parvient pas non plus à distinguer les liaisons qui sont toujours réalisées par tout locuteur et dans tout contexte (appelées obligatoires ou catégoriques selon les auteurs cf. Partie 1.Chapitre 4.2.1) de celles dont la réalisation varie en fonction de la combinaison de différents facteurs linguistiques (liaisons facultatives ou variables selon les auteurs)[27].

Cela nous amènera à étudier l'ensemble des variables d'analyse de la liaison (Mallet 2008; Durand et Lyche sous presse). Nous nous attacherons tout d'abord à décrire les critères syntaxiques qui influencent la réalisation de la liaison, ce qui nous permettra de faire la différence entre liaisons antéposées et postposées à une catégorie syntaxique principale, puis nous décrirons les autres variables notamment phonologiques, prosodiques, sociolinguistiques et lexicales qui ont été prises en compte dans l'étude de la liaison.

1. Liaisons morphosyntaxiques et phonologiques

Nous avons évoqué une distinction traditionnellement établie entre d'une part des liaisons morphosyntaxiques, correspondant à la réalisation de consonnes qui ont historiquement une valeur fonctionnelle et d'autre part des liaisons phonologiques dont la production de la CL s'appuie sur des consonnes lexicales. Le rôle fonctionnel des liaisons semble assurer, comme

[27] Nous nous attacherons dans le quatrième chapitre (cf. Partie 1.Chapitre 4) à décrire les classifications et les terminologies utilisées dans la littérature. Dans ce chapitre préliminaire il nous paraît intéressant d'étudier les variables d'analyse de la liaison sans pour le moment définir les typologies de liaison dont le contenu ne fait pas toujours l'objet d'un consensus.

dans le passé (cf. Partie 1.Chapitre 5), une plus grande productivité. En effet, comme Grammont (1914) l'avait déjà souligné au début du XXème siècle, « les nouvelles générations tendent à ne garder que les liaisons qui jouent un rôle grammatical, particulièrement celles qui servent à distinguer le singulier du pluriel » (Grammont 1914, p. 130). Cette capacité à distinguer le singulier du pluriel se manifesterait selon Delattre (1947) par une absence de liaison au singulier. Cependant, Delattre (1947) précise que la réalisation de la CL en tant que marque du pluriel correspond dans certains contextes (ex. « grands amis ») à une tendance et non pas à une obligation.

En effet, la valeur morphologique et donc la variable syntaxique ne permettent pas d'opérer une séparation nette entre les liaisons qui sont toujours réalisées (par tout locuteur et dans tout contexte) et celles dont la réalisation varie en fonction de la combinaison de différents facteurs (phonologiques, syntaxiques, prosodiques, lexicaux et sociolinguistiques etc.). Il suffit en effet de considérer la liaison après un nom pluriel (ex. « des enfants intelligents ») pour se rendre compte que la correspondance entre liaison fonctionnelle et liaisons toujours réalisées n'est pas vérifiée[28].

La fonction morphologique de la liaison en [z] au pluriel est réduite à cause de la variabilité de la réalisation de la liaison dans ce contexte où l'influence de l'orthographe n'est pas anodine (cf. Partie 1.Chapitre 6.3).

> « Ce rôle structural qu'il *[le « s » correspondant à la CL du pluriel]* remplit dans la langue écrite semble toutefois très précaire dans la langue parlée. Déjà, le fait même que la liaison du nom pluriel est facultative nous renseigne sur la valeur diminuée de cette fonction. Delattre préfère donner à ce 's' une fonction morphologique rendant possible (mais pas indispensable) l'opposition du pluriel et du singulier ». (Ågren 1973, p. 124)

Ces observations invitent donc à prendre en compte des variables linguistiques autres que le facteur syntaxique. Celles-ci pourraient non seulement être utilisées en tant que descripteurs de la variabilité de la liaison mais aussi comme facteurs nécessaires pour isoler les cas où la liaison est toujours réalisée (liaisons catégoriques ou obligatoires selon les auteurs).

[28] Nous approfondirons cet aspect dans le Paragraphe 2.3.2.

2. Analyse des variables linguistiques

2.1. Variables phonologiques

Dans le chapitre précédent (cf. Partie 1.Chapitre 1), nous avons décrit la liaison comme un phénomène de *sandhi* externe soumis à des conditionnements à la fois morphosyntaxiques et phonologiques qui, selon les descriptions traditionnelles (Fouché 1959 ; Delattre 1966), déterminent la prononciation des consonnes finales de certains mots. Celles-ci apparaissent donc quand les deux mots (M1 et M2) sont étroitement liés syntaxiquement et lorsque le deuxième mot (M2) commence par une voyelle, d'où la condition phonologique qui influence la réalisation de la liaison.

2.1.1. Liaisons enchaînées ou non-enchaînées

Selon la description de Fouché (1959), la liaison à l'instar de l'enchaînement des consonnes fixes détermine dans la plupart des cas une resyllabation de la CL qui « se soude à la voyelle initiale du second [*mot*] pour former une syllabe avec elle » (Fouché 1959, p. 434). La CL est donc enchaînée à savoir produite comme attaque de la première syllabe du M2 (ex. « petit enfant » [pə.ti.tɑ̃.fɑ̃]). Comme nous l'avons déjà observé (cf. Partie 1.Chapitre 1.2.1), le phénomène de resyllabation est le résultat en français d'une dynamique ancienne de syllabation ouverte, très répandue dans tout le groupe des langues romanes favorisant la formation et la stabilisation d'une alternance de syllabes CV. La liaison se présente comme « un optimum d'équilibre entre la dynamique d'enchaînement généralisé et la dynamique de syllabation ouverte » (Laks 2005a, p. 162) [29]. Cependant, bien que, comme nous l'avons dit, la tendance à la syllabation ouverte soit prédominante, les CL ainsi que les consonnes fixes[30] peuvent ne pas être sujettes à resyllabation et se trouver en position de coda de syllabe [pə.tit.ɑ̃.fɑ̃]. Ce type de réalisations non-enchaînées, bien qu'elles soient marginales par rapport aux formes resyllabées, ont été attestées et ont été l'objet principal de l'étude d'Encrevé (1988) qui en propose une

[29] C'est pour cette raison que la liaison, comme nous le verrons dans le cinquième chapitre (cf. Partie 1.Chapitre 5), a été longtemps considérée comme un moyen d'éviter le *hiatus*, conformément au cadre esthétique du vers (Morin 2005). Cette fonction anti-hiatique du discours esthétique sur la poésie sera ensuite empruntée par des modèles phonologiques plus récents (Steriade 1999) et utilisée comme un « moyen mécanique d'assurer que les consonnes de liaison apparaissent devant un mot à initiale vocalique » (Morin 2005, p. 16). Cette position, comme nous le verrons dans le cinquième chapitre (cf. Partie 1.Chapitre 5) sera remise en discussion par plusieurs auteurs (Clédat 1917; Encrevé 1988; Morin 2005).

[30] Encrevé (1988) souligne que cet aspect est souvent négligé par les orthoépistes qui décrivent la resyllabation des consonnes fixes comme un fait invariable. Cependant, des cas de consonnes fixes non-enchaînées avaient été également attestés par Malécot (1975).

description dans le cadre d'une analyse multilinéaire[31]. L'analyse phonétique réalisée sur des données extraites des interventions radiophoniques et télévisées de vingt-et-un dirigeants politiques des années 1978-1981 confirme que « la présence de la consonne de liaison et sa position dans la structure syllabique sont deux phénomènes dissociables ». De plus l'auteur montre que les non-enchaînements des CL semblent être le plus souvent suivis par un coup de glotte qui précède la voyelle appartenant au M2.

> « Ce qui est caractéristique du non-enchaînement de la consonne de liaison, c'est qu'il y a 'quelque chose' entre C [consonne] et V [voyelle] qui concrétise la non-resyllabation de la CL. Mais ce 'quelque chose' peut varier […] peut s'agir comme dans l'exemple précédent [j'avais un rêve [ʒɑvɛzʔɛ̃ʁɛv]] que l'on perçoit nettement du fait de la brutalité de l'attaque vocalique. Mais une occlusion glottale très semblable […] peut être perçue davantage comme une légère pause si l'attaque dans laquelle se résout la tenue glottale est plus douce [que je jugeais essentiel [kəʒəʒyʒɛz(ʔ)esɑ̃sjɛl]) ». (Encrevé 1988, p. 32‑35)

La CL non enchaînée pourrait aussi, lorsqu'elle est occlusive, être réalisée comme aspirée (« qui sont en vérité les siens » [kisɔ̃tʰ(ʔ)ɑ̃veritelesjɛ̃]) ou suivie d'un *schwa* (« il faut en être à la fois » [ilfotʔəɑ̃nɛtʁa]). L'analyse de ces données amène Encrevé (1988) à considérer la nécessité d'intégrer le cas des liaisons non-enchaînées à l'analyse formelle de la liaison proposant un modèle multilinéaire (cf. Partie 1.Chapitre 5.3.2.4.1) qui sera par la suite l'objet de nombreuses critiques.

Ces formes de liaison non-enchaînées ont été déjà signalées par d'autres auteurs mais le plus souvent stigmatisées (Passy 1890 ; Cohen 1963 cités par Galazzi 2015b) ou associées à la présence d'un accent d'insistance (Coustenoble et Armstrong 1934 ; Delattre 1955)[32]. En particulier, Passy (1890 cité par Galazzi 2015b) qualifie de « risible » la réalisation de la liaison non-enchaînée[33]. En revanche, Coustenoble et Armstrong (1934) considèrent ces formes comme un des moyens (le moins fréquent) de réaliser un accent d'insistance (*intensive stress*) où « la consonne de liaison est prononcée, mais elle ne fonctionne pas comme une consonne de liaison, le locuteur insérant à sa place la plosive glottale qui commence la syllabe sur laquelle porte l'emphase » (Coustenoble et Armstrong 1934, p. 142-143). Delattre (1955) cite également

[31] Nous nous attacherons à présenter les détails de cette analyse dans le cinquième chapitre (cf. Partie 1.Chapitre 5).

[32] En revanche Encrevé (1988) signale que, d'après l'analyse phonétique menée sur les cas de non-enchaînement enregistrés dans son corpus, le coup de glotte ne s'accompagne d'aucun accent particulier sur la première syllabe du M2. « […] on ne peut donc aucunement parler ici d'accent d'insistance (ni de quelque accent que ce soit) » (Encrevé 1988, p. 32).

[33] « Rien de plus risible qu'une liaison faite mal à propos. *C'est une idée*, prononcé (sèt, unidé), fait croire qu'on a le hoquet. Un professeur prononçait des phrases comme : "*La première est excessivement facile*", en s'interrompant après est: *la première est, têxcessivement facile*. La première fois que nous l'entendîmes, ce fut un éclat de rire général ». (Passy 1890, p. 55 cité par Galazzi 2015b, p. 31)

ce phénomène parmi les facteurs prosodiques signalant la possibilité qu'un accent d'insistance affecte la réalisation de la liaison. Celui-ci pourrait selon l'auteur entraîner soit la non-réalisation de la CL (accompagnée d'un allongement vocalique) soit une discontinuité entre la CL et la voyelle suivante (sans pour autant qu'un coup de glotte se produise).

> « […] l'accent d'insistance peut affecter la liaison de diverses manières. Ainsi l'exclamation : 'c'est Impossible' se prononcera soit en omettant la liaison et en prolongeant la voyelle du verbe [sɛːɛ̃pɔsibl], soit de telle manière que le 't' se réunisse implosivement à la première syllabe : [sɛt ɛ̃pɔsibl] ». (Delattre 1955, p. 59)

Quant à l'étude d'Ågren (1973), bien qu'il n'enregistre qu'un seul exemple de coup de glotte, il signale qu'une partie des pauses-hésitations attestées dans son corpus suivent une CL donnant l'impression que celles-ci sont « incorporées au mot précédent »[34].

Morin et Kaye (1982) présentent également onze cas de CL (extraits des émissions radiophoniques et télévisées du Québec des années 1977-1980) suivis d'une « pause intonationnelle […] qui peut dans certains cas être renforcée par une occlusive glottale faible mais dépourvue de valeur emphatique comme dans les exemples suivants […] (c) 'qui était-t/ un grand ethnologue' (Jacques Soustelle, CBFT, 25.09.78) ». Cependant, les auteurs précisent que ce phénomène, bien qu'il soit « fréquemment observé dans les discussions formelles en France » est peu attesté au Québec. Les résultats de l'enquête de Lucci (1983) signalent également la présence dans le discours de certains hommes politiques français de réalisations de liaisons non-enchaînées du type (« c'était-t incroyable »).

D'après Encrevé (1988) ces formes, bien qu'elles soient connues par les travaux précédents, ont été peu étudiées et souvent traitées comme des erreurs de performance ou rapidement analysées comme des productions de liaisons affectées par des phénomènes accentuels.

Cependant, l'examen plus récent d'un grand corpus de français parlé (parole spontanée et lecture) (« Phonologie du Français Contemporain » cf. Partie 2.Chapitre 1) (Durand et al. 2009) a remis en discussion le statut phonologique de la liaison non-enchaînée (Mallet 2008; Durand et al. 2011) montrant un nombre très restreint de liaisons non-enchaînées (0,35% des liaisons totales attestées). Celui-ci est d'autant plus limité en parole spontanée où seulement une dizaine d'exemples indiscutables ont été enregistrés. Si les auteurs ne nient pas l'existence de ces formes de liaison, ils les considèrent comme des arguments insuffisants pour pouvoir justifier le développement d'une théorie phonologique. Ils suggèrent donc que les liaisons non-

[34] Les autres exemples des pauses-hésitations correspondent à des cas de CL précédées d'une pause (cf. Partie 1.Chapitre 2.2.5.4).

enchaînées relèvent de l'influence de la forme orthographique des mots laquelle pourrait s'avérer plus importante dans un style de parole publique, tel que celui étudié par Encrevé (1988).

> « Nous ne mettons pas en doute l'existence de la liaison non enchaînée. Il suffit d'assister à une conférence, de participer à une réunion publique ou d'écouter les médias pour en entendre des occurrences. La question qui reste posée par nos données extensives est celle de l'intégration de ce phénomène à l'analyse phonologique du français moderne. La liaison non enchaînée exige-telle une révision des modèles et de la théorie phonologique ? Au contraire, n'est-elle pas le signe tangible de la perméabilité des compétences et connaissances phonologique et orthographique ? L'effet Buben (en écho à Buben 1935) a été maintes fois documenté (cf. Chevrot & Malderez 1999 ; Laks 2005a). Il conduit à traiter le non-enchaînement comme marginal par rapport au système robuste de liaison ». (Durand et al. 2011, p. 34‑35)

Cette dernière hypothèse serait, d'après Encrevé, Scheer et Wauquier (2015), invalidée par la présence de 10 occurrences totales de liaisons non-enchaînées observées dans les productions de deux enfants pré-lecteurs (dont un enfant bilingue français-suédois et un enfant suédophone apprenant du français L2) (Splendido 2014, p. 246-247).

Pour résumer, l'ensemble de ces données montrent la présence d'une forme de réalisation de la liaison « sociologiquement qualitative » qui serait attestée principalement chez les hommes politiques des années 1970-1980[35] et plus en général chez les professionnels de la parole publique en particulier dans des contextes où la liaison est réalisée de façon variable[36]. Cependant, le statut phonologique de ces formes de réalisation de la liaison fait encore l'objet d'un débat parmi les spécialistes[37].

[35] À partir d'une analyse fine des données de parole des hommes politiques des années 1908 et 1996 (corpus HPOL1), Laks (2005c) a montré que, après une augmentation de liaisons non-enchaînées enregistrée dans les années 1970-1980 (correspondant assez précisément à la période d'observation d'Encrevé 1988), le phénomène régresse à des niveaux marginaux au cours des années suivantes (cf. Paragraphe 2.6.1.2). Ces fluctuations (liées principalement aux variations du phénomène socio-stylistique de la liaison non-enchaînée) confirmeraient selon l'auteur la primauté des variables diastratiques et diaphasiques par rapport aux facteurs diachroniques.

[36] Encrevé (1988), contrairement à Ågren (1973) n'a pas observé de réalisations non-enchaînées des CL dans des contextes de liaison catégorique. De plus, il souligne que dans un contexte de liaison variable, le caractère occlusif de la CL favoriserait le non-enchaînement de la CL.

[37] Étant donné que le statut phonologique des CL non-enchaînées est remis en question par certains auteurs (cf. Durand et al. 2011) et que l'attachement lexical ou phonologique des CL (cf. Partie 1.Chapitre 5.2) ne fait pas encore à ce jour l'objet d'un consensus, nous nous interrogeons sur la validité d'une distinction entre « liaisons non-enchaînées » (Encrevé 1988) et les cas où la CL est précédée d'une pause-hésitation (Ågren 1973). En effet cette opposition repose sur une approche qui considère la CL comme une consonne finale du M1 (même si flottante) (cf. Partie 1.Chapitre 5.2.1). En raison de ces observations, il nous paraît plus prudent de revenir à une description plus objective telle que celle qui avait été proposée par Ågren en 1973, qui distinguait les CL précédées ou suivies d'une « pause-hésitation » temporelle ou intonationnelle. Néanmoins nous continuerons à utiliser le terme « liaison non-enchaînée » pour faire référence aux seuls cas décrits par Encrevé (1988), à savoir des CL suivies d'une pause et non-resyllabées en tant qu'attaque de la voyelle suivante. En revanche nous décrirons les cas de réalisations d'une CL resyllabée à la voyelle suivante mais précédée d'une pause-hésitation (Ågren 1973) comme des occurrences de liaison où la CL est précédée d'une fracture temporelle ou intonationnelle.

2.2. Variables phonétiques

Contrairement aux défenseurs des modèles phonologiques qui voient dans l'évitement du *hiatus* le principe explicatif de la réalisation, Delattre (1955) présente les variables d'ordre phonétique comme des facteurs secondaires capables seulement d'expliquer « des différences qui ne sont attribuables ni à des différences de style, ni a des différences d'union ». D'après Delattre (1955) la liaison serait plus fréquente : (a) après voyelle qu'après consonne pour des raisons articulatoires (« il est plus aisé d'articuler une consonne que deux ») et aussi parce que « le besoin de liaison se fait moins sentir lorsqu'il y a déjà enchaînement […] puisqu'il n'y a pas de *hiatus* à éviter » (Delattre 1955, p. 46)[38] ; (b) « lorsqu'il s'agit d'éviter un *hiatus* entre deux voyelles de timbre différent » (ex. « Nous avons été » *versus* « Tu as été ») ; (c) après une consonne qu'après deux (ex. « Des noms amusants » *versus* « Des contes amusants »).

Les hypothèses (b) et (c) formulées par Delattre (1955) ont été testées par Ågren (1973) à partir d'une analyse d'occurrences de noms pluriels suivis d'un adjectif ou d'un participe passé, extraits de son corpus. Celles-ci ont été réparties en trois groupes : (1) les noms se terminant par une voyelle; (2) les noms se terminant par une consonne ; (3) les noms se terminant par deux consonnes. L'écart de réalisation de la liaison entre les trois groupes semblerait être très faible et indiquer, contrairement aux intuitions de Delattre (1955), que la liaison est produite aussi fréquemment après consonne qu'après voyelle et que pour des raisons articulatoires l'insertion d'un *schwa* serait possible en cas de liaison en /z/ après deux consonnes.

2.3. Variables syntaxiques

2.3.1. Degrés de cohésion syntaxique

Le facteur syntaxique est souvent décrit comme le plus déterminant dans la réalisation de la liaison. Comme nous l'avons vu dans le chapitre précédent (cf. Partie 1.Chapitre 1.2.3) la cohésion syntaxique entre les mots de liaison constitue une condition *sine qua non* sa réalisation ne pourrait pas avoir lieu. Comme le souligne Delattre (1966) l'union syntaxique étroite entre certains mots a permis de maintenir « l'enchaînement ancien » (Delattre 1966, p. 40). Ågren fait également observer que « sans cohésion entre deux mots, il n'y aura pas de liaison. Donc, plus

[38] En revanche, d'après Clédat (1890) : « ne fait-on guère la liaison que si le nom se termine par une consonne (ou par un e muet d'orthographe ordinaire) ; on dira facilement : *dé trou énorm'*, mais : *tigr'z énorm'* ». (Clédat 1890, p. 31)

l'union des mots est forte, plus les chances sont grandes pour que la liaison se fasse » (Ågren 1973, p. 15).

Toutefois comme le fait remarquer Côté (2013, p. 157) la notion de cohésion reste un concept élusif dans la littérature. Celle-ci a reçu plusieurs définitions: syntaxique (entre autres Damourette et Pichon 1968; Delattre 1955), prosodique (cf. Côté 2012), lexicale (Bybee 2005; Chevrot, Chabanal et Dugua 2007) et statistique (Côté 2013)[39] (cf. Partie 2.Chapitre 2.2.3.1).

Sur le plan syntaxique, les grammairiens Damourette et Pichon (1968) décrivaient la liaison comme un phénomène de « mue de contigüité » résultat du contact et des liens de nature différente entre les mots, d'une part « ligative » et de l'autre « séparative ».

> « La mue de contigüité est celle qui résulte du contact entre les mots. En effet, dans la phrase française, il n'y a pas d'arrêt forcé entre les mots. Le mode d'attaque toujours relativement doux des consonnes françaises permet aux consonnes qui terminent le mot précédent de s'appuyer, à moins d'une pause significative, sur la voyelle du mot suivant, les mots formant ainsi […] une chaîne ininterrompue dans laquelle leur sémantisme seul permet de les isoler [..] les mots sont donc sujets dans leur contact avec les mots subséquents, à deux ordres de muances, les unes ligatives, les autres séparatives ». (Damourette et Pichon 1968, vol. I, p. 193, 203)

Ils opèrent donc une différence entre deux types de mots (« mot franc » et « mot serf ») en français selon les diverses « muances » auxquelles ils peuvent être soumis. D'une part, ils considèrent les « mots francs » comme plus nombreux que les « mots serfs » et susceptibles de se trouver dans trois situations différentes : (a) « situation libre » dans laquelle le mot peut facultativement revêtir des « muances ligatives » ou « des muances séparatives » ; (b) la situation constructive dans laquelle il est obligé de prendre une « muance ligative » ; (c) la situation obstructive dans laquelle il est obligé de prendre une « muance séparative ». En revanche, les mots serfs ne peuvent se trouver que dans des situations constructives et obstructives. Cette distinction d'état amène les auteurs à traiter les cas de réalisation de la liaison comme des marques d'une situation constructive qui caractérise les liens entre les mots en liaison. Le principe de cohésion syntagmatique a été particulièrement mis en évidence par les études de Delattre (1947), qui ont fait de la variable syntaxique l'un des critères principaux de description de la réalisation de la liaison. Cette approche a donné lieu, dans la plupart des travaux à des analyses des contextes de la liaison basées sur une étude des catégories morphosyntaxiques d'appartenance des deux mots de liaison. Ceux-ci ont été décrits de façon détaillée pour la première fois par Delattre (1947) (cf. Tableau 1), constituant la base pour la

[39] Nous reviendrons sur les autres définitions de cohésion dans les pages qui suivent (cf. Partie 1.Chapitre 2.2.4.1).

définition des classifications que nous nous attacherons à étudier dans le quatrième chapitre (cf. Partie 1.Chapitre 4)[40].

	Contexte de liaison potentielle	**Exemple**
NOM	article défini - nom	*les enfants*
	article indéfini - nom	*un enfant*
	article contracté - nom	*aux enfants*
	article partitif - nom	*des enfants*
	adj. démonstratif - nom	*ces enfants*
	adj. possessif - nom	*son ami*
	adj. numéral - nom	*deux amis*
	adj. interrogatif - nom	*quelles histoires*
	adj. indéfini - nom	*plusieurs hommes*
	adj. qualificatif - nom	un *grand arbre*
	déterminatif - nom	*plusieurs autres*
	déterminatif - adjectif	*leurs éternels* regrets
	nom pluriel - adjectif	des *maisons immenses*
	nom pluriel - verbe	*les trains* ont du retard
	nom pluriel - invariable	des *plans à* faire
VERBE	pronom personnel - verbe	*vous êtes*
	pronom personnel - pronom personnel - verbe	*vous y* êtes
	verbe - pronom personnel	*vient-elle*
	verbe - pronom personnel- pronom personnel	allez-*vous-en*
	c'est (impersonnel) - M2	*c'est impossible*
	il est (impersonnel) – M2	*il est évident*
	pronom (pas personnel) - M2	*d'autres arriveront*
	pronom pers. postposé - M2	avons-*nous un* livre
	verbe - M2	ils *sont heureux*
	M1- participe passé	ils n'ont *rien acheté*
INVARIABLE	préposition monosyllabique - M2	*dans un* an
	préposition polysyllabique - M2	*depuis un* an
	adverbe monosyllabique - M2	*très utile*
	adverbe polysyllabique - M2	*souvent absent*
	conjonction monosyllabique – M2	*puis on* partira
	adjectifs pluriels- M2 invariable	*bons ou* mauvais

[40] Cependant, comme Mallet (2008) le fait observer, dans ce type de description « la caractérisation syntaxique demeure seconde et peut s'exprimer dans une économie minimale : la consécution de deux catégories » (Mallet 2008, p. 64) (cf. Partie 1.Chapitre 4.4).

	pronom - M2 invariable adverbe polysyllabique - M2 invariable	les *miens aussi* il a *toujours à* redire
	groupes figés	les *États-Unis*

Tableau 1 : Contextes morphosyntaxiques de liaison potentielle (extraits du tableau détaillé de Delattre 1947)

Delattre (1955) à la suite de Gougenheim (1935)[41] considère le degré d'étroitesse dans l'union des mots comme le « premier principe et seul nécessaire : la liaison se fait dans la mesure où l'usage a consacré l'extrême étroitesse d'union de deux mots ou classes de mots » (Delattre 1966, p. 40). En 1955 il établit un classement des contextes en fonction de leur degré de cohésion en les situant sur une échelle allant de 10 à 1, soit de la cohésion la plus forte à la plus faible :

10: Entre le déterminatif et le nom: « des enfants » ;

10: Entre l'adjectif et le nom: « de beaux enfants » ;

10: Entre le pronom personnel et le verbe: « ils ont » ;

10: Entre le verbe et le pronom personnel: « ont-ils » ;

9: Entre l'adverbe et le modifié: « tellement aimable » ;

8: Entre la préposition et son complément: « pendant un jour » ;

7: Entre l'auxiliaire et le participe passé: « vous avez aidé » ;

6: Entre l'auxiliaire et l'infinitif: « vous allez aider » ;

5: Entre le nom et l'adjectif: « des enfants intelligents » ;

4: Entre le verbe et son complément: « il désirait un cadeau » ;

3: Entre le pronom et le verbe: « les miens attendront » ;

2: Entre le nom et le verbe: « les enfants attendront » ;

1: Entre la conjonction et ce qui suit: « pourtant il était là » ;

1: Entre la conjonction et ce qui précède: « il sortait et ne rentrait plus ».

Delattre (1955) essaie ainsi de définir une hiérarchie du degré de cohésion syntaxique bien qu'il précise que les exemples donnés ont été choisis pour faire coïncider la fréquence de réalisation de la liaison avec le degré d'union. En effet Delattre (1955), parmi d'autres, souligne qu'un même contexte syntaxique peut présenter des fréquences de liaisons différentes (cf. Partie 1.Chapitre 4.4). Il en résulte que le degré de cohésion syntaxique des liaisons n'est pas seulement le résultat d'une combinaison de classes morphosyntaxiques. Pour cette raison Côté

[41] Gougenheim (1935) avait déjà essayé de caractériser le lien syntaxique entre des mots se trouvant dans un contexte où la liaison est toujours réalisée. Celui-ci correspondrait au rapport existant « entre les mots qui ont une fonction grammaticale (mots-outils) et les mots pleins avec lesquels ils font corps » (Gougenheim 1935, p. 41).

(2013) proposera de considérer la cohésion d'une séquence de liaison non seulement en termes syntaxiques mais également statistiques en s'appuyant sur la probabilité transitionnelle des séquences de liaison. L'idée de base est que la productivité d'une liaison entre un M1 et un M2 (*W1* et *W2* dans Côté 2013) (ex. « quand il ») est conditionnée par la probabilité transitionnelle qu'un M2 (ex. « il ») ou sa catégorie morphosyntaxique (ex. pronom personnel sujet) suive le M1 (ex. « quand »).

> « […] liaison remains productive with words that surface almost exclusively in the same construction and whose transition probability to a given category of words approaches 100% […] The productivity of liaison between W1 and (the category of) W2 is determined by the transition probability from W1 to (the category of) W2 ». (Côté 2013, p. 163)

Si la définition de la cohésion d'une séquence de liaison ne fait pas l'objet d'un consensus, toute la littérature reconnaît que la liaison est sensible à différents facteurs non seulement syntaxiques[42]. Parmi ceux-ci Delattre (1955) cite des facteurs stylistiques, prosodiques, phonétiques et historiques.

Cependant, d'autres auteurs, comme par exemple Anderson (1975), considèrent que la cohésion syntaxique qui caractérise les contextes où la liaison est toujours réalisée n'est pas sujette à l'influence d'autres variables linguistiques, ce qui confirmerait le rôle fonctionnel de ce type de liaison.

> « [...] I maintain that the relevant syntactic relationships go beyond mere lexical characteristics in determining the occurrence of obligatory liaison. Of primary importance here is the degree and consistency with which a special class of sequences including modifier-noun, clitic pronoun-verb and verb-clitic pronoun evidence liaison ». (Anderson 1975, p. 849)

Anderson (1975) fait ici référence aux contextes de liaison auxquels Delattre (1955) avait assigné un indice de cohésion 10, à savoir respectivement aux liaisons : (a) entre les déterminants et le nom et entre l'adjectif et le nom qu'elle considère appartenant à la catégorie des « modifieurs - nom » ; (b) entre le pronom clitique et le verbe et entre le verbe et le pronom clitique inclus par Anderson (1975) dans les classes « pronom clitique - verbe » et « verbe - pronom clitique ». Anderson (1975) essaie d'établir un lien entre ces contextes ayant une cohésion syntaxique maximale et pour lesquels la liaison semblerait être toujours réalisée indépendamment des variables autres que celles syntaxiques. S'appuyant sur des tests syntaxiques, Anderson (1975) considère que les classes « pronom clitique - verbe », « verbe -

[42] Nous reviendrons sur cet aspect dans les pages qui suivent, lors de l'étude des autres variables d'analyse de la liaison.

pronom clitique » et la classe « modifieurs - nom » partagent des propriétés syntaxiques[43] communes telles que l'impossibilité d'introduire des interruptions syntaxiques (comme par exemple des expressions parenthétiques *« ce, me semble-t-il, professeur s'intéresse trop aux détails »). Ces comportements (qui reflètent le fort degré de cohésion syntaxique et prosodique existant entre ces éléments) permettraient selon Anderson (1975), de distinguer les contextes ci-dessus, pour lesquels la liaison est toujours réalisée, d'autres contextes où la production de la liaison est soumise à l'influence de facteurs autres que syntaxiques. Autrement dit, Anderson (1975), bien qu'il ne remette pas en question l'ensemble de la hiérarchie établie par Delattre (1955), limite la variation de la liaison aux contextes ayant un degré d'étroitesse inférieur à 10.

2.3.2. Liaisons antéposées et postposées à une tête syntaxique

Le degré de cohésion syntaxique qui assure une production invariable de la liaison semblerait donc être associé à un rapport de dépendance syntaxique ainsi que prosodique du M1 de liaison au M2, qui constitue la tête syntaxique. Déjà en 1947, Delattre (1947) avait décrit cette relation en signalant que la liaison était souvent réalisée « en passant d'un mot moins important à un mot plus important » comme dans les exemples suivants : « nous arrivons » ; « les amis » ; « deux élèves » ; « en entrant ». Il signale une seule exception à cette tendance à savoir la liaison entre le verbe et le pronom postposé (ex. « est-il » ; « a-t-il? » ; « sont-ils? » ; «entrent-ils? » ; « allez-y ») (Delattre 1947, p. 149). Il en va de même pour Grammont (1914), d'après lequel les mots « accessoires » se lient toujours quand ils sont préposés mais pas lorsqu'ils sont postposés.

Ces observations semblent suggérer une possible correspondance entre la position du M1 par rapport à une catégorie principale et le fait que la liaison soit toujours réalisée. Cependant, comme nous le verrons dans le quatrième chapitre (cf. Partie 1.Chapitre 4.3.2.3.1 ; Partie 1.Chapitre 4.3.2.3.2), il existe des contextes (tels que la liaison après les adverbes, prépositions et adjectifs) où la liaison est variable bien que les M1 dépendent syntaxiquement du mot suivant. La position de la liaison par rapport à une catégorie principale ne constitue donc pas un critère permettant d'isoler les cas où la liaison est toujours réalisée. Cependant, Laks (2005b) suggère d'évaluer le poids de ce paramètre en prenant également en compte l'opposition (dont nous avons retracé les origines historiques dans le premier chapitre cf. Partie 1.Chapitre 5) entre d'une part les liaisons fonctionnelles et d'autre part les liaisons phonologiques (cf. Partie

[43] Cependant, les épreuves syntaxiques ne sont pas vérifiées pour le contexte « adjectif-nom » (qu'Anderson avait initialement décrit comme appartenant à la classe « modifieur - nom ») qui sera décrit par les études suivantes comme un site où la liaison est réalisée de façon variable (cf. Partie 1.Chapitre 4.3.2.3.2).

1.Chapitre 2.1). Il fait donc la distinction entre liaisons préposées à une catégorie principale ayant une valeur fonctionnelle[44] (que l'auteur définit comme « liaison de nombre » et « liaison de personne ») et postposées à une catégorie principale. Les premières constitueraient des marques morphologiques et seraient en tant que telles toujours réalisées. En revanche, les liaisons postposées à une catégorie principale, s'appuyant sur des consonnes lexicales, seraient tributaires de la représentation orthographique des mots que nous nous attacherons à étudier dans le sixième chapitre (cf. Partie 1.Chapitre 6.3)[45].

Pour résumer, bien que l'antéposition à une catégorie principale ne constitue pas un critère descriptif suffisant, la prise en compte à la fois du rapport de dépendance syntaxique au mot suivant et de la valeur fonctionnelle de la liaison permet d'identifier des contextes de liaison traités par tous les auteurs (Fouché 1959; Delattre 1947; Encrevé 1988; Durand et Lyche 2008; Durand et al. 2011) comme des cas où la liaison est toujours réalisée (à savoir les contextes déterminant - nom ; pronom clitique - verbe ; verbe - clitique que nous avons évoqués dans le Tableau 1 (cf. Paragraphe 2.3).

Toutefois, parmi les séquences où la liaison est toujours réalisée il existe également des contextes où les M1 ne se trouvent pas dans une position antéposée à une catégorie principale et où la CL ne possède aucune valeur fonctionnelle. Il s'agit des expressions figées (ex. « de temps en temps ») dont l'analyse implique la prise en compte des variables lexicales.

2.4. Variables lexicales

La définition syntaxique des contextes de liaison basée sur une analyse linéaire des catégories morphosyntaxiques des mots de liaison paraît insuffisante non seulement pour caractériser la variation de la liaison mais également pour identifier les cas où la liaison est toujours réalisée. Certains auteurs se sont donc attachés à étudier les variables lexicales pouvant conditionner la production de la liaison à savoir la longueur du M1 et M2, leur fréquence d'occurrence ainsi que la fréquence de cooccurrence de leur combinaison.

[44] Cela revient à dire qu'il existe certains contextes de liaison antéposés à une catégorie principale qui ne possèdent pas de valeur fonctionnelle (dont la CL renvoie à une consonne lexicale) dans lesquels la liaison est variable. C'est le cas par exemple de la liaison entre un adjectif et un nom au singulier. Bien que celui-ci corresponde à une liaison antéposée à une catégorie principale, il n'a pas de rôle morphologique.

[45] Ces observations permettent d'expliquer la variabilité de la réalisation de la liaison après des noms pluriels (représentant la tête syntaxique) que nous avons déjà mise en évidence dans les pages précédentes (cf. Paragraphe 1).

2.4.1. Cohésion syntagmatique comme fréquence de cooccurrence des mots de liaison

Comme nous l'avons souligné dans le paragraphe précédent (cf. Paragraphe 2.4), le degré d'étroitesse n'est pas seulement établi entre des mots ayant une fonction grammaticale et les mots pleins dont ils dépendent, mais peut également être lié à des phénomènes de nature sémantique et lexicale.

À la suite de Grammont (1933), Laks (2005b) fait observer que le degré d'union syntagmatique des mots de liaison est aussi bien sémantique et il cite deux exemples de contextes de liaison ayant des comportements différents[46].

> « Pour en prendre deux exemples, ce n'est certainement pas sur la base d'une analyse phonologique ou même morphologique que l'on peut expliquer la différence aspectuelle entre *vas-y voir* (i.e. tente !) *et va / y voir* (i.e. va voir là), *allez-y comprendre quelque chose* où la liaison est assez fréquente et plus facile et *allez / y prendre quelque chose* où elle l'est moins et stylistiquement très marquée ». (Laks 2005b, p. 110)

L'auteur explique que ces exemples nécessitent une analyse sémantico-pragmatique pour rendre compte de la non-réalisation de la liaison devant la particule « y » ayant une valeur locative. De façon plus générale Âgren (1973) avait déjà fait observer que la liaison tend à se conserver dans les contextes qui apparaissent le plus souvent.

C'est dans ce cadre que s'inscrivent les études de Bybee (2005) et Chevrot et al. (2007) qui proposent de considérer la cohésion syntaxique comme « le résultat direct de la fréquence de cooccurrence : les mots qui sont souvent employés ensemble semblent liés par une plus forte cohésion » (Bybee 2005, p. 24). Cette conception renvoie aux théories lexicalistes du Modèle Basé sur l'Usage (MBU)[47] selon lequel toute réalisation de liaison est stockée au sein de constructions constituées par des séquences M1-M2. En particulier Chevrot et al. (2007) proposeront de considérer la cohésion d'une séquence de liaison comme le résultat du rapport entre le nombre d'occurrences du M1 dans un contexte donné et le nombre de différents M2 susceptibles de se trouver après M1 à l'intérieur d'une séquence de liaison.

Cette analyse de la cohésion syntagmatique comme conséquence de la fréquence n'est pas nouvelle. Âgren (1973) avait déjà observé le lien étroit existant entre la cohésion et la fréquence de cooccurrence selon lequel « une forte cohésion suppose une fréquence d'emploi élevée » (Âgren 1973, p. 28). La prise en compte de la fréquence de cooccurrence comme

[46] Delattre (1996) cite également les deux exemples ; « Coupez-*vous en* deux » *versus* « Coupez-*vous // en* deux ».

[47] Nous nous attacherons à décrire dans le détail ce modèle dans le cinquième chapitre (cf. Partie 1.Chapitre 5.3.4).

facteur pouvant influencer le degré de cohésion syntagmatique pourrait permettre d'expliquer la présence de liaisons postposées à une tête syntaxique dont la CL n'a pas de rôle morphologique.

> « Synchroniquement, la liaison postposée à une catégorie principale est impossible si elle s'appuie sur une consonne lexicale, donc non fonctionnelle: **le temps͡ était compté*, **l'univers͡ immense*. Pourtant la construction lexicalisée il travaille *nuit͡ et jour* fait apparaître une liaison catégorique postposée s'appuyant sur une consonne lexicale (/t/). Il faut en rechercher la source dans la diachronie qui a figé l'expression alors que dans la forme discursivement construite la liaison est bien catégoriquement interdite ». (Laks 2005a)

Dans ce sens l'analyse de la fréquence de cooccurrence qui caractérise les constructions figées permet d'identifier un contexte où la liaison n'ayant aucun rôle fonctionnel est pourtant traditionnellement considérée comme toujours produite.

Nous pouvons donc observer qu'une analyse conjointe des facteurs syntaxiques et lexicaux permet d'isoler les contextes de liaison que toute la littérature reconnaît comme toujours réalisés, notamment : (a) déterminant - nom ; (b) pronom clitique - verbe ; (c) verbe - clitique (d) les expressions figées.

Quant aux autres contextes où la liaison est réalisée de façon variable, les effets de la fréquence de cooccurrence seront mis en avant par les travaux de Fougeron et al. (2001). Ces auteurs, à partir d'un corpus de cinq heures de parole, ont mené une étude statistique globale de facteurs de nature différente (stylistique, syntaxique et lexicale) pouvant jouer un rôle dans la réalisation de la liaison. Les résultats de cette recherche ont montré l'existence d'une incidence statistiquement significative de la fréquence de cooccurrence du M1 et du M2 sur la réalisation de la liaison.

> « La fréquence de la séquence seule est donc importante, indépendamment de la fréquence du lieur. Les liaisons apparaissent donc plus fréquemment lorsque les lieurs et les liés sont des mots fréquents et quand ils forment une séquence de mots fréquents dans la langue ». (Fougeron et al. 2001, p. 9)

Cependant ces auteurs soulignent que ces données doivent être interprétées en tenant compte du fait que d'après l'étude statistique menée, les caractéristiques de fréquence sont fortement dépendantes de celles de longueur des mots de liaison (cf. Paragraphe 2.4.3).

2.4.2. Fréquence des mots de liaison

Ce lien étroit entre fréquence et longueur des mots avait déjà été souligné par Ågren (1973), lequel, comme nous l'avons observé dans le paragraphe précédent (cf. Paragraphe 2.4.1), avait considéré la fréquence comme un facteur pouvant influencer le degré de cohésion syntagmatique de la séquence de liaison. La fréquence est donc considérée par Ågren (1973),

et avant lui par Langlard (1928), comme un paramètre déterminant dans la réalisation de la liaison dont la survivance dans la prononciation moderne serait aussi tributaire. Cependant, il renonce à examiner objectivement l'influence de ce facteur qu'il estime non dissociable de la cohésion de la séquence de la liaison ainsi que de la longueur des mots de liaison.

> « Il est cependant hors de ma portée de pousser plus loin l'examen de ce facteur et il faut constater que la fréquence d'emploi ne pourra pas être isolée des autres facteurs. Elle doit être en rapport étroit d'une part avec la cohésion (une forte cohésion suppose une fréquence d'emploi élevée), d'autre part avec la longueur des mots liants ». (Ågren 1973, p. 28)

Des études récentes menées sur des corpus de parole semi-spontanée ont obtenu des résultats qui semblent suggérer une possible influence de la fréquence des M1 et M2 dans la réalisation de la liaison. En particulier, l'étude menée par Adda et al. (1999) sur deux corpus (l'un de parole semi-spontanée MASK, l'autre de lecture oralisée BREF) a montré que le taux de réalisation de la liaison varie sensiblement en fonction de la fréquence des M1 et des M2. Ces auteurs ont montré qu'en réduisant au fur et à mesure l'échantillon du corpus aux mots les plus fréquents, le pourcentage de production de la liaison augmente de façon significative. Autrement dit, un nombre limité des mots les plus fréquents permet de rendre compte de plus de la moitié des réalisations de la liaison ce qui rejoint la tendance observée sous le nom de loi de Zipf (1949)[48].

> « For MASK (spontaneous task-oriented speech), the 10 most frequent words account for 30% of the corpus, whereas for read newspaper speech they cover about 20% of the data. The 100 most frequent words cover 80% of the MASK data, but slightly less than 50% of BREF. The coverage of the BREF data is close to linear on the logarithmic scale (Zipf's law) ». (Adda-Decker et al. 1999, p. 3)

Les résultats des études menées par Fougeron et al. (2001) semblent également indiquer une corrélation positive significative ($p < 0,01$) entre le taux de réalisation de liaison et la fréquence du M1 et du M2. Toutefois, comme nous l'avons souligné dans le paragraphe précédent (cf. Paragraphe 2.4.1), les résultats de cette étude montrent que les caractéristiques de longueur et de fréquence sont difficilement dissociables les unes des autres. Pour cette raison Fougeron et al. (2001) invitent à la prudence quant aux conclusions qui peuvent être tirées. L'influence de la fréquence des mots de liaison semble en effet être étroitement corrélée à la longueur du M1[49] en accord avec la loi de Zipf (1949).

[48] Nous reviendrons sur cet aspect dans la deuxième partie de la présente étude (cf. Partie 2.Chapitre 2.1.2).

[49] Dans le sens contraire De Jong (1988) affirme que :

« […] ce n'est pas le facteur longueur de mot (monosyllabes *versus* polysyllabes) qui joue un rôle. Les formes polysyllabiques sont les formes les moins fréquentes. Le facteur longueur de mot en nombre de syllabes ne fait que refléter le facteur de fréquence de mot ». (De Jong 1988, p. 115)

> « […] comme le prédit la loi de Zipf, les mots courts qu'ils soient lieurs ou liés sont aussi les plus fréquents. Ainsi pour ce qui concerne les mots liés, l'influence de la fréquence est similaire à celle observée pour la longueur : les lieurs les plus fréquents, les monosyllabiques, ont un taux de réalisation de liaison plus important. Par contre, si la fréquence des mots liés diminue avec leur allongement, il apparaît que les mots moins fréquents et longs (3-4 syll) ne sont pas moins liés que les autres (avec par ex. plus de 70% de liaisons pour les noms) ». (Fougeron et al. 2001, p. 9)

2.4.3. Longueur des mots de liaison

Delattre (1947) avait déjà mis en évidence que la liaison est produite « d'autant plus que le premier des deux mots est plus court » (Delattre 1947, p. 24). En particulier il avait comparé de façon qualitative la différence de réalisations observée entre les liaisons produites après des prépositions monosyllabiques et polysyllabiques (« en un jour » *versus* « depuis un jour ») et après des adverbes monosyllabiques et polysyllabiques (« très utile » *versus* « extrêmement utile »). L'étude d'Encrevé (1988), menée auprès de vingt et un hommes politiques français, confirmera cette intuition. Tout contexte confondu, les liaisons après des éléments polysyllabiques ne sont réalisées qu'à hauteur de 29% des cas alors qu'après un monosyllabique le taux de production atteint environ 77% des cas. Les résultats de l'étude de Mallet (2008) menée sur le corpus « Phonologie du Français Contemporain » (Durand et al. 2009) (cf. Partie 2.Chapitre 1.1) vont dans la même direction et indiquent que les monosyllabes lient dans 66,3% des contextes et les polysyllabes dans 7,4% des contextes. Il en va de même pour des travaux (De Jong 1988; Durand et Lyche 2008; Mallet 2008) ayant mené des analyses des formes fléchies du verbe « être ». Les formes verbales monosyllabiques ont tendance à être réalisées plus fréquemment avec liaison que celles polysyllabiques. Les résultats de l'étude statistique menée par Fougeron et al. (2001) précédemment évoquée, confirment également que la fréquence de réalisation de la liaison est plus importante lorsque le M1 est court. Cependant, les auteurs soulignent que cette tendance intéresse particulièrement les conjonctions, les déterminants, les prépositions, les pronoms et les verbes après lesquels les liaisons sont réalisées à plus de 60%.

Durand et al. (2011) partagent en partie cette conclusion dans la mesure où les mots grammaticaux qui présentent un taux plus important de liaison sont aussi monosyllabiques. Cependant, les auteurs en comparant la fréquence de réalisation de la liaison après les formes verbales « sont » et « était », montrent que bien que la première (« sont ») soit sensiblement moins représentée dans le corpus par rapport à la deuxième (« était »), elle opère la liaison deux fois plus fréquemment. Ces observations démontreraient d'après Durand et al (2011) le rôle joué par le nombre de syllabes dans la production de la liaison.

Quant à l'influence de la longueur du M2 sur la réalisation de la liaison, très peu de travaux ont été consacrés à l'étude de cet aspect. Fougeron et al. (2001) ont observé une tendance inverse par rapport à celle enregistrée pour les M1 bien qu'ils précisent que celle-ci est très réduite sauf dans les cas des adjectifs.

> « […] les mots monosyllabiques sont moins souvent liés que les 2 à 4 syllabes. Toutefois, il apparaît que cet effet est aussi fonction de la catégorie syntaxique : les adjectifs, noms, pronom et verbes monosyllabiques de notre corpus sont liés à plus de 50%. De plus, l'effet de longueur du mot lié apparaît négligeable lorsque l'on compare les proportions de réalisation de liaison pour une même catégorie syntaxique aux différentes longueurs. Par exemple, les noms sont liés à environ 80% indépendamment de leur longueur. Seuls les adjectifs montrent une diminution de leur pourcentage de liaison avec une augmentation de leur longueur ». (Fougeron et al. 2001, p. 8)

De plus, Fourgeron et al. (2001) notent que les variables lexicales de longueur des mots de liaison ne sont pas sans rapport avec l'organisation prosodique de l'énoncé en groupes accentuels.

2.5. Variables prosodiques

Comme nous l'avons vu dans les paragraphes précédents, les facteurs syntaxiques, lexicaux et prosodiques semblent être interdépendants les uns des autres. Les facteurs lexicaux de fréquence et longueur des mots de liaison sont strictement corrélés à des variables prosodiques pourtant souvent moins abordées par la littérature. Néanmoins, certains travaux ont été consacrés à l'étude de l'influence des facteurs prosodiques sur la réalisation de la liaison tels que la longueur du groupe syntaxique auquel les M1 et M2 appartiennent, l'intonation de la phrase, le débit de parole ainsi que la présence d'un accent d'insistance et d'une pause.

2.5.1. La longueur du groupe syntaxique auquel appartiennent le M1 et le M2

La longueur du groupe syntaxique semble être strictement corrélée aux contraintes rythmiques qui gouvernent la longueur du groupe accentuel. Delattre (1955) sera le premier à souligner l'incidence de ce facteur dans la réalisation de la liaison en précisant que plus long est le groupe syntaxique du M1 et/ou du M2, moins la liaison aura de possibilités d'être réalisée. Il fournit plusieurs exemples pour argumenter son hypothèse.

> « Dans *Les enfants attendront*, la possibilité de lier est plus forte que dans *Les plus petits des enfants attendront*. Le même résultat s'observerait si au lieu d'allonger le groupe du sujet on allongeait le groupe du verbe. On liera plus fréquemment dans *Les enfants attendront* que dans *Les enfants attendront longtemps leurs parents* ». (Delattre 1955, p. 46)

L'étude de cette variable a été ensuite reprise par des travaux successifs. Ågren (1973) avait accordé une place importante aux conditionnements rythmiques. La longueur du groupe syntaxique du M2 semble également jouer un rôle dans les résultats de l'étude de Morin et Kaye (1982). Ces auteurs ont enregistré des taux de liaison plus élevés avec des compléments courts.

> « It [*liaison*] is also conditioned by length factors [...] liaison seems definitely more difficult in sentence *Ils mangent après qu'ils aient fini leur travail* than in sentence *Ils travaillent d'abord et mangent après* ». (Morin et Kaye 1982, p. 307‑308)

Plus intéressantes sont les données de l'étude statistique menée par Fougeron et al. (2001) qui ont analysé l'influence de la longueur de la séquence de liaison sur la réalisation de la liaison. Les auteurs observent que dans des séquences courtes de deux syllabes les liaisons sont réalisées plus fréquemment que dans les séquences plus longues. En effet, il semblerait qu'au-delà du seuil de cinq syllabes, la fréquence de liaison tende à se réduire.

2.5.2. L'intonation de la phrase

En ce qui concerne l'effet de l'intonation de la phrase sur la production de la liaison, Delattre (1955) se limite à signaler les effets qu'il estime « marqués » de l'intonation interrogative sur la diminution de la production de la liaison. Il souligne la différence entre des réalisations de liaisons fréquentes dans un énoncé tel que « quand on arrivera » et l'absence de liaisons dans une production similaire ayant une intonation interrogative « quand arrivera-t-il?». Il en va de même pour la liaison dans l'énonciation : « les enfants attendront longtemps » qu'il juge très fréquente par rapport à celle qui peut être produite dans la question « les enfants attendront-ils longtemps ? » (Delattre 1955, p. 46).

En revanche, le travail de Lucci (1983), mené sur un corpus de parole spontanée et semi-spontanée de vingt locuteurs, a analysé les effets de la variation intonative de l'énoncé (correspondant à la perception d'un « déplacement vers le haut sur la dernière syllabe du premier mot [*de liaison à savoir M1*] ») sur la réalisation de la liaison (Lucci 1983, p. 241-242)[50].

Les résultats de cette étude sembleraient confirmer les intuitions de Delattre (1955) indiquant que la réalisation de la liaison s'accompagne le plus souvent d'une absence de variabilité prosodique. Cela amène Lucci (1983) à affirmer que la liaison « fonctionne comme une marque de non rupture ou de continuité prosodique » (Lucci 1983, p. 247).

[50] L'auteur précise que cette variation intonative représente un paramètre prosodique de l'accent du groupe rythmique qui peut se réaliser aussi bien par un allongement syllabique que par une augmentation de l'intensité.

2.5.3. L'accent d'insistance

La réalisation de la liaison serait en outre sensible, selon Delattre (1955) à la présence d'un accent d'insistance portant sur la première syllabe du M2. Celui-ci pourrait, dans un énoncé exclamatif tel que « c'est impossible ! », affecter la réalisation de la liaison « soit en omettant la liaison et en prolongeant la voyelle du verbe [seɛ̃ːposibl], soit de telle manière que le [t] se réunisse implosivement à la première syllabe: [set ɛ̃posibl] »[51] (Delattre 1955, p. 46). Lucci (1983) fait également référence à ce type de phénomène qu'il qualifie de systématique dans son étude et qu'il définit comme « accent didactique » (Lucci 1983, p. 242). Il signale que dans ce cas la voyelle est généralement précédée d'un coup de glotte après lequel la liaison est le plus souvent non réalisée. Nous avons vu comment la présence de cet accent d'insistance a été par la suite remise en question par l'étude d'Encrevé (1988) (cf. Paragraphe 2.1.1).

2.5.4. Pauses

Comme nous l'avons vu dans les pages précédentes, la liaison constitue « une marque de continuité prosodique » normalement considérée comme un phénomène se produisant à l'intérieur d'un groupe rythmique en dehors duquel la liaison ne peut pas être réalisée (Grammont 1933; Fouché 1959; Léon 2007). Cependant, certains auteurs (Vinay 1970; Ågren 1973; Lucci 1983; Encrevé 1988) ont noté que la présence d'une pause entre les mots de liaison n'entraîne pas forcement l'absence de production de la liaison.

L'étude d'Ågren (1973), notamment, menée sur un corpus de parole semi-spontanée d'enregistrements radiophoniques, montre la présence de liaisons réalisées avant et après des pauses-hésitations à hauteur de 5%. En particulier les liaisons entre noms et adjectifs (ex. « enfants intelligents ») présentent un taux important (25%) de ce type de réalisation. Des liaisons précédées d'une pause hésitation ou d'une pause d'intonation sont également attestées par l'étude de Morin et Kaye (1982) conduite sur un corpus de parole semi-spontanée au Québec. Ces auteurs, contrairement à l'hypothèse ensuite défendue par Encrevé (1988) (cf. Paragraphe 2.1.1), traitent ces pauses-hésitations comme le résultat d'un déficit de performance orale n'ayant aucune valeur stylistique.

2.5.5. Débit de parole

Un nombre très réduit de travaux (Adda-Decker et al. 2012; Fougeron et al. 2001) ont été consacrés à l'étude de l'influence du débit de parole, facteur pourtant fortement corrélé aux

[51] Ce contexte, comme nous l'avons déjà observé (cf. Paragraphe Partie 1.Chapitre 2.2.1.1), avait été présenté par Encrevé (1988) comme un cas possible de liaison non enchaînée.

variables stylistiques souvent analysées dans la littérature. En particulier le travail d'Adda-Decker et al. (2012) a analysé les effets du débit de parole sur la réalisation de la liaison à partir d'un corpus de parole spontanée familière (NCCFr à savoir *Nijmegen Corpus of Casual French*) caractérisé par des réductions temporelles fréquentes (Adda-Decker et Snoeren 2011). Ces auteurs émettent l'hypothèse que dans ce type de parole la réalisation de liaison pourrait être affectée par un débit de parole[52] rapide conduisant à une réduction du nombre de syllabes lorsque la liaison n'est pas réalisée. Autrement dit, en cas de non-réalisation de la liaison le contact entre les deux voyelles appartenant à deux syllabes consécutives (V#V) pourrait donner lieu à une fusion de celles-ci dans un seul segment vocalique ou semi-vocalique (appartenant à une seule syllabe). Les résultats d'une analyse de régression menée dans cette étude montrent la présence d'une corrélation négative entre le débit et le taux de réalisation des liaisons sauf pour un sous-ensemble de contextes de liaison incluant les mots « chez », « dans » et « est » pour lesquels une forte variation inter-individuelle a été enregistrée.

En revanche, les résultats de l'étude de Fougeron et al. (2001) ont montré[53] que le débit de parole n'a d'effet sur la réalisation des liaisons que pour un locuteur sur dix. Les auteurs en concluent que, contrairement à d'autres phénomènes comme l'élision, la liaison n'est pas affectée par les *fast speech rules*. En dépit de leurs résultats contradictoires, les deux travaux cités ci-dessus ont le mérite d'avoir souligné l'importance du débit de parole comme facteur non seulement prosodique mais aussi comme élément faisant partie des variables stylistiques d'analyse de la liaison.

2.6. Variables sociolinguistiques

L'analyse des variables sociolinguistiques qui affectent la réalisation de la liaison a été l'objet principal des analyses variationnistes, contrairement aux travaux de la phonologie générative et de syntaxe formelle, que nous nous attacherons à décrire dans le prochain chapitre (cf. Partie 1.Chapitre 3.2.2). Les usages variables de la liaison sont multiples et ils intéressent l'ensemble des dimensions de la variation inter-locuteurs et intra-locuteurs du français (Coseriu 1969; Gadet 2007).

Parmi les axes de variation inter-locuteurs nous pouvons citer la variation diachronique (selon le temps), diatopique (selon l'espace géographique) et diastratique (selon l'âge, le sexe et le niveau socio-culturel des locuteurs). Les facteurs de la variation intra-locuteurs seront

[52] Le débit de parole a été calculé au niveau du mot et ne tient donc pas compte des silences.

[53] À partir d'une comparaison des productions de liaison de lectures oralisées (ayant un débit de parole normal et rapide).

également pris en compte. Parmi ceux-ci nous comptons la variation diaphasique (selon la situation stylistique et fonctionnelle). Ces dimensions sont bien évidemment le résultat d'un traitement classificatoire visant, pour des raisons descriptives, à regrouper un ensemble de traits de variation qui ne sont pas discontinus mais plutôt interdépendants.

Cependant, il convient de rappeler que cette variation n'intéresse pas les liaisons qui sont « invariablement réalisées par tous les locuteurs » (Encrevé 1988, p. 48) telles que nous les avons décrites dans les pages précédentes. En effet celles-ci ne constituent pas un « lieu de variation linguistique, pas plus social que stylistique » ne créant aucune distinction sociale ni positive ni négative (Encrevé 1988, p. 49).

En revanche tous les auteurs s'accordent à considérer la réalisation des liaisons sujettes à variation comme un indicateur sociolinguistique explicite. La répartition sociale et communicative de ce phénomène oblige, d'après Encrevé (1988), à rechercher et étudier la liaison auprès des « locuteurs socialement dominants » et « dans des conditions de langue *surveillée* et non de conversation familière » (Encrevé 1988, p. 49-50).

2.6.1. Variation inter-locuteur

2.6.1.1 Variation diastratique

La liaison représente donc un important indicateur sociolinguistique dont la réalisation ou la non-réalisation détermine une distinction sociale positive ou négative.

> « La liaison est un indicateur social explicite, un des rares lieux de la langue où les plus antivariationnistes des linguistes ont été amenés à reconnaître la variation sociale et l'hétérogénéité linguistique ». (Encrevé 1988, p. 30)

Dans ce sens la liaison a été définie par Encrevé comme « un phénomène sociolinguistique inversé » puisque les locuteurs les plus scolarisés présentent le plus large système de variation. Il en découle que les liaisons les plus intéressantes d'un point de vue sociolinguistique sont aussi les plus rares (Gadet 1989).

Cependant, comme le souligne Mallet (2008), les contours des profils socioculturels des locuteurs sont souvent flous, s'appuyant sur des dénominations (ex. « bourgeoisie », « société cultivée », « classe populaire », « classe favorisée / défavorisée », « hommes politiques ») qui relèvent de cadres de niveaux d'analyses différents (classe sociale / professionnelle; classe cultivée / populaire etc.). Cet état de description serait à attribuer à une difficulté propre de la sociolinguistique de définir les critères d'identification socio-démographique des locuteurs. Ceux qui sont le plus souvent retenus correspondent en effet à des indices sociaux quantifiables tels que le sexe, l'âge et les indices de position sociale proprement dits (niveau d'étude,

profession etc.) (Gadet 2007, p. 91). Ces difficultés associées à la définition d'un profil socioculturel font place à une réflexion sur la possibilité effective de considérer la liaison comme un indicateur pouvant permettre « de classer socialement un locuteur ».

Quant à la variable sociodémographique du sexe, les résultats semblent être très contradictoires et suggèrent que ce facteur n'a pas d'effet significatif sur la fréquence de réalisation de la liaison. En effet, si d'une part les études de Malécot (1975) et Booij et De Jong (1987) signalent un taux plus important de production de liaison chez les femmes, d'autre part l'étude de Ashby (1981) présente des résultats inverses et les travaux de Mallet (2008) et Durand et al. (2011) ne signalent aucune différence statistiquement significative entre les taux de liaisons réalisées entre hommes et femmes. Cependant l'étude de Laks (2009) menée auprès des hommes et des femmes politiques entre les années 1999 et 2011 montre que « les femmes sont en avance sur les hommes dans le nivellement de la liaison par le bas ».

En ce qui concerne l'indicateur du niveau d'étude, l'analyse conduite par Mallet (2008, p. 191-192)[54] à partir des données extraites du corpus « Phonologie du Français Contemporain » (Durand, Laks et Lyche 2009) (cf. Partie 2.Chapitre 1.1) semble être en contradiction avec l'hypothèse selon laquelle les locuteurs les plus scolarisés réalisent le plus de liaisons (Gadet 1989, p. 52-53).

Cela pourrait s'expliquer d'après Mallet (2008) par le fait qu'une scolarité obligatoire diffusée (où l'apprentissage de l'orthographe est central) a assuré la généralisation et l'imposition de l'écrit (Laks 2005b, p. 114). Durand et al. (2011) avancent l'idée que le manque de significativité enregistré pourrait être attribué à « un nivellement des usages entre les différents groupes de locuteurs suggérant des pratiques communes dans la conversation ordinaire » (Durand et al. 2011, p. 73). Des écarts plus importants entre divers niveaux d'étude auraient donc pu être enregistrés si les mêmes sujets avaient été placés dans des situations de forte tension linguistique telles que la prise de parole publique.

En revanche, concernant l'indicateur de l'âge, toutes les études (Malecot 1975; Ashby 1981; Mallet 2008) obtiennent des résultats comparables qui indiquent que la fréquence de liaison est corrélée à la variable d'âge étant donné que la liaison est plus souvent produite chez des sujets plus âgés que chez des locuteurs plus jeunes. En particulier, les données de l'étude de Mallet (2008) indiquent que le taux de production de la liaison est proportionnel à quatre tranches d'âge (< 30 ans ; 30 - 45 ans ; 46 - 60 ans ; 60 ans). La différence semblerait être évidente dans une comparaison (tous styles de parole confondus) entre les locuteurs de plus de

[54] Taux de réalisation par nombre d'années de scolarité : moins de 14 ans de scolarité (46,46%) ; entre 14 et 20 ans de scolarité (48,91%) ; plus de 20 ans (47,27%) ; scolarité inconnue (46,30%).

60 ans et les trois autres tranches. En revanche, celle-ci serait plus marquée en parole semi-spontanée qu'en lecture chez les sujets âgés de 20 à 40 ans que chez les jeunes de moins de 20 ans. Durand et al. (2011) semblent suggérer deux types d'interprétations possibles pour ces résultats. D'une part, ils pourraient indiquer une réduction progressive au niveau diachronique de la liaison souvent évoquée et attribuée à l'évolution du français[55]. D'autre part, ces données pourraient être associées à une tendance d'*age grading* selon laquelle les sujets les plus âgés préserveraient les usages linguistiques.

Malgré l'analyse de l'ensemble de ces facteurs, comme Encrevé (1988, p. 258) et Mallet (2008, p. 203) le soulignent, la variation de la liaison ne peut pas être saisie par une simple analyse de variables inter-locuteurs mais demande également la prise en compte d'une variation intra-individuelle à l'intérieur d'un même discours résultant d'une pluralité de facteurs personnels et situationnels difficilement prévisibles tels que le degré de tension (de surveillance, de formalité), la fatigue, l'estimation que fait le locuteur de la réception de son discours de la part de l'auditoire et le stress généré par l'appareil enregistreur.

2.6.1.2 Variation diachronique

Un nombre réduit de travaux a été consacré à l'étude de cette dimension de variation. Certains auteurs (Grammont 1914; Bauche 1920; Guiraud 1965) considèrent « sur des bases impressionnistes ou normatives » que le taux de réalisation se serait historiquement affaibli. Si Grammont (1914, p. 130) voit dans « les nouvelles générations » la cause de cet affaiblissement, d'autres (Bauche 1920; Guiraud 1965) l'attribuent à une variation diastratique. Selon Bauche (1920), « la liaison se fait de moins en moins en français […]. En langue populaire, hormis certains cas déterminés, elles ont presque complètement disparu » (Bauche 1920, p. 52). Guiraud (1965) considère également que la liaison « est réduite au minimum dans le français populaire […] en français populaire on ne fait plus la liaison » (Guiraud 1965, p. 106‑107).

Les résultats des études empiriques montrent que la fréquence de liaison est corrélée à la variable d'âge et elle est plus souvent produite chez les locuteurs les plus âgés que chez les locuteurs les plus jeunes (Malecot 1975; Ashby 1981; Mallet 2008). Laks (2009) a montré que l'étude de la parole publique constitue un « observatoire privilégié » pour analyser les changements diachroniques et les dynamiques longues qui affectent le phénomène de la liaison[56]. Les « corpus d'hommes et femmes politiques 1 » (dorénavant HPOL1) (Laks 2009)

[55] Celle-ci, comme nous le verrons dans le paragraphe suivant, sera remise en discussion par Laks (2005c).

[56] Pour illustrer ce point il signale par exemple qu'au début du 20ème siècle, la liaison variable avec les noms propres était encore possible alors que « dans les années 1960 Chevalier et al. (1964) la prohibent

et « corpus d'hommes et femmes politiques 2 » (dorénavant HPOL2) (Laks 2014) sur lesquels il s'appuie comprennent les discours prononcés par des hommes et des femmes politiques français entre les années 1908 et 1996 (corpus HPOL1) et 1999 et 2011 (corpus HPOL2).

En 2005, il compare les données extraites du premier de ces deux corpus aux résultats d'autres corpus relevant de trois périodes historiques différentes : (a) les données de parole semi-spontanée du corpus « Phonologie du Français Contemporain » (Durand, Laks et Lyche 2009) (cf. Partie 2.Chapitre 1.1) ; (b) les résultats de l'enquête menée à Villejuif (Laks 1977; Laks 1980; Laks 1983) ; (c) les données collectées par Brunot en 1912 et 1913. Grâce à la comparaison de l'ensemble de ces corpus, Laks (2005c) montre que bien que l'analyse du taux global de liaisons réalisées confirme l'hypothèse traditionnelle d'une tendance historique à l'affaiblissement quantitatif de la liaison, une analyse plus fine des données remet en discussion cette conclusion. Cette étude montre en effet « une organisation diastratique qui l'emporte de beaucoup sur la tendance diachronique à la non liaison » (Laks 2005c, p. 6). En effet le nombre de liaisons variables réalisées (38,7%) par les ouvriers nés au XIXème siècle, attesté par Brunot, serait seulement légèrement plus important que celui des adolescents de Villejuif (28,2%) enregistré soixante ans plus tard et en tout cas inférieur au taux de liaisons réalisées par les hommes politiques pour la même période (corpus HPOL1). De plus, les résultats du corpus « Phonologie du Français Contemporain » (cf. Partie 2.Chapitre 1.1), représentatif de l'usage contemporain de la liaison, montrent un taux de réalisation variable comparable à ce qu'il était chez les ouvriers un siècle auparavant (40,7%). L'auteur en conclut que contrairement aux hypothèses traditionnelles, la liaison semble être caractérisée par une certaine stabilité historique selon laquelle la variation diastratrique et diaphasique l'emporterait sur le changement diachronique.

> « […] contre les contempteurs d'un usage qui affaiblirait continûment la liaison, une très grande stabilité du poids des facteurs sociaux et stylistiques qui la contrôlent : loin de s'affaiblir la liaison facultative reste globalement stable dans sa variabilité ». (Laks 2005c, p. 6)

Quant aux changements historiques qui ont intéressé le discours des dirigeants politiques, l'auteur observe que la fréquence de réalisation de la liaison variable, après une régression régulière qui aurait eu lieu jusqu'aux années 1970, connaît jusqu'aux années 1980 une forte remontée apportant « un démenti cinglant aux orthoépistes de tous ordres qui voient dans l'affaiblissement fantasmé de la liaison le signe même de la décadence de la langue française ». L'analyse fine de ces données a permis de mettre en évidence que l'augmentation enregistrée

explicitement, recommandant par exemple une prononciation non liée du nom de l'ancien Président de la République Vincent Auriol » (Laks 2014).

dans les années 1970-1980 correspond assez précisément à la période où le phénomène de non-enchaînement de la liaison, attesté par Encrevé (1988), atteint le seuil de significativité de 10% pour ensuite régresser à des niveaux marginaux dans les années suivantes. Ces fluctuations (liées principalement aux variations du phénomène socio-stylistique de la liaison non-enchaînée) (cf. Paragraphe 2.1.1) confirmeraient selon l'auteur la primauté des variables diastratiques et diaphasiques par rapport aux facteurs diachroniques.

2.6.1.3 Variation diatopique

Dans le cadre de cette dimension de variation, l'examen du corpus « Phonologie du Français Contemporain » (Durand, Laks et Lyche 2009) (cf. Partie 2.Chapitre 1.1) s'avère crucial. En effet les données extraites de ce corpus permettent de comparer les réalisations enregistrées dans différents points d'enquête de l'ensemble de l'espace francophone[57]. L'étude contrastive des résultats de deux points d'enquêtes représentatifs respectivement des variétés du nord et du sud de la France a permis de remettre en discussion l'hypothèse formulée (qui ne se fonde d'ailleurs sur aucune étude quantitative) par Brun (1931 cité par Durand et Lyche 2009) d'après laquelle une fréquence plus faible de liaisons serait attestée chez les sujets de Marseille[58]. Néanmoins, les auteurs observent une légère différence statistique dans le cas de quelques contextes de liaisons variables en faveur des régions du sud de la France qui pourrait être liée au substrat occitan. Cette analyse a été élargie par Coquillon et al. (2010) à 67 points d'enquêtes des régions du nord de la France qui ont été comparés à 67 points d'enquêtes des régions du sud de la France. L'examen des contextes de liaison variable enregistrés en parole semi-spontanée amène les auteurs à observer une homogénéité globale de la réalisation de la liaison entre nord et sud.

En revanche, en ce qui concerne les variétés de français laurentien Côté (2013) et Eychenne et Walker (2010) observent que « la liaison est globalement nettement moins réalisée que dans le FR [*variétés de français de France*], ce qui se comprend aisément dans un contexte où la norme orthographique est moins prégnante »[59] (Eychenne et Walker 2010, p. 257).

Cela pourrait s'expliquer par le fait que la liaison n'est jamais réalisée après le pronom « ils » et qu'elle est variable après le pronom « on » (Côté 2012). D'autres particularités qui

[57] Afrique, Belgique, France, Louisiane, Québec et Suisse romande (cf. Partie 2.Chapitre 1).

[58] « Les liaisons sont beaucoup moins fréquentes qu'en français commun. Cette négligence, ainsi que la paresse à articuler les groupes de consonnes donne au parler du provençal ce caractère de vulgarité qui choque le nouveau-venu » (Brun 1931 cité par Durand et Lyche 2008).

[59] Côté (2015) souligne en effet que l'influence de la norme est réduite dans les variétés nord-américaines (français du Québec ; français de Louisiane).

caractérisent cette variété sont la liaison en [t] après les séquences verbales « je suis » et « tu es » et la liaison en [l] après le pronom « ça » (Côté 2012).

Quant aux variétés de français d'Afrique[60] Bordal et Lyche (2008) montrent également des taux de réalisation de la liaison variable moins importants[61]. À titre d'exemple Côté (2015) signale que la liaison variable après « est » est peu fréquente en Afrique (14%) (Bordal et Lyche 2008 ; Boutin, Gess et Guèye 2012) et encore moins en Louisiane (Boutin et Lyche 2014) alors qu'elle est plus étendue au Québec (Côté 2012 ; 2013 ; 2014) et en Europe (Durand et Lyche 2008 ; Durand et al. 2011). Boutin et Lyche (2014) (en comparant les productions des locuteurs de la Louisiane et de la Côte d'Ivoire) soulignent que ces différences pourraient s'expliquer par un contact fréquent des variétés nord-américaines et africaines avec des langues à accent de mot (anglais, langues africaines).

> « Nous observons […] que les deux terrains participent au mouvement général de l'amenuisement de la liaison variable en français […] des facteurs prosodiques de langues en contact sont susceptibles de faciliter cette tendance ». (Boutin et Lyche 2014, p. 301)

2.6.2. Variation intra-locuteur

2.6.2.1 Variation diaphasique

Le niveau stylistique du discours est considéré par la plupart des auteurs (Delattre 1955; Ågren 1973; Anderson 1975) comme l'un des facteurs les plus déterminants dans la réalisation de la liaison variable. Delattre (1955) fait observer que le choix du style exigé par la situation de communication peut faire varier significativement la fréquence de réalisation de la liaison. Il distingue quatre styles se situant entre l'oral spontané (conversation familière, conversation soignée) et l'écrit oralisé (conférence, récitation des vers). Quant à Fouché (1959), il ne considère que deux registres (l'un soutenu et l'autre courant) sans faire référence à d'autres styles tout en signalant que lorsque « le ton s'abaisse et devient plus familier, on lie de moins en moins ». Ågren (1973) a également observé que les styles les plus soignés, qu'il a pu analyser à partir de son corpus d'émissions radiophoniques, étaient compris parmi ceux qui présentaient le taux plus élevé de liaisons variables réalisées. En revanche, l'étude de Malécot (1975), limitée au style de la « conversation familière de la bourgeoisie cultivée » de Paris, n'a pas relevé d'importantes variations stylistiques[62].

[60] En particulier les variétés du Burkina Faso, le Mali, de la République Centrafricaine et de la Côte d'Ivoire.

[61] La liaison variable semblerait être peu fréquente après « c'est » et « est » mais également après « très » et « chez ».

[62] Cela pourrait confirmer l'hypothèse avancée par Durand et al. (2011) selon laquelle la conversation soignée décrite par ces études se rapprocherait davantage des styles très formels que de la conversation ordinaire.

Ces registres qualifiés de « soutenus, soignés » sont pourtant difficilement descriptibles. Il apparaît ainsi délicat de comparer les études et surtout de les situer dans le cadre d'une représentation traditionnelle de registres de langue qui distingue : registre soutenu (recherché, soigné, contrôlé) ; registre standard (courant, commun) ; registre familier (relâché, spontané) et registre populaire (argotique) (Mallet 2008).

Cette problématique est sans doute liée au fait que ces dénominations de registre ne relèvent pas d'une évidence mais plutôt de classifications établies par des linguistes.

> « […] les dénominations de français familier, populaire, des jeunes… sont des termes d'experts, ou de la reprise publique. C'est pourtant bien ces classifications ordinaires qu'épouse la notion de variété, qui n'est pas une évidence mais une construction. Elle suppose que les traits variables puissent converger en un tout cohérent, sur la base duquel il serait possible de constituer des objets cernables et énumérables ». (Gadet 2007, p. 22)

C'est la raison pour laquelle plusieurs auteurs ont étudié la variation de la liaison en fonction du registre s'appuyant sur des situations et des modalités de production différentes, telles que la lecture (donnant lieu à un style plus soigné) et la conversation spontanée (relevant d'un registre plus familier). Dans ce sens les dimensions diaphasique et diamésique interagissent fortement entre elles et les registres ou styles varient aussi bien en fonction « des circonstances, de la situation et des participants à l'échange (style situationnel, en réponse à certaines composantes d'une situation), que selon les ressources déployées dans les interactions » (Gadet 2007).

2.6.2.2 Variation diamésique

Comme nous venons de le voir, certains auteurs (Lucci 1983; Fougeron et al. 2001; Mallet 2008) ont comparé la fréquence des liaisons réalisées en lecture et conversation (parole spontanée ou semi-spontanée). Cette approche permet d'étudier la variation de la liaison selon le registre résultant de situations et de modes de communication différents.

Lucci (1983) analyse les productions dans quatre types situationnels qu'il qualifie également de socio-culturels: la conférence, la lecture, l'interview et « la conversation de locuteurs socio-culturellement défavorisés » (Lucci 1983, p. 243). Il convient de souligner que dans le cas du dernier style situationnel l'auteur a étudié des locuteurs « défavorisés », ce qui laisse émerger une volonté de faire correspondre à la variation stylistique une diversité de profils socio-culturels. Les résultats indiquent une différence de fréquence de liaison entre les styles situationnels d'écrit oralisé (conférence et lecture) et ceux de parole semi-spontanée (interview et conversation).

Fougeron et al. (2001) ont utilisé un texte créé à partir d'extraits d'articles du journal « le Monde » dont la lecture a été comparée aux productions de conversation spontanée en style familier entre les locuteurs et les expérimentateurs. Les résultats de ce travail ont montré une variabilité inter-individuelle mettant en évidence un taux plus important de liaisons produites en situation de lecture qu'en conversation.

Mallet (2008), dans le cadre du projet « Phonologie du Français Contemporain » (Durand, Laks et Lyche 2009) (cf. Partie 2.Chapitre 1), compare les productions des liaisons observées dans la lecture d'un texte construit et celles relevant de deux styles situationnels de conversation (une discussion informelle et une entrevue formelle). Cette méthodologie essaie de rapprocher au maximum les trois registres décrits ci-dessous de ceux définis par Delattre (1955) :

> « […] (i) la lecture du texte qui se rapproche du style « conférence » proposé par Delattre (1966), (ii) la conversation libre ou informelle qui correspond au registre « familier » ; et (iii) la conversation guidée qui donne un ton plus formel au dialogue et qui se rapproche donc du registre « soigné » de Delattre ». (Mallet 2008, p. 148)

Les résultats du corpus « Phonologie du Français Contemporain » (Mallet 2008; Durand et al. 2011) confirment une différence entre lecture et conversations, sans pour autant pouvoir distinguer la fréquence de réalisation de la liaison entre la conversation informelle et l'entrevue formelle.

> « Dans la conversation guidée, nous observons 9 301 liaisons enchaînées (le modèle par défaut) sur 20 908 codages, soit 43,13% des cas. Dans la conversation libre, on observe 7 503 liaisons enchaînées sur 17 393 codages, soit 44,52 %. La différence entre les deux n'est pas statistiquement significative ». (Durand et al. 2011, p. 68)

Cela pourrait s'expliquer selon les auteurs aussi bien par des difficultés d'ordre méthodologique (incapacité de gestion de deux registres distincts qui ne correspondent pas aux registres souhaités)[63] que par le fait que la liaison ne constitue pas un indicateur permettant de distinguer les deux registres. Il serait donc possible que le style de parole attesté dans le corpus « Phonologie du Français Contemporain » (Durand, Laks et Lyche 2009) (cf. Partie 2.Chapitre 1) soit relativement uniforme du point de vue de la liaison et que, pour observer des différences diaphasiques plus marquées, il s'avère nécessaire de prendre en compte « des situations beaucoup plus *formelles* comme les prises de parole publiques ».

[63] Cependant, les auteurs soulignent que des travaux menés au laboratoire LIMSI indiqueraient que les deux types de parole spontanée présentent des propriétés phonétiques partiellement différentes (Boula de Mareüil, Adda-Decker et Woehrling 2007).

3. Conclusion

À l'issue de ce bilan non exhaustif des variables qui influencent la production de la liaison, il apparaît que même si les facteurs phonologiques et syntaxiques constituent les conditions nécessaires pour la réalisation de la liaison, ils n'en constituent qu'une partie des dimensions de variation.

En effet non seulement ils ne permettent pas d'expliquer le comportement hétérogène de la liaison mais ils ne parviennent pas non plus à distinguer les liaisons qui sont toujours réalisées par tout locuteur et dans tout contexte (appelées obligatoires ou catégoriques selon les auteurs cf. Partie 1.Chapitre 4.2.1) de celles dont la réalisation est variable (liaisons facultatives ou variables selon les auteurs).

En effet, comme nous l'avons vu dans les pages précédentes, l'analyse du degré d'étroitesse entre les mots demande parfois le recours à des facteurs sémantiques et lexicaux (Laks 2005a). D'autres auteurs (Bybee 2001a; Bybee 2005) n'hésiteront pas à considérer la cohésion syntaxique comme le résultat même de la fréquence de cooccurrence des mots de liaison.

Ces deux aspects ont en effet été jugés indissociables par Ågren (1973) qui les met également en relation avec un autre facteur lexical à savoir la longueur des mots de liaison. Celui-ci est à son tour en rapport direct avec l'organisation prosodique de l'énoncé en groupe accentuel. L'ensemble de ces observations suggère que les facteurs syntaxiques, (Damourette et Pichon 1911-1927 ; Delattre 1966), sémantiques et rythmiques (Grammont, 1914) et lexicaux (Ågren1973; De Jong 1991; Bybee 2005) évoqués pour rendre compte du comportement de la liaison sont interdépendants les uns des autres.

De plus à l'ensemble de ces variables linguistiques internes s'ajoute l'éventail des dimensions de variation sociolinguistique.

Si la variation inter-locuteur liée à l'âge et au sexe du locuteur semble être peu marquée dans la production de la liaison, en revanche la liaison représente un important indicateur social et stylistique dont la réalisation ou non-réalisation détermine une distinction sociolinguistique positive ou négative établie en fonction de la norme linguistique en vigueur.

Cet aspect met en évidence la nécessité, dans le cadre d'une description des usages de la liaison, de tenter d'identifier les caractéristiques de la prononciation de référence ainsi que les approches à adopter par rapport à cette norme. La diversité des variables d'étude de la liaison

(en particulier celles sociolinguistiques), nous obligent à rechercher une norme prononciation de référence et sur l'approche à utiliser pour décrire les usages de la liaison.

Chapitre 3. Normes de prononciation et approches méthodologiques pour l'étude de la liaison

Étant donnée la pluralité des variables en jeu dans l'analyse de la liaison (en particulier le rôle crucial de la dimension sociolinguistique) (cf. Partie 1.Chapitre 2.2.6), il nous paraît important de nous interroger sur le modèle de prononciation de référence et sur l'approche à utiliser pour décrire les usages de la liaison.

Cette question a des origines très anciennes (Laks 2014) et concerne en particulier l'usage de la liaison variable, qui comme nous l'avons vu dans le chapitre précédent (cf. Partie 1.Chapitre 2.2.6.1.1), constitue un fort indicateur sociolinguistique. Dans ce chapitre nous présenterons quelques modèles de prononciation de référence du français ainsi que les différentes approches d'étude de la liaison. Cette démarche nous permettra de mieux comprendre la diversité du contenu des classifications de la liaison (qui fera l'objet du prochain chapitre cf. Partie 1.Chapitre 4.3) basées sur des normes d'usage qui, comme le souligne Le Prieult (2006, p. 126), cèdent « aux valeurs historiques dominantes (morales, esthétiques, politiques) ».

1. Quel français de référence pour la description de la liaison ?

La liaison étant un phénomène socio-stylistique relativement marqué, le choix de la norme de prononciation de référence joue un rôle déterminant dans la description et la classification de ses usages. Or, dans le cas de la liaison, cette norme, comme le souligne Laks (2002) ne peut pas se limiter à préciser les aspects orthoépiques purement lexicaux (à savoir la prononciation des mots isolés) mais doit inclure également les phénomènes qui se réalisent à la jonction entre ces mots. La liaison constitue comme nous l'avons vu dans le chapitre précédent (cf.Partie 1.Chapitre 2) un phénomène extrêmement variable qui comme le rappelle Laks (2002) « ne se laisse pas traiter dans le vocabulaire prescriptif de la règle obligatoire et de l'exception » et dont le bon usage a longtemps représenté « un affaire de goût, d'éducation et de mesure » (Laks 2002, p. 7).

Le modèle de prononciation de référence, comme nous le verrons dans les pages qui suivent, varie en fonction des valeurs et des groupes sociaux dominants. En particulier, la norme phonétique de prononciation a été soumise au cours de diverses périodes à plusieurs révisions

qui relèvent d'un processus de « démocratisation » et de « décentralisation » de cette notion (Borrel et Billières 1989). En effet, si dans un premier temps la norme de référence correspondait au français parlé par un groupe social très limité, à savoir la bourgeoisie parisienne, au fur et à mesure, d'autres modèles socialement et géographiquement plus larges se sont imposés. Cependant, comme l'observent Borrell et Billières (1989), la norme de prononciation sera toujours géographiquement limitée au territoire de l'Hexagone en excluant implicitement les usages d'autres variétés de l'espace francophone. De plus, en tant que référence, elle possède un caractère prestigieux et presque politique qui lui vient de son appartenance à un « groupe dominant » et qui fait de la norme de prononciation « une affaire de classe » (Morin 2000, p. 96). Qu'il s'agisse d'une classe privilégiée et cultivée ou d'une classe dirigeante, la définition des contours de ces groupes sociaux de référence est complexe à l'époque moderne (Laks 2002). Cette difficulté, comme le souligne Morin (2000), vient de l'absence d'un consensus quant à la norme explicitement établie et reconnue par la communauté.

En effet, la notion de norme de prononciation reçoit souvent différentes dénominations et définitions, selon les auteurs, ce qui montre la difficulté d'identifier, au fil des périodes, les caractéristiques phonétiques et phonologiques de ces modèles (Morin 2000 ; Francard et al. 2000 ; Laks 2002 ; Lyche 2010).

> « […] les phonologues sont en effet immédiatement confrontés à la diversité géographique, sociale et stylistique de la langue, à la variabilité des usages, et à l'absence d'une norme stable, explicite et parfaitement définie. Si la langue écrite connaît, depuis le milieu du XIXe siècle au moins, une norme grammaticale de référence explicite et quasiment fixe, tel n'est pas le cas de l'oral. La variété des usages y est première et la norme, implicite, mouvante et toujours difficile à dégager, seconde ». (Laks 2002, p. 5)

Dans les pages qui suivent, nous allons retracer un parcours (non exhaustif) des modèles de prononciation du français contemporain au XXème siècle correspondant dans un premier temps aux variétés de français parlé par une classe sociale restreinte (haute bourgeoisie parisienne). Ceux-ci seront ensuite remplacés par des modèles basés sur des « usages moyens » tels que le « français standard » (Léon 1968) et le « français standardisé » (Carton et al. 1983) (cf. Borrel et Billières 1989 ; Galazzi 2015b).

1.1. Le français de la « bonne société parisienne »

Le modèle de prononciation de français qui a été jusqu'aux années 60 adopté comme référence était celui de la haute bourgeoisie de Paris. Cette couche sociale reçoit toutefois, selon les auteurs, des définitions différentes qui montrent toute l'hétérogénéité et la variabilité stylistique qui caractérisent cette « bonne société parisienne » (Martinon 1913). Si certains

auteurs, parmi lesquels Pichon (1938) et Grammont (1914), font au début du XXème siècle référence à une prononciation à laquelle on accède par « échelons de générations », d'autres (Fouché 1959) recherchent ce modèle dans « la prononciation en usage dans une conversation 'soignée' chez les Parisiens cultivés »[64].

Pour les premiers, comme Pichon (1938), « ce n'est pas le mérite personnel d'un individu, ce n'est pas son instruction, ni ses titres universitaires, ni sa culture intellectuelle et morale » qui permettent au locuteur d'être classé parmi ceux qui possèdent « la parlure dite bourgeoise, ou mieux optimale » (Pichon 1938, p. 108). C'est donc, comme l'observe l'auteur une question purement de classe et d'appartenance aux « plus vieilles familles parisiennes dont sont issues les officiers généraux et les évêques ».

En revanche, à partir de la moitié du XXème siècle, certains auteurs commencent à faire référence à une classe sociale dont le prestige semble être lié à des qualités personnelles telles que la culture (Fouché 1959) ou les mérites qui ont permis à ce groupe de s'affirmer comme « classe sociale dirigeante de la capitale » (Malécot 1977, p. 1)[65]. On voit donc ici une ouverture vers la norme de prononciation de la moyenne et de la petite bourgeoisie parisiennes soutenue par une partie des auteurs relevant de l'« intelligentsia républicaine laïque ou de gauche ».

> « C'est ainsi que Fouché, qui en 1936 parle encore de *français parisien de la bourgeoisie cultivée,* ne mentionne plus explicitement dans son *Traité de prononciation française* (1956) la classe sociale dont il décrit l'usage et se contente de dire qu'il décrit le *français parisien cultivé*, laissant croire que celui-ci peut aussi être utilisé par une partie de la moyenne et de la petite bourgeoisie ». (Morin 2000, p. 94)

Il est intéressant de remarquer que ces auteurs (Grammont 1914; Fouché 1939; Fouché 1959) qui défendent et légitiment la primauté des usages de la bourgeoisie parisienne n'appartiennent pas le plus souvent à cette classe et que pourtant ils les décrivent plus au moins convaincus de savoir les maîtriser ou d'en connaître les traits principaux.

Cette tradition sera perpétuée jusqu'aux années soixante[66] et s'appuiera sur une description d'une norme sociologique et idéologique plus que sur des enquêtes scientifiques. C'est justement le développement d'études phonético-phonologiques à partir de la deuxième moitié du XXème siècle qui permettra de dépasser ces modèles de prononciations. En

[64] Fouché (1939) reconnaît quatre types de prononciations possibles au sein de cette classe : « le débit de conversation familière, celui de la conversation soignée (qui est celui de la lecture), celui de la conférence ou du sermon, et enfin celui de la diction » (Fouché 1939, p. 211).

[65] La norme de prononciation pour Malécot (1977) est celle « des Parisiens authentiques (de naissance, d'éducation, etc.) parmi les chefs d'entreprises, hauts fonctionnaires, cadres supérieurs et tous ceux qui ont des situations de responsabilité dans les professions libérales » (Malécot 1977, p. 1).

[66] Bien qu'au cours des années qui précèdent d'autres auteurs (Martinon 1913; Bruneau 1931) eussent déjà contesté le modèle basé sur la prononciation de Paris en proposant d'élargir la norme aux variétés parlées par le reste du nord de la France.

particulier, l'étude de Martinet (1971) sur la prononciation du français contemporain remettra en discussion la primauté du français parlé par « la bonne société parisienne » montrant la pluralité des prononciations ainsi que la diversité des systèmes phonologiques présents sur le territoire français, dont le français parisien ne constitue qu'une variété.

Cependant, Martinet (Martinet et Walter 1973), comme le souligne Morin (2000, p. 126-127), propose aussi, comme ses prédécesseurs, une norme d'usage basée sur une variété de la classe sociale à laquelle l'auteur lui-même[67] appartient. Il fait notamment référence aux usages des personnes caractérisées par une « mobilité géographique avec prédominance du séjour dans la Capitale » (Martinet et Walter 1973, p. 18) dont la prononciation a été épurée « des bizarreries et des localismes ».

> « […] on devrait accorder la préférence aux personnes d'origine et d'attaches provinciales, car une certaine mobilité géographique est susceptible d'éliminer les bizarreries et les *localismes* ». (Martinet 1964, p. 131)

Malgré ces observations, il est évident que la norme de la « bonne société » parisienne commence à perdre son autorité et sera bientôt rejetée en faveur d'un modèle basé sur la prononciation de la langue des professionnels de la parole publique.

1.2. Le français des medias : « le français standard »

C'est en opposition à la fois au modèle de prononciation parisienne et à celui proposé par Martinet (Martinet 1964; Martinet et Walter 1973) qu'à partir des années 60 jusqu'au milieu des années 80 va s'imposer une nouvelle norme définie par Léon (1968) comme « français standard ». Celle-ci serait représentée par la prononciation des professionnels de la parole publique tels que les annonceurs et les interviewers radio-télévisés.

> « […] il existe une prononciation standard dont le niveau moyen est *grosso modo* représenté par les annonceurs et les interviewers de la radio… ; dans l'ensemble, leur prononciation reflète l'usage moyen, sans recherche (pour plaire au grand public) et sans familiarité excessive (à cause du micro) ». (Léon 1968, p. 68)

Cette notion de « français standard » a le mérite de mettre en évidence le rôle joué par les medias dans la diffusion des modèles de prononciation. Cependant, le concept de « niveau moyen » de prononciation, comme le souligne Morin (2000) à la suite de Warnant (1987) serait « un terme creux » qui recouvre toute variété de français non méridional et qui ne dit rien quant aux critères linguistiques qui ont fait préférer tel ou tel annonceur ou interviewer. La notion de

[67] Né en Savoie, d'une famille savoyarde, qui a déménagé de la région d'origine pour s'installer définitivement à Paris à l'âge de 11 ans.

« français standard » reste donc relativement ambigüe à mi-chemin entre une prononciation parisienne épurée des extrêmes et un modèle résultant du nivellement de la prononciation en France métropolitaine. Malgré les critiques avancées à cette notion, l'uniformisation de la prononciation autour des grandes villes a été à plusieurs reprises mise en avant par différentes enquêtes (Walter et Martinet 1982).

1.3. Le français « qui passe inaperçu » : « le français standardisé »

Une autre conception de norme de prononciation émergera à partir du milieu des années 1980 et correspondra à une variété qui ne présente rien que l'on puisse remarquer et qui la fasse passer comme « inaperçue ».

> « Les bonnes prononciations sont celles qui passent inaperçues, les mauvaises, celles qui soudain vous rappellent, ne serait-ce qu'à un niveau très inférieur de la conscience, que votre interlocuteur est de telle origine, nationale, géographique ou sociale ». (Martinet et Walter 1973, p. 17)

Cette notion avait déjà était proposée par Dubois (1965) qui parlait d'un « français neutralisé » pouvant assurer la plus large intercompréhension. Cependant le concept de « français standardisé » sera définitivement suggéré par Carton et al. (1983).

> « Il semble maintenant qu'il existe, dans toutes les régions de France, et non seulement à Paris, une prononciation commune acceptée partout qu'on pourrait appeler standardisée. On la définira d'abord linguistiquement, par des caractères communs, opposés aux variantes qui excluent leurs auteurs du français standardisé et qui marquent un accent régional ». (Carton et al. 1983, p. 77)

Néanmoins, comme le soulignent certains auteurs (Warnant 1987 ; Morin 2000), il est difficile dans ce contexte d'établir des critères objectifs capables de définir une prononciation qui passe inaperçue.

De plus, il convient de souligner que malgré les processus de « démocratisation » et de « décentralisation » qui ont caractérisé l'évolution de la norme de prononciation, celle-ci reste encore une variété de « prestige » socialement et géographiquement limitée (Borrel et Billières 1989). En effet l'élargissement social de la norme n'a pas atteint les classes sociales défavorisées et moins cultivées qui représentent toujours des « exclus d'office » (Morin 2000, p. 99). Il en va de même, comme nous l'avons déjà observé, pour les locuteurs possédant une variété de prononciation francophone autre que celle de la France hexagonale qui sont « absents » dans la description des modèles de prononciation plus récents.

Pour résumer, au terme de ce bref parcours de l'évolution de la norme de prononciation, il paraît évident, comme le soulignent Laks (2002) et Galazzi (2008a ; 2015b), que la notion de français de référence possède des contours encore flous et difficiles à saisir[68]. À la suite de Morin (2000), Laks (2002) rappelle que cette norme hétérogène est « une construction purement doxique » qui relève d'une accumulation de traditions anciennes, d'intuitions personnelles des phonologues et qui manque de grandes études systématiques, représentatives et solides de la variété d'usages. Cet aspect méthodologique ainsi que la norme de prononciation seront l'objet d'une profonde évolution.

2. Approches méthodologiques d'étude de la liaison

Dans les pages qui suivent nous examinerons les méthodologies utilisées pour la description et l'étude de la norme de prononciation de la liaison et la façon selon laquelle elles ont évolué au cours du XXème siècle. L'étude de la liaison a en effet connu le passage d'une tradition méthodologique basée sur les intuitions des phonologues et linguistes au développement de travaux s'appuyant sur des résultats d'enquêtes de plus en plus systématiques.

2.1. Approches normatives

Au début du XXème siècle, la norme de prononciation de la liaison sera décrite par plusieurs pédagogues et phonéticiens (parmi d'autres Martinon 1913 ; Grammont 1914 ; Grevisse 1936 ; Fouché 1959 ; Carton 1974 ; Malmberg 1969). Le modèle de prononciation de référence correspond encore à la variété de langue parlée par les classes sociales dominantes, celle en usage dans « la bonne société parisienne, constituée essentiellement par les représentants des vieilles familles et de la bourgeoisie » (Grammont 1914, p. 1) ou « dans une conversation soignée chez les Parisiens cultivés nés vers la fin du XIXème siècle ou plus tard » (Fouché 1959, p. 3).

Dans ce sens, l'argumentation du discours normatif se base sur des jugements de grammaticalité qui, comme le rappelle Le Prieult (2006, p. 126), cèdent aux jugements sociologiques qui opposent l'usage commun au parler de l'élite. En ligne avec cette tradition normative, ces auteurs décrivent les bons usages et proscrivent les fautes. Il est intéressant

[68] En revanche, Francard 2001 (cité par Galazzi 2015b) distingue le français standard des français de référence dans la mesure où « le standard reste un modèle de correction abstrait (la langue normée), à côté des variétés attestées (les français de référence) » (Francard 2001, cité par Galazzi 2015b).

d'observer comment ces prescriptions, qui règlent la variation de la liaison, se basent le plus souvent chez les normatifs sur des évaluations subjectives employant, comme le soulignent Auroux et Weil (1991), un vocabulaire qui relève de la normativité morale.

> « Ainsi le vocabulaire des évaluations linguistiques est largement emprunté au vocabulaire moral (*bien parler, mal parler, une faute, bon / mauvais usage*, etc. …) et des pratiques prescriptives (il ne faut pas dire / il faut dire…). La question de la normativité dans les sciences du langage passe par une réflexion sur leurs analogies formelles avec la morale, quoique bien évidemment elle ne s'y puisse réduire ». (Auroux et Weil 1991, p. 82-83)

Les prescriptions de bon usage définies par ces auteurs décrivent de façon précise l'emploi de la liaison afin d'éviter « les fautes parfois grossières que certains [*apprenants de français et locuteurs natifs*] commettent » (Fouché 1959, p. II).

En particulier, Grevisse (1936) cite des exemples de fautes de liaison qui comportent l'introduction par analogie d'une CL dans des contextes où elle n'a pas de correspondances graphiques (cf. Partie 1.Chapitre 4.2.4).

> « […] ce sont des fautes appelées familièrement *cuirs* (quand la consonne introduite est un [t]) et velours (quand c'est un [z]) ». (Grevisse 1936, p. 50)

Cet auteur distingue les liaisons obligatoires (qui doivent toujours être réalisées) de celles qui sont simplement recommandées (par le bon usage) et les liaisons qui ne se font pas. De plus, chez Grevisse (1936) la question de la variation de la liaison est abordée de façon marginale.

> « On entend beaucoup moins de liaisons dans la conversation ordinaire que dans le discours soigné et dans la lecture à voix haute (surtout celle de la poésie). Les Parisiens ont tendance à abandonner les liaisons qui se maintiennent mieux en province et en Belgique. Les orthophonistes ne sont pas toujours d'accord entre eux sur les cas particuliers, selon le milieu qu'ils observent et qu'ils prennent comme modèle ». (Grevisse 1936, p. 52)

Quant à Fouché (1959), il distingue « les cas où on fait la liaison » des « cas où on ne fait pas la liaison ».

> « La question [*quand fait-on la liaison*] n'est pas simple, car dans le français dit 'correct' il y a plusieurs variétés à distinguer, et dans chacune d'elles l'usage n'est pas nécessairement le même ». (Fouché 1959, p. 437)

Dans la préface, Fouché (1959) signale en effet que le cas de liaison a été le seul où il a été « absolument nécessaire » d'élargir l'étude à d'autres registres stylistiques : la conversation courante et le style soutenu. Ceux-ci correspondent respectivement à un style de « conversation familière » (que Grevisse avait déjà qualifié de « conversation ordinaire ») et au style des lectures oralisées à voix haute, des déclamations ou des lectures à voix basse d'un texte plus au moins littéraire.

De ces considérations, nous pouvons conclure que l'observation des usages oraux occupe une place plus que secondaire dans la description de la liaison. Les observations formulées s'appuient sur des connaissances personnelles et intuitives de la norme difficilement contrôlables. Les exemples utilisés par ces auteurs, comme le remarque Encrevé (1988), sont donc le plus souvent construits ou extraits de la langue écrite et employés pour des visées prescriptives et orthoépiques.

> « Fouché ne pratiquait pas d'analyse quantitative d'un corpus fermé, et, loin d'une moyenne 'objective', il construit une moyenne qualitative, c'est-à-dire normative, prescriptive : il s'agit d'indiquer le bon usage ». (Encrevé 1988, p. 44)

La description de liaison est réalisée sous forme de liste de règles et d'exceptions ayant pour but de corriger et d'améliorer la prononciation des « étudiants et des professeurs étrangers désireux de bien connaître notre langue » (Fouché 1959).

Nous présentons ci-dessous un exemple de formulation de règle qui régit la prononciation de la liaison entre un substantif pluriel et les conjonctions « et » et « ou » suivantes.

> « [*on fait la liaison*] entre un substantif pluriel et les conjonctions *et* et *ou* suivies immédiatement d'un autre substantif pluriel, lorsque le premier substantif se termine en *-ns*, *-nds,-nts* ou *-es* n'est pas précédé d'un déterminant et que les deux substantifs ne sont pas en fonction d'attribut *Raconter monts-z-et merveilles* […] mais sans liaison *Expéditeurs et producteurs se réunirent* ». (Fouché 1959, p. 469)

Comme l'observent Durand et Lyche (2003), la définition sociale de la norme chez ces auteurs exerce encore une forte influence qui empêche toute distinction entre description et prescription. Cependant, Encrevé (1988) reconnaît à ces travaux prescriptifs le mérite d'avoir élaboré une analyse fine et une première classification des contextes de réalisation de la liaison selon le registre du locuteur.

2.2. Approches descriptives

2.2.1. Premières études

Contrairement aux travaux précédents, la notion de variation des usages linguistiques sera au centre des études successives de la liaison (Delattre 1966 ; Ågren1973 ; Laks 1977 ; Morin et Kaye 1982 ; Lucci 1983 ; Malécot 1975 ; De Jong, 1988 ; Encrevé, 1988). En effet, au cours de la deuxième moitié du XXème siècle la préoccupation de définir les règles de bon usage de la liaison laissera la place à des approches plus descriptives basées sur des analyses quantitatives. Même les travaux non basés sur des analyses quantitatives de corpus (Delattre 1966 ; Léon 1992) s'éloignent des objectifs purement normatifs de la tradition précédente et

s'efforcent de comprendre le fonctionnement complexe de la liaison compte tenu d'une plus grande variété de facteurs, autres que ceux d'ordre socio-culturel.

Delattre (1966) est le premier à essayer d'analyser et de rendre compte de la totalité des facteurs linguistiques en jeu dans la réalisation de la liaison. Il analyse en particulier l'influence de cinq facteurs qu'il considère comme les plus importants dans la variation de la liaison : facteur stylistique, syntaxique, prosodique, phonétique et historique. Le facteur stylistique, considéré par Delattre (1966) comme l'un des plus déterminants dans la réalisation de la liaison, avait déjà été mis en évidence par Fouché (1959). Le degré d'union syntaxique entre le M1 et le M2 constitue une autre variable cruciale évoquée par Delattre (1966). Cependant, comme nous l'avons déjà souligné (cf. Partie 1.Chapitre 2.2.3.1), l'auteur observe que la réalisation de la liaison n'est pas seulement influencée par la relation syntaxique existant entre le M1 et le M2 mais qu'elle peut être conditionnée par d'autres facteurs tels que la prosodie de la phrase, la longueur des mots (facteur prosodique), le contexte phonétique (facteur phonétique) ou encore par des raisons purement historiques. L'étude de Delattre (1966) aboutira à un premier traitement classificatoire et à une tripartition de la liaison qui sera l'objet de plusieurs révisions opérées par les études successives (cf. Partie 1.Chapitre 4).

Quant à Léon (1992), il propose d'établir une classification de la liaison (correspondant à celle déjà définie par Delattre 1966) sur la base d'une règle de cohérence syntagmatique, critère étroitement lié à l'appartenance au groupe rythmique. À l'instar de Delattre (1966), Léon (1992) reconnaît le rôle joué par d'autres facteurs tels que la longueur du mot. De plus l'auteur souligne l'importance du rôle distinctif que la liaison joue dans certaines paires minimales (ex. « les héros » *versus* « les zéros ») ou dans certains contextes où elle constitue la seule marque qui permet de distinguer le pluriel du singulier (ex. « il aime » *versus* « ils aiment »). Finalement, Léon (1992) rappelle, comme l'avait fait son prédécesseur, le rôle phonostylistique joué par la liaison, soulignant sa fonction identificatrice en tant qu'indice de parlers conservateurs ainsi que sa fonction impressive, considérant la liaison plus comme une marque de style que comme un indice d'appartenance à une classe sociale.

Dans l'ouvrage conçu comme aide-mémoire d'orthoépie, Léon (1966) établit deux règles de prononciation de la liaison liées à l'accentuation des mots à l'intérieur du groupe rythmique. Dans la première règle l'auteur signale que la liaison ne doit pas être réalisée après un mot accentué (ex. « Il est grand | aussi ») alors que dans la deuxième il établit que la liaison doit être produite d'un mot inaccentué à un mot accentué, c'est-à-dire à l'intérieur d'un groupe rythmique. Ensuite, au sein des liaisons produites de manière variable, l'auteur suggère de définir le degré de possibilité de réalisation de la liaison (« liaisons peu fréquentes / normales »

et « liaisons anormales / possibles ») sur la base du style de parole (conversation familière, conversation soignée, style oratoire et style poétique).

Les études de Léon (1992, 1966) et surtout celles de Delattre (1966) constitueront les bases de nombreuses hypothèses qui seront reprises et développées par les travaux successifs.

Toutefois, comme le souligne Encrevé (1988), bien que les travaux de Delattre (1966) et Léon (1992, 1966) aient le mérite d'avoir fait avancer l'étude de la liaison vers une analyse basée sur une pluralité de facteurs (qui seront source d'inspiration pour toutes les études à venir), ils se heurtent aux limites d'une étude privée d'une base empirique solide.

> « Delattre ne dit rien de ses sources mais elles semblent du même ordre (ainsi ses exemples d'homophonie sont tirés directement du manuel de Nyrop, paru en 1902), même si l'intention descriptive est plus nette et la variation, à l'intérieur du ton désigné, plus précisément observée ». (Encrevé 1988, p. 44)

À partir de la deuxième moitié du XXème, de premières analyses quantitatives seront réalisées grâce à la constitution de corpus oraux. C'est le cas de l'étude de Ågren (1973) correspondant à 34 extraits d'émissions radiophoniques (« Tribune de Paris » et « Club des jeunes ») enregistrées entre les années 1960 et 1961. Le corpus compte 40 heures d'enregistrements d'entretien, de discussions et de présentations, dont les locuteurs étaient principalement des journalistes et des hommes politiques. Il en va de même pour l'étude menée par De Jong (1988) dix ans après. La ressource orale sur laquelle il s'appuie est le corpus d'Orléans ESLO1 (Michel Blanc et Biggs 1971), dont il annote et étudie 16000 contextes de liaison.

En particulier, il analyse les entretiens semi-formels de 45 locuteurs repartis entre hommes et femmes, en trois tranches d'âge et en cinq groupes selon les facteurs socioéconomiques et socioculturels.

Quant aux travaux conduits par Malécot (1975), les données ont été extraites de « 50 enregistrements de conversations avec des Parisiens cultivés » réalisés à leur insu, qui ont été étudiés quantitativement en prenant en compte à la fois les facteurs syntaxiques, phonologiques et sociolinguistiques.

En revanche, l'étude de Lucci (1983) propose d'observer systématiquement la variabilité de la liaison selon différentes situations communicatives et socio-culturelles du français parlé. Cependant, l'auteur semble faire coïncider le paramètre socio-culturel avec le type de situation communicative. Il associe respectivement le style de l'interview aux locuteurs qu'il définit comme « intellectuels » et la conversation familière aux locuteurs ayant un niveau faible d'instruction.

> « Nous avons […] observé le fonctionnement des liaisons dites 'facultatives', de manière comparative, dans plusieurs types situationnels ou socio-culturels de français oral, tels que : la conférence, la lecture, l'interview (caractéristique d'intellectuels) ou la conversation familière (locuteurs peu instruits) ». (Lucci 1983, p. 236)

En effet Lucci (1983) précise que les deux variables (situationnelle et socio-culturelle) sont liées « car l'acte de la conférence ou de lecture à voix haute n'est familier qu'aux locuteurs les plus scolarisés » (Lucci 1983, p. 259).

L'étude d'Encrevé (1988), contrairement aux précédentes, s'appuie sur un corpus très spécifique qui regroupe les discours enregistrés de 21 hommes politiques français entre 1978 et 1981. L'ensemble du corpus compte 10816 sites de liaisons potentielles, à partir desquels Encrevé (1988) mène une analyse sociolinguistique visant à mettre en relation les origines socioculturelles des locuteurs et le type de réalisation de la liaison.

Ces travaux sont les précurseurs des études plus systématiques de la liaison qui, comme nous le verrons dans les pages qui suivent, aboutiront à la constitution d'une large base de données pour l'analyse de la liaison. Ces études inscrivent dans une démarche de collection des données qui se distingue de la méthodologie des études normatives basées sur la référence à des exemples le plus souvent construits par introspection. L'opposition entre ces deux approches en linguistique, comme nous le verrons dans le cinquième chapitre (cf. Partie 1.Chapitre 5.3.1), amènera Laks (2008) à distinguer la linguistique du *datum* (à savoir la linguistique du corpus) de la linguistique de l'*exemplum* « basée sur un petit nombre d'exemples, plus au moins récurrents, dont les principes de construction et de recollection restent scientifiquement non fondés ».

2.2.2. Etudes systématiques de liaison

Bien que les études quantitatives menées au cours de la deuxième moitié du XXème siècle aient joué un rôle déterminant dans l'analyse de la liaison, elles souffrent d'une certaine limitation dans la collecte des données. En effet, comme le font observer Eychenne et al. (2014), ces travaux ne traitent qu'une partie sociolinguistique de la population (journalistes, hommes politiques, bourgeoisie cultivée de Paris etc.) et souvent un seul type d'interaction communicative (conversation radiophonique, discours publique et conversation familiale).

> « Les enregistrements d'Ågren portent sur des entretiens radiophoniques, ce qui, par définition, nous prive de renseignements précieux sur l'origine et le portrait socio-économique du locuteur; le corpus de De Jong ne garantit pas non plus un ancrage profond et durable dans la ville d'Orléans ; et celui d'Encrevé vise une population très particulière dont le discours n'est certainement pas caractéristique de la population dans sa totalité ». (Eychenne et al. 2014, p. 40)

C'est à la suite de ces considérations et d'une volonté de développer une étude plus systématique de la liaison que le programme « Phonologie du Français Contemporain » (dorénavant PFC) (Durand et al. 2009) (cf. Partie 2.Chapitre 1) se met en place au XXIème siècle. Ce projet, bien qu'il ne puisse pas être défini comme écologique (puisqu'il ne considère que certaines des situations de communication possibles), a pour objectif de constituer une base de données empirique qui représente un corpus de référence du français. En effet, le projet PFC, à l'aide d'un protocole d'enquête unique, a été en mesure de récolter des données de français oral bien plus riches que celles des études précédentes. Le corpus comprend à ce jour 38 points d'enquêtes dans l'espace francophone, soit 396 locuteurs dont les enregistrements ont été transcrits et codés. Dans le sillage des études sociolinguistiques basées sur les méthodes des réseaux denses (Milroy 1980), pour chaque point d'enquête une dizaine de locuteurs, (équilibrés entre hommes et femmes et repartis en trois tranches d'âge) ont été enregistrés et soumis à quatre types différents de tâches : deux tâches de lecture et deux entretiens (une conversation semi-dirigée et une conversation libre)[69]. L'ensemble de ces données recueillies, transcrites et codées (selon un système de codage n'intégrant aucun *a priori* théorique) correspondent à 53561 sites de liaison potentielle codés. Les analyses menées à partir de cette ressource ont donc permis de valider ou d'infirmer sur une plus grande échelle certaines des tendances observées par les premières études quantitatives menées sur la liaison. Notamment, comme nous l'avons vu dans le chapitre précédent (cf. Partie 1.Chapitre 2.2.6), si d'une part le corpus PFC a permis de confirmer l'importance de la variation sociolinguistique, de l'autre part il a remis en discussion la pertinence dans la réalisation de la liaison de facteurs tels que le sexe ou le niveau d'études (Eychenne et al. 2014). De plus, comme nous le verrons dans le chapitre suivant (cf. Partie 1.Chapitre 3.2.2), l'analyse de la liaison à partir d'une plus large base empirique a conduit à la révision des taxinomies établies par les études précédentes et plus en général à une redéfinition de la notion de variabilité de la liaison.

3. Conclusion

La définition d'une norme de prononciation à utiliser dans l'analyse des usages de la liaison est loin d'être anodine, d'autant plus dans le cadre d'un travail ayant pour objectif l'étude de l'acquisition de la liaison en français langue étrangère. Le bref parcours des modèles de prononciation adoptés au cours du XXème siècle nous a permis de mettre en évidence que la

[69] Nous nous attacherons à décrire le projet PFC et sa méthodologie dans la deuxième partie de ce travail (cf, Partie 2.Chapitre 1.1.3).

norme de prononciation a évolué en fonction des différentes valeurs et groupes sociaux dominants suivant un processus de « démocratisation» et « décentration » (Borrell et Billières 1989). Néanmoins à la suite de ces changements, qui ont imposé une norme socialement et géographiquement plus large, les caractéristiques phonético-phonologiques des modèles de prononciation ainsi que les groupes sociaux de référence sont difficilement identifiables (Morin 2000).

Comme la norme de référence, les approches méthodologiques ont connu une évolution d'une tradition normative (Grammont 1914; Fouché 1959) d'étude de la liaison caractérisée par le recours à des données issues de l'introspection du linguiste, au développement de travaux s'appuyant sur des résultats d'enquêtes (Ågren1973; Laks 1977; Morin et Kaye 1982; Lucci 1983; De Jong 1988; Encrevé 1988) de plus en plus systématiques (Durand, Laks et Lyche 2009). L'apport de ces recherches empiriques a contribué au renouvèlement des descriptions et des traitements classificatoires de la liaison (Durand et Lyche 2008; Mallet 2008; Durand et al. 2011) auxquels le prochain chapitre sera consacré.

Chapitre 4. Typologies et classifications de la liaison

Comme nous l'avons vu dans le chapitre précédent la liaison a fait l'objet de plusieurs travaux ayant recours à des approches méthodologiques et à des modèles de prononciation différents. Les descriptions de la liaison ont donc été soumises à plusieurs révisions en fonction des méthodologies et des normes de prononciation retenues.

Dans ce chapitre nous nous attacherons à passer en revue les principaux traitements classificatoires de la liaison ainsi que les limites explicatives (Mallet 2008 ; Côté 2013 ; Barreca 2014) de ces descriptions qui ne parviennent pas à rendre compte des comportements hétérogènes de la liaison.

1. Types et tokens de liaison

Le contexte de réalisation de la liaison, comme nous l'avons précisé dans les chapitres précédents (cf. Partie 1.Chapitre 1.2.3), est constitué d'une séquence de deux mots (que nous avons définis M1 et M2). La catégorie morphosyntaxique d'appartenance de ces mots permet d'identifier le « type »[70] de liaison comme par exemple la liaison entre un déterminant (DET) et le nom suivant (NOM). Ces généralisations permettent de regrouper l'ensemble des occurrences de liaison (« tokens ») dont les M1 et M2 relèvent des mêmes catégories morphosyntaxiques (ex. « les amis », « les arbres » etc.)

Pour cette raison le type de liaison, comme nous le verrons dans les pages suivantes, a été adopté dans toute la littérature, comme le critère principal pour classifier les contextes de liaison. Cependant ce facteur n'est pas le seul à influencer la réalisation de la liaison. Divers paramètres linguistiques que nous avons évoqués dans les chapitres précédents (cf. Partie 1.Chapitre 2), affectent la fréquence de production de la liaison et déterminent une variabilité interne à chaque type de liaison.

De plus, comme nous l'avons vu dans le chapitre précédent (cf. Partie 1.Chapitre 3.2), la liaison a été l'objet de plusieurs travaux dont la diversité des approches méthodologiques et les modèles de prononciation de référence ont influencé les descriptions et les classifications que nous nous attacherons à étudier dans les pages qui suivent.

[70] Nous utiliserons ce terme en tant que synonyme de « contexte » de liaison constitué par les catégories morphosyntaxiques d'appartenance du M1 et du M2 (ex. déterminant + nom).

2. Tripartition des contextes de liaison

Malgré cette hétérogénéité des traitements classificatoires, tous les auteurs s'accordent à répartir les contextes de liaison en trois typologies : (a) les liaisons qui se réalisent toujours ; (b) celles dont la production est variable de façon plus au moins importante en fonction des facteurs que nous avons évoqués dans le deuxième chapitre (cf. Partie 1.Chapitre 2) ; (c) celles qui ne devraient jamais se produire quel que soient le contexte et le locuteur.

Cette tripartition a été originairement proposée par Delattre (1947) et sera en fait reprise dans toutes les études successives et même dans les travaux à caractère plus normatif qui n'ont pu que reconnaître la place fondamentale de la variation dans l'étude de ce phénomène.

> « La partition traditionnelle entre liaisons obligatoires, facultatives et interdites, reprise à juste titre par tous les phonologues modernes, témoigne bien que pour la liaison, même les tenants les plus stricts de l'homogénéité du bon usage n'ont pas pu renvoyer la variation à la ténèbre de la performance ou de l'agrammaticalité. Il a fallu lui reconnaître une place dans la compétence, et ce sont les liaisons facultatives qui ont principalement retenu l'attention ». (Encrevé 1988, p. 46)

Cette tripartition basée sur des arguments essentiellement syntaxiques a toujours été l'objet d'un consensus bien que les dénominations ainsi que le contenu des classes varient selon les auteurs.

2.1. Différentes dénominations

Les trois typologies de liaison que nous avons évoquées ont reçu deux différents ordres de dénominations. En effet selon les auteurs, nous pouvons rencontrer des alternances entre les termes : « liaison obligatoire » / « liaison catégorique » ; « liaison facultative » / « liaison variable » et « liaison interdite » / « liaison erratique ».

Le choix de l'une ou de l'autre terminologie révèle l'approche adoptée par les auteurs lors de leur analyse. En effet, comme le souligne Mallet (2008), la distinction entre ces deux classes de dénominations correspond à l'opposition entre les deux approches linguistiques différentes que nous avons évoquées dans le chapitre précédent (cf. Partie 1.Chapitre 3.2), à savoir les approches normatives et descriptives. Encrevé (1988) sera le premier à remettre en cause explicitement la terminologie adoptée par l'approche normative ainsi que sa méthodologie (cf. Partie 1.Chapitre 3.2.1).

> « Délaissant le normatif pour le descriptif, il faudrait en toute rigueur rebaptiser selon le lexique de Labov les liaisons obligatoires, facultatives et interdites en liaisons invariables (ou catégoriques), variables et erratiques. […] nous utiliserons les deux vocabulaires, mais, […],

> c'est toujours le second qu'il faut entendre dans le premier : ce que nous visons, c'est la réalité de la pratique actuelle, et non les prescriptions qui peuvent être faites dans les traités ». (Encrevé 1988, p. 33)

En ligne avec la linguistique variationniste, il propose de modifier non seulement les définitions de la liaison mais il adoptera également une approche méthodologique descriptive basée sur une observation directe « de la réalité de la pratique actuelle ». Dans les prochains paragraphes, nous nous attacherons à introduire brièvement ces trois classes dont le contenu a fait l'objet de plusieurs révisions.

2.2. Liaisons catégoriques

La plupart des auteurs s'accordent à reconnaître la validité de la classe des liaisons catégoriques en tant que liaisons invariablement réalisées.

> « Les liaisons obligatoires ne sont donc pas un lieu de variation linguistique, pas plus sociale que stylistique : tous les locuteurs les réalisent toujours, et les « fautes » sur la liaison (les liaisons « mal-t-à propos ») ne se produisent pas dans ce domaine ». (Encrevé 1988, p. 49)

Ce caractère régulier de la liaison relève, comme l'observent Durand et al. (2011), de la présence d' « un noyau relativement stable et étroit » de liaisons catégoriques. En effet, comme nous le verrons dans les pages qui suivent, ces auteurs, contrairement à d'autres travaux (Fouché 1959; Delattre 1947), restreignent considérablement le nombre de contextes de liaison catégorique. Autrement dit, le caractère obligatoire des liaisons catégoriques ne fait pas en réalité l'objet d'un consensus. En effet, le contenu de cette typologie a subi des modifications selon les modèles de prononciation retenus par les différents travaux. Un même type de liaison peut donc être traité comme catégorique ou variable d'une étude à l'autre.

2.3. Liaisons variables

Quant à la classe des liaisons variables, la pluralité de facteurs susceptibles d'influencer la réalisation de la liaison rend sa prédiction difficile (Boula de Mareüil, Adda-Decker et Gendner 2003; Fougeron et al. 2001) (cf. Paragraphe 4). Toute la complexité de l'analyse des liaisons variables réside donc dans la prévision de la possibilité plus ou moins forte que ces liaisons soient produites. Delattre (1966) illustre bien cet aspect qui constitue le nœud central de l'étude de la liaison.

> « Les liaisons dites facultatives sont celles que l'on peut aussi bien omettre sans pour cela dépasser les bornes du langage accepté […]. Il est tout aussi correct de dire 'vous ave(z) entendu' que 'vous ave zentendu', mais encore faut-il savoir dans quelles circonstances dire l'un ou non l'autre […]. L'art de faire ou d'omettre des liaisons facultatives permet de garder

ses distance à l'égard d'Untel, ou au contraire de briser la glace et de se faire rapidement un ami ». (Delattre 1966, p. 66)

2.4. Entre erreurs de performance et « fausses liaisons »

2.4.1. Liaisons catégoriques non réalisées

Les liaisons catégoriques ne constituent pas « un lieu de variation linguistique » et leur réalisation ne représente donc pas un indicateur socio-stylistique. Néanmoins, elles peuvent être occasionnellement non produites et ainsi affecter l'intelligibilité de l'énoncé. Comme le fait remarquer Mallet (2008), étant donné la nature obligatoire de la classe des liaisons catégoriques, un nombre restreint de travaux a été consacré aux productions agrammaticales résultant de la non-réalisation des liaisons catégoriques. Déjà Encrevé (1988) avait noté dans son corpus des non-réalisations de la liaison dans les séquences « les États européens », « trois ans » et « nous arrivons ». L'étude des données extraites du corpus PFC (Durand, Laks et Lyche 2009) montre également la présence de cas de liaisons catégoriques non réalisées, traités par Durand et al. (2011) comme des erreurs de performance n'ayant aucune incidence sur le fonctionnement général de la liaison catégorique.

> « Comme dans tout corpus de grande taille, on trouve quelques exemples plus surprenants de liaison catégorique non réalisée. Ce type d'erreur, que l'on dira de performance, ne conduit pas à remettre en cause le réglage de la liaison catégorique ». (Durand et al. 2011, p. 41)

2.4.2. Liaisons erratiques

Contrairement aux liaisons catégoriques non réalisées, une grande place a été accordée à la description des liaisons erratiques[71]. Cela pourrait s'expliquer par la forte insécurité linguistique qui accompagne la maîtrise de la variation de la liaison (Encrevé 1988). La conscience du risque de commettre des fautes justifie donc le besoin de définir une typologie de liaisons interdites. Comme le fait observer Encrevé (1988), Fouché consacre une plus grande partie de l'analyse de la liaison à la description des « cas où on ne fait pas la liaison » qu'à celle des cas où elle est réalisée. Parmi ces liaisons erratiques, que Gadet (1989, p. 56) qualifie de « liaisons par excès », on compte des liaisons qui ne devraient pas être faites (à savoir des liaisons entre deux groupes rythmiques différents, entre « termes disjoints » devant les citations ou après la conjonction « et ») ou des liaisons théoriquement possibles mais peu souhaitables (ex. « vous êtes trop homme de terre ») (Nyrop 1934). Les raisons qui justifient

[71] Ces contextes seront signalés par deux barres obliques positionnées entre les mots de liaison (ex. « plus // haut ».

l'agrammaticalité de ces liaisons sont donc multiples et vont des arguments rythmiques aux arguments euphoniques et étymologiques.

2.4.2.1 Raisons prosodiques : liaison erratique entre deux groupes rythmiques et après la conjonction « et »

Traditionnellement, on considère comme erratiques toutes les liaisons qui se réalisent en dehors du groupe rythmique (cf. Partie 1.Chapitre 5).

> « On lie dans l'intérieur d'un élément rythmique, on ne lie pas d'un élément rythmique au suivant, en d'autres termes : on lie d'une syllabe inaccentuée sur la suivante, on ne lie pas d'une syllabe accentuée ». (Grammont 1914, p. 129)

> « Dans tous les cas où la liaison est interdite, son absence marque une rupture accentuelle, intonative, une *limite* sémantique ou dans le cas du H aspiré, une trace de séparation étymologique ». (Léon 2007, p. 153)

En réalité, comme nous l'avons déjà observé, d'autres études (Encrevé 1988; Morin et Kaye 1982; Post 2000) ont montré que la liaison peut être également réalisée après une pause (temporelle ou intonationnelle). Il en résulte que sa production ne serait pas forcement conditionnée par les frontières des groupes accentuels. Néanmoins, celles-ci semblent effectivement bloquer la réalisation de la liaison dans certains contextes. Léon (2007) présente l'exemple de l'interrogation avec inversion, qui d'après l'auteur détermine la formation de deux groupes rythmiques qui empêche la réalisation de la liaison (ex. « A-t-on // essayé? »).

Il en va de même pour le type de liaison erratique après la conjonction « et ». Dans ce cas, Delattre (1947) (dans le cadre d'une approche didactique) suggère de ne pas isoler ce contexte comme une pure exception mais au contraire de présenter la conjonction « et » comme un élément qui « en réunissant deux groupes de mots, ne s'unit régulièrement ni à l'un ni à l'autre » (Delattre 1947, p. 148).

2.4.2.2 Raisons euphoniques

Il s'agit ici de liaisons qui sont théoriquement possibles mais qui pour des raisons d'euphonies ne sont pas produites afin d'éviter « le calembour 'zozo' » (Gadet 1989). Le risque des cacophonies peut donc conduire à éviter des liaisons dont la production pourrait produire des effets comiques (ex. «vous êtes trop homme de terre » (Nyrop 1934) ; « vous feriez mieux d'aller garder vos oies et vos veaux » Queneau cité par Gadet 1989).

2.4.2.3 Raisons « d'intelligibilité » : liaison erratique avec des « termes disjoints » ou des citations

Pour des raisons d'intelligibilité du discours, la liaison n'est pas produite avant une citation ou interjections (« des // ah ») et entre deux « termes disjoints » (Léon 2007, p. 154) comme par exemple « cent // onze ». Cette définition est employée par Léon (2007) pour expliquer l'absence de liaison dans cette séquence en comparaison à des unités telles que « dix-huit », « vingt-huit » à l'intérieur desquelles la liaison est toujours produite. Comme le fait justement remarquer Léon (2007) la basse fréquence de « cent // onze » pourrait justifier l'absence de production dans cette séquence[72].

2.4.2.4 Contraintes étymologiques : devant « h » aspiré

Tous les auteurs signalent la liaison devant M2 commençant par « h » aspiré comme un site de liaison erratique. Ce dernier est certainement le plus connu et appris mais également le plus arbitraire du point de vue du locuteur et encore plus de l'apprenant. La référence à une origine étymologique serait en fait le seul élément capable de bloquer la liaison dans ces séquences. En raison de leur méconnaissance des mots commençant par « h » aspiré, les locuteurs sont parfois amenés à réaliser des productions agrammaticales.

2.4.3. « Fausses liaisons » chez les natifs

Il en va autrement pour les formes de liaison traditionnellement appelées « fausses liaisons », « liaisons fautives », « pataquès » (Domergue 1805, p. 465)[73] ou encore « cuirs» et « velours » (cf. note 74) qui ont été longtemps stigmatisées par les grammaires normatives (Fouché 1959 ; Grevisse 1936). Parmi celles-ci nous pouvons distinguer des formes de liaison correspondant à : (a) des substitutions de la CL cible (ex. « gros [t] enfant »); (b) des insertions d'une CL non attendue et non présente dans la graphie (ex. « quat' [z] enfants ») (Desrochers 1994).

Les premières représentent des formes de liaison où la CL cible est remplacée par une autre consonne lorsque la liaison est produite. En revanche, les deuxièmes constituent des liaisons non attendues produites (le plus souvent avec une CL /z/ et /t/)[74] où les CL ne s'appuient

[72] Les « termes disjoints » ainsi que les liaisons erratiques devant des mots commençant par « h » aspiré sont traités par Delattre (1947, p. 152) comme des « liaisons spéciales qu'on classe généralement comme 'exceptions' ».

[73] Le mot « pataquès » relève de la légende décrite par Domergue (1805, p. 465) (cf. Partie 1.Chapitre 1.6).

[74] Dans la tradition prescriptive les mots « cuirs » et « velours » font respectivement référence aux erreurs de liaison en [t] et [z]. Il n'existe pas de termes spécifiques pour désigner les « fausses liaisons » associées à d'autres CL ce qui montre leur faible fréquence par rapport aux erreurs en [t] et [z] (Desrochers 1994).

sur aucune consonne graphique. Ces formes sont donc le résultat de l'insertion d'une CL épenthétique entre deux mots qui ne sont pas susceptibles d'être liaisonnés.

Bien que ces formes soient « fautives », comme l'observe Gadet (1989 ; 2007), on peut généralement comprendre leur comportement et rechercher les phénomènes (souvent de nature hypercorrective) qui sont à l'origine de leur production. C'est dans cette direction que vont les travaux de Desrochers (1994) et Gadet (1989 ; 2007) qui s'opposent à la stigmatisation de ces « fausses liaisons ». Ces auteurs accordent à ces formes un intérêt sociolinguistique et linguistique pouvant apporter des éléments d'analyse importants pour l'étude du statut des CL (cf. Partie 1.Chapitre 5.2). En particulier, Desrochers (1994) propose de ré-analyser l'ensemble des « fausses liaisons » comme le produit de généralisations morphologiques (« erreurs morphologiques »), d'analogies lexicales (« erreurs lexicales ») ou comme des formes où l'insertion de la CL inattendue serait conditionnée par la prégnance de mémoire (« erreurs à distance »).

Les « erreurs morphologiques » et « les erreurs lexicales » sont donc respectivement liées à des généralisations du marquage du pluriel (selon lesquelles la CL inattendue fonctionne comme un « préfixe autonome de pluriel » (Frei 1929, p. 119)[75] (« les chemins de fer [z] algériens ») et/ou à des régularisations du paradigme lexical (« cinq [z] hommes ») par analogie avec « deux hommes » relevant de la même catégorie.

En revanche « les erreurs à distance » (ex. « bien mieux [n] isolé ») (Côté 2005) correspondent à des formes de liaison non attendue où la CL est séparée du M1 auquel appartient la consonne graphique, par d'autres éléments lexicaux.

Ces formes « hypercorrectives » de réalisations peuvent relever, comme le souligne Gadet (1989, p. 56), d'un niveau d'analyse syntagmatique résultant d' « une prégnance discursive interne à la chaine » ou paradigmatique « par prégnance de mémoire » (comme dans le cas des « erreurs lexicales »).

L'analyse sur de grands corpus, comme le projet PFC, montre que ces formes sont statistiquement minoritaires et ne constituent que 0,053% de l'ensemble des liaisons attestées tous styles confondus. De plus, elles seraient plus limitées en parole semi-spontanée qu'en lecture (Mallet 2008, p. 182-185; Durand et al. 2011, p. 36). Les cas d'épenthèse (14 occurrences) seraient plus productifs que ceux de substitution de la CL (6 occurrences). Parmi ceux-ci la réalisation « cent [z] euros » semble être assez récurrente.

[75] Gadet (1989) souligne à juste titre comment ce phénomène serait à l'origine de la formation du verbe « zieuter » (« entre quat' [z] yeux »).

Sur la base de l'observation de la marginalité de ces phénomènes les auteurs suggèrent qu'en dehors d'un contexte de tension linguistique particulier le phénomène de la liaison reste régulier.

> « […] lorsque les sujets parlants sont placés dans des situations sans tension linguistique particulière, la liaison reste un phénomène régulier et les locuteurs ont à leur disposition des répertoires bien réglés. En revanche, la parole publique dans ses divers états (interviews radiophoniques, cours, etc.) crée des contextes de planification complexe et d'hétéro-surveillance accrue générateurs d'erreurs et d'écarts par rapport à la norme ». (Durand et al. 2011, p. 38)

Malgré leur marginalité, ces formes s'avèrent, comme nous le verrons dans le cinquième chapitre (cf. Partie 1.Chapitre 5.2), extrêmement précieuses pour l'étude du statut des CL[76].

3. Classifications

Dans le chapitre précédent (cf. Partie 1.Chapitre 5), nous avons brièvement retracé les approches méthodologiques et les modèles de prononciation qui ont autrefois caractérisé les études consacrées à la liaison.

Si dans un premier temps des études normatives (Grevisse 1936; Fouché 1959) basées sur les « connaissances linguistiques intimes » du locuteur natif ont imposé une approche prescriptive (qui renvoient le plus souvent à la norme de prononciation basée sur les usages de la bourgeoise parisienne), au fil des années 80 une méthodologie variationniste sera mise en place par des travaux à caractère descriptif (Delattre 1966; Encrevé 1988). Ceux-ci s'appuient sur des données issues des études de corpus (cf. Partie 1.Chapitre 3.2) et font référence à des modèles de prononciation qui ne se limitent plus aux usages d'un groupe social privilégié (cf. Partie 1.Chapitre 3.1.1). Ces travaux poseront les bases pour le développement dans les années 90 d'études plus systématiques de la liaison, issues de l'analyse des corpus linguistiques (Durand, Laks et Lyche 2009; Eychenne et al. 2014) (cf. Partie 1.Chapitre 3.2.2.2). À cette pluralité de méthodologies correspond, comme nous le verrons dans les pages qui suivent, une diversité des traitements de la liaison au sein des classifications.

Les premières études normatives menées par Grevisse (1936) et Fouché (1959) adoptent une approche prescriptive introduisant les premières taxinomies de la liaison, correspondant à des listes de types de liaisons définis sur la base des critères syntaxiques et stylistiques. Celles-

[76] En effet, si les « erreurs morphologiques » laissent supposer que la liaison fonctionne comme un préfixe morphologique (rattaché au M2) (Morin et Kaye 1982) en revanche les « erreurs à distance » constituent un argument solide pour des travaux qui défendent le statut épenthétique et donc autonome de la CL (Côté 2005, p. 71).

ci seront suivies par des classifications basées sur des études descriptives telles que celle de Delattre (1947 ; 1966) et Encrevé (1988) qui donneront lieu aux tripartitions traditionnelles de la liaison (obligatoire / catégorique ; facultative / variable ; et interdite / erratique) qui constituent encore aujourd'hui les références pour l'étude de la liaison. Ces classifications seront ensuite l'objet de révisions fondées sur des études quantitatives menées sur de grands corpus de français oral qui ont permis d'affiner la description de la liaison.

3.1. Classifications normatives

3.1.1. Grevisse (1936)

Le travail de Grevisse (1936) s'inscrit dans le courant des descriptions normatives. Cet auteur, avant que les définitions de la liaison (obligatoire / catégorique ; facultative / variable ; interdite / erratique) ne soient adoptées, distinguait déjà les liaisons obligatoires des celles « recommandées » et des cas où la liaison ne se fait pas.

Contrairement à Fouché (1959), il ne précise pas (sauf dans le cas des liaisons entre déterminant et nom) la classe morphosyntaxique du M2. Son analyse est donc moins fine par rapport à celle proposée par Fouché (1959) (cf. Paragraphe 3.1.2).

Parmi les contextes de liaisons qu'il définit comme obligatoires, on retrouve la liaison entre adjectif et nom. La liaison est également obligatoire entre les adverbes (de degré ou de négation) et : (a) un autre adverbe (« pas aujourd'hui » ; « plus ici ») ; (b) un verbe à l'infinitif (« ne jamais oublier ») ; (c) un adjectif (« tout entier » ; « trop heureux » et « assez ouvert »). Il en va de même pour les prépositions monosyllabiques qui lient de façon obligatoire (« dans un cas » ;« sous un arbre ») (Grevisse 1936, p. 52). La liaison après les formes verbales impersonnelles « il est » et « c'est » est également présentée comme obligatoire (« c'est évident » ; « c'est à voir »). Cependant, le caractère « obligatoire » de la liaison dans ces types de liaison est en réalité relatif dans la mesure où il découle d'un jugement purement normatif et d'un choix stylistique opéré par l'auteur. Ici en effet, contrairement aux études successives (Delattre 1966), la liaison obligatoire n'est pas encore celle qui est réalisée par tous les locuteurs, dans toutes les situations de communication et dont la non-réalisation détermine l'agrammaticalité de l'énoncé. La définition de la nature catégorique de la liaison s'appuie sur les usages des classes socio-culturelles les plus élevées présentées comme des normes de prononciation de référence à suivre. L'auteur est donc conscient de la variabilité à laquelle ces types de liaison sont sujets dans d'autres variétés du français. En effet Grevisse (1936) précise que seule une partie des liaisons définies comme « obligatoires » sont réalisées même dans la

« langue populaire ». Des pratiques relevant des variétés de prestige définissent donc le « bon usage » de la liaison.

> « Dans certains des cas envisagés dans a), même la langue populaire fait la liaison : 'les années', 'il nous entend', 'dit-il', 'nous avons', 'vis-à-vis', 'en Asie' ». (Grevisse 1936, p. 52)

En revanche, un nombre restreint de types de liaison appartiennent à la classe que l'auteur qualifie comme des « liaisons recommandées ». Parmi elles, on retrouve la liaison : (a) entre le verbe et le nom (ex. « je suis homme ») ou l'adjectif attribut (ex. « nous sommes heureux ») ; (b) entre les auxiliaires à la troisième personne et le participe passé (ex. « il est allé ») ; (c) après « quand » et « dont » (ex. « quand on voit »). Ces contextes sont donc le résultat d'une sélection des types de liaison variable dont la réalisation n'est dans ce cas pas prescrite mais simplement souhaitée. Quant aux types où la liaison ne se fait pas, Grevisse (1936) fait référence aux cas de pause ou de disjonctions après lesquels la liaison n'est pas réalisée.

Contrairement à l'étude de Fouché (1959) (cf. Paragraphe 3.1.2), la description proposée par Grevisse (1936) se veut plus synthétique et propose un traitement classificatoire et un schéma tripartite qu'on retrouvera dans les études suivantes.

3.1.2. Fouché (1959)

Bien que l'approche de Fouché soit normative, prescrivant les usages d'une « conversation soignée chez les Parisiens cultivés » et visant un public d'apprenants, l'auteur reconnaît la nécessité d'une prise en compte de la variation dans la description de la liaison. En effet il précise que le cas de la liaison a été le seul où il a été « absolument nécessaire » d'élargir l'étude à d'autres variétés stylistiques : la conversation courante et le style soutenu. L'auteur distingue donc deux registres : la « conversation familière » et une variété correspondant au style des lectures oralisées à voix haute, des déclamations ou des lectures à voix basse d'un texte plus au moins littéraire. À partir de ces deux styles, Fouché (1959) oppose deux typologies de liaison : « les cas où on fait la liaison » et les « cas où on ne fait pas la liaison ».

L'analyse menée par Fouché (1959) aboutit à une description des types de liaison où les classes morphosyntaxiques du M1 et du M2 sont le plus souvent précisées. Il en résulte une liste riche de règles, exceptions et remarques difficilement résumables dans une classification. Comme nous l'avons déjà souligné dans le chapitre précédent (cf. Partie 1.Chapitre 3.2.1), chaque présentation est accompagnée d'exemples qui correspondent le plus souvent à des séquences construites n'ayant qu'une valeur illustrative.

Cette étude pourrait donc se présenter comme une analyse fine des liaisons. Cependant, les mots de liaison appartenant à la même classe morphosyntaxique sont souvent regroupés

sous une même règle sans que des différences de comportements entre ces éléments ne soient mises en évidence. C'est le cas par exemple de l'étude de la liaison après des adverbes monosyllabiques et polysyllabiques pour laquelle Fouché (1959) propose un traitement unitaire, illustré par la règle ci-dessous[77].

> « [*on fait la liaison*] entre les adverbes *droit, beaucoup, fort, ne… pas, pis, plus, ne … plus, point (ne…point), puis, tant, tout, très* et un mot suivant ». (Fouché 1959, p. 475)

Il en va de même pour la description de la liaison après les prépositions. Bien que l'auteur signale scrupuleusement toute exception, il ne fait référence à aucune variation pouvant affecter la réalisation de la liaison après les prépositions polysyllabiques[78].

> « [*on fait la liaison*] entre les prépositions *après, chez, depuis, dès, durant, en, pendant, sans, sous, suivant* ou les locutions prépositives *d'après, chez* et un mot suivant ». (Fouché 1959, p. 475)

Le manque d'une différenciation de comportements entre les composants de ces classes morphosyntaxiques est probablement à attribuer d'une part à une plus grande attention et préoccupation de l'auteur pour l'étude des exceptions (ce qui expliquerait, tout au long de l'ouvrage, la présence massive de notes et remarques). D'autre part, nous ne pouvons pas exclure que cette absence soit liée à une volonté précise de l'auteur de limiter son analyse à la variété de la conversation soignée, attestée chez les Parisiens cultivés.

3.2. Classifications descriptives

3.2.1. Delattre (1947 ; 1966)

La tripartition de la liaison (entre liaison obligatoire / catégorique, facultative / variable et interdite / erratique)[79] a été originairement proposée par Delattre (1947) et sera ensuite utilisée dans toutes les études successives. Bien que cette description ne s'appuie pas encore sur des données empiriques, Delattre (1947) s'éloigne des approches purement normatives des études précédentes. D'une part l'analyse de la liaison catégorique sera animée par une volonté de décrire, non plus les usages d'une variété socio-culturelle de prestige, mais plutôt les contextes où la non-réalisation de la liaison pourrait donner lieu à la production d'énoncés agrammaticaux. Comme l'observe Encrevé (1988), grâce à Delattre (1947) la notion de

[77] Sauf dans le cas des adverbes « bien », « mieux » et « trop ».

[78] Cette différence de comportements sera plusieurs fois mise en évidence par les études successives (entre autres Delattre 1947 ; 1966 ; Encrevé 1988).

[79] En particulier Delattre (1947) introduit les termes de « liaison obligatoire », « liaison facultative » et « liaison interdite ». Nous continuerons à utiliser ici la terminologie proposée par Encrevé (1988) (« liaison catégorique », « liaison variable » et « liaison erratique ») que nous avons adoptée au début de ce chapitre et qui relève d'une approche moins normative.

« liaison obligatoire » a donc évolué désignant un contexte de liaison réalisée par tous les locuteurs toute classe sociale confondue.

D'autre part, l'analyse de la variation de la liaison ne sera plus abordée en termes de prescription ou recommandation. Delattre (1947) introduit le concept de « liaison facultative » voulant désigner avec ce terme, toute liaison « que l'on peut aussi bien faire qu'omettre sans pour cela dépasser les bornes du langage accepté » (Delattre 1955, p. 43). Cette approche innovatrice posera les bases pour des études successives de la liaison, et aboutira à une première analyse des « tendances » qui caractérisent la réalisation de la liaison (Delattre 1955). Dans ce contexte, Delattre (1955), comme nous l'avons vu dans le chapitre précédent (cf. Partie 1.Chapitre 3.2.2), développera une étude des facteurs (syntaxique, prosodique, stylistique, historique, phonétique etc.) pouvant influencer la réalisation de la liaison.

En effet, bien que l'auteur établisse une classification de la liaison basée sur un registre de « conversation soignée courante »[80] il met en garde le lecteur quant au caractère absolu de cette tripartition.

> « Il est évident que les distinctions *obligatoires*, *facultatives*, *interdites* ne sont pas absolues. Elles varient selon le style. Dans les exemples qui suivent, elles se rapportent, autant que possible, au style de la *conversation soignée courante*, style encore variable selon les individus et le milieu où ils se trouvent ». (Delattre 1947, p. 152)

À partir de ces typologies, Delattre (1947) distingue quatre types de liaison selon la catégorie morphosyntaxique des unités auxquelles « se rapportent » les liaisons (cf. Figure 1): (a) le nom ; (b) le verbe ; (c) les mots invariables (prépositions et adverbes) ; (d) et finalement une catégorie de « liaisons spéciales qu'on classe généralement comme 'exceptions'». Ce dernier type comprend les cas des liaisons catégoriques à l'intérieur des groupes figés (cf. Tableau 1, Partie 1.Chapitre 2.2.3).

[80] Delattre (1947) contrairement à ses prédécesseurs, précise que cette conversation soignée constitue un style qui est toutefois « variable selon les individus et le milieu où ils se trouvent » (Delattre 1947, p. 152).

TABLEAU SIMPLIFIÉ

	OBLIGATOIRES	FACULTATIVES	INTERDITES
NOM	*déterminatif* + {*nom* / *pronom* / *adjectif*} vos enfants deux autres un ancien ami	*nom pluriel* + des soldats anglais ses plans ont réussi	*nom singulier* + un soldat anglais son plan a réussi
VERBE	*pronom personnel* + *verbe* ils ont compris nous en avons *verbe* + *pronom personnel* ont-ils compris allons-y	*verbe* + je vais essayer j'avais entendu dire vous êtes invité il commençait à lire	
INVARIABLES	*invariables monosyllabiques* + en une journée très intéressant	*invariables polysyllabiques* + pendant un jour toujours utile	*et* + et on l'a fait
SPÉCIALES	*formes figées* comment allez-vous les Etats-Unis accent aigu tout à coup de temps en temps		*h aspiré* des héros en haut + *un, huit, onze et dérivés* la cent huitième en onze jours

Figure 1: Tableau simplifié de la classification de Delattre (1947 ; 1966)

Toutefois, le critère d'appartenance à la catégorie morphosyntaxique ne constitue pas, contrairement aux descriptions précédentes, le seul critère de classification. L'analyse de la liaison prend également en compte des variables syntaxiques telles que la flexion des mots (singulier, pluriel et invariable), la fonction syntaxique (sujet, complément par exemple dans les cas des pronoms) ainsi que des facteurs lexicaux comme par exemple le nombre de syllabes du M1 (monosyllabique ou polysyllabique) (voir tableau détaillé Delattre 1947, p. 153-157).

Cette grille tripartie constitue une référence pour les classifications proposées par les études successives (Ågren 1973; Morin et Kaye 1982; Encrevé 1988; Mallet 2008; Durand et Lyche 2008; Durand et al. 2011) qui en ont modifié ou confirmé le contenu à partir de données empiriques.

3.2.2. Léon (1966)

Dans une démarche didactique[81] Léon (1966) propose deux règles de nature prosodique selon lesquelles la liaison ne se réalise pas en fin de groupe rythmique (ex. « il est grand //aussi ») (règle 1) mais entre un mot inaccentué et un mot accentué (à savoir à l'intérieur d'un groupe rythmique ex. « c'est un grand ami ») (règle 2). C'est dans le cadre de ces deux règles que Léon (1966) à la suite de Delattre (1947) fait la différence entre liaisons obligatoires, facultatives et interdites en soulignant que les premières correspondent uniquement à celles qu'on « fait toujours quel que soit le style parlé » (Léon 1966, p. 119).

> « La fréquence de rencontre de deux mots ou catégories grammaticales de mots a déterminé, dans le cadre large de deux règles ci-dessus, 3 types essentiels de liaison […] ». (Léon 1966, p. 119)

En effet Léon semble accorder la priorité aux facteurs de nature prosodique lorsqu'il propose de considérer que toutes les liaisons obligatoires « dépendent de la règle II » (cf. règle 2 ci-dessus) et qu'elles se réalisent de façon générale entre un « déterminant » (article, adjectif, auxiliaire, pronom, à savoir « mots inaccentuées ») et un « déterminé » (nom, pronom, formes pleines du verbe à savoir « mots accentués ») (Léon 1966, p. 120).

3.2.3. Encrevé (1988)

Encrevé (1988) conserve la tripartition traditionnelle établie par Delattre (1947 ; 1966) qu'il considère comme pertinente. Toutefois, il souligne que cette taxinomie nécessite quelques modifications concernant à la fois les définitions et le contenu de la classification. En effet, comme nous l'avons déjà évoqué (cf. Partie 1.Chapitre 3.2.2.1), d'une part Encrevé (1988) critique la terminologie adoptée par Delattre (1966) qu'il considère comme trop normative et lui préfère des définitions plus descriptives telles que liaison invariable (ou catégorique), variable et erratique. D'autre part, Encrevé (1988) souligne que les résultats des études quantitatives menées par Ågren (1973) et Malécot (1977) et Laks (1977)[82] invitent à modifier partiellement le contenu de la classification et la rendre ainsi plus conforme aux usages réels.

> « Pratiquement négligées par les phonologues, ces enquêtes présentent pourtant une information plus sûre, plus précise et plus récente que les ouvrages de référence : la prise en compte des résultats de Ågren et de Malécot paraît indispensable par exemple pour le classement en liaisons obligatoires et liaisons facultatives ». (Encrevé, 1988 : 45)

[81] Nous reviendrons sur cet aspect de l'ouvrage de Léon (1966) dans la troisème partie de ce travail (cf. Partie 3.Chapitre 3).

[82] Qu'il considère comme « indispensables » bien qu'ils soient souvent négligés par les phonologues.

Le caractère « obligatoire » des liaisons catégoriques est donc ici renforcé et validé par les résultats des travaux empiriques.

> « […] tous les auteurs s'accordent pour considérer qu'elle [*la tripartition de Delattre révisée*] caractérise l'usage de toutes les variétés de locuteurs : pour la traiter comme un fait grammatical représentatif de toute la communauté française. Les liaisons obligatoires sont, de l'avis général, invariablement réalisées par tous les locuteurs ». (Encrevé 1988, p. 48)

Les résultats de ces études empiriques ont conduit d'une part à la révision de deux types de liaison (originairement considérés comme catégorique par Delattre 1947) : (a) après adverbes monosyllabiques ; (b) après prépositions monosyllabiques. D'autre part ils ont posé les bases pour le développement d'une analyse plus fine de la liaison après les adjectifs épithètes antéposés au nom.

3.2.3.1 La variabilité de la liaison après les mots invariables (adverbes et prépositions) monosyllabiques

Encrevé (1988), s'appuyant sur les résultats des études de Malécot (1977) et Ågren (1973), intègre le type de liaison après les adverbes monosyllabiques à la classe des liaisons variables. En effet, les résultats de l'enquête de Malécot (1977) indiquent que des adverbes comme « très » et « plus » présentent un taux de réalisation de la liaison de 63%. La fréquence de liaison serait plus faible après l'adverbe « pas » (37%). Le résultats du corpus de Encrevé (1988) recueillis à partir de l'analyse des débats télévisés entre Giscard d'Estaing et Mitterrand, vont également dans cette direction et montrent la présence d'un grand écart de réalisation de liaison entre les adverbes monosyllabiques.

> « […] V. Giscard d'Estaing ne réalise que 16,7% (1974) et 37% (1981) de liaisons après pas, F. Mitterrand 63,2% (1974) et 91,3% (1981) ; la liaison après *plus* (adverbe de manière) est réalisée à 100% par F. Mitterrand (1974 et 1981) mais seulement à 66,6 %(1974) et 50% (1981) par son interlocuteur ». (Encrevé 1988, p. 64)

Quant à la fréquence de réalisation de la liaison après les prépositions monosyllabiques, elle semble être extrêmement variable d'une étude à l'autre. L'enquête d'Ashby (1981) signale que la liaison est produite dans ce contexte dans 49% des cas alors que Malécot (1977) indique un taux de réalisation beaucoup plus important (92%). En revanche, les résultats du corpus d'Encrevé (1988) ne concernent que les prépositions « dans » et « sans » après lesquelles la liaison est toujours réalisée.

Bien que ces données paraissent discordantes, elles démontrent, contrairement à la description proposée par Delattre (1947), l'hétérogénéité des comportements associés à la

liaison après les mots invariables monosyllabiques mettant en avant le poids d'autres facteurs[83] dans la classification de la liaison.

Encrevé (1988) s'interroge sur les causes qui peuvent être à l'origine de cette révision qui a conduit à une restriction des types de liaison catégorique initialement définis par Delattre (1947).

> « Soit que la position de Delattre ait été descriptivement inexacte, soit que la situation ait évolué, le domaine des liaisons invariablement réalisées par les « locuteurs cultivés » s'est restreint: les mots invariables monosyllabiques ont rejoint les mots invariables polysyllabiques dans la catégorie des liaisons réalisées variablement ». (Encrevé 1988, p. 48)

En effet il précise dans une note que « si modification il y a eu, c'est chez les locuteurs cultivés dans la mesure où Ågren et Malécot ont observé chez eux un domaine de liaisons obligatoires plus restreint que celui que décrivait Delattre » (Encrevé 1988, p. 54).

Il est aussi probable, comme le prétend Encrevé (1988), que la description initiale de Delattre (1947) était incorrecte étant donné qu'elle ne s'appuyait sur aucune enquête expérimentale[84].

3.2.3.2 Premières données orales sur la liaison après adjectifs épithètes antéposés

Si Encrevé (1988) n'hésite pas à modifier le classement de la liaison après les mots invariables (adverbes et prépositions) monosyllabiques, il se limite en revanche à observer des « flottements » dans l'usage de la liaison après les adjectifs antéposés aux noms. En effet, au début de son travail de révision de la classification tripartite de Delattre (1947), il précise que : « la modification porte sur un seul point […] les mots invariables monosyllabiques ont rejoint les mots invariables polysyllabiques dans la catégorie des liaisons réalisées variablement. Pour le reste il n'y a pas de changements très notoires » (Encrevé 1988, p. 48).

Cependant, Encrevé (1988) pose les bases pour le développement d'une analyse plus précise de la variabilité de réalisation de la liaison après l'adjectif épithète antéposé (ex. « grand arbre ») qui prend en compte à la fois les facteurs lexicaux (fréquence de l'adjectif), syntaxiques (tendance à l'antéposition ou postposition de l'adjectif) et stylistiques (nombre d'adjectifs

[83] Nous reviendrons sur cet aspect dans la deuxième partie de ce travail (cf. Partie 2.Chapitre 2.2.2.2.8).

[84] En effet Encrevé (1988) fait observer que Delattre (1947) ne précise pas ses sources mais qu'elles semblent du même ordre que celles des études normatives (ex. exemples tirés du manuel de Nyrop paru en 1902), « même si l'intention descriptive est plus nette et la variation, à l'intérieur du ton désigné, plus précisément observée » (Encrevé 1988, p. 44).

susceptibles d'être antéposés selon le style). Ce type de liaison avait été traité par Delattre (1947) comme un contexte relevant de la catégorie des liaisons catégoriques[85].

À la suite des analyses quantitatives comme celle menée par Laks (1975) auprès d'un groupe de jeunes à Villejuif et des observations réalisées par Encrevé dans l'usage courant, Encrevé (1988) indique que dans ce contexte la liaison n'est pas toujours susceptible d'être réalisée.

> « Si, dans la langue 'surveillée' on peut antéposer en principe tous les adjectifs qualificatifs, il n'est pas sûr que l'obligation de liaison porte sur tous les adjectifs antéposés ». (Encrevé 1988, p. 49)

En particulier Encrevé (1988), et d'autres (entre autres Benzitoun 2014), s'interrogent sur les caractéristiques distributionnelles des adjectifs qualificatifs et ils en concluent que, contrairement au registre plus soutenu, « dans l'usage courant la liste des adjectifs masculins antéposables est limitée pour tous les locuteurs » (Encrevé 1988, p. 49).

Laks (1975) avait déjà remarqué cette tendance chez les adolescents de Villejuif auprès desquels il n'avait observé qu'un groupe très restreint d'adjectifs antéposés (« petit », « gros » et « grand »). Après ce nombre extrêmement limité d'adjectifs la liaison est toujours réalisée chez les adolescents de Villejuif.

À la suite de ces résultats Encrevé (1988 : 49) restreint la liste d'adjectifs susceptibles d'être antéposés dans l'usage courant aux lexèmes suivants :

- « grand », « petit », « gros » et « bon » ;
- les adjectifs se terminant à la pause par une voyelle nasale : « prochain », « ancien »;
- les adjectifs terminés dans l'orthographe par un « r » : « premier » et « dernier » ;
- les adjectifs à consonne finale fixe : « admirable », « cher » et « brave » ;
- et les adjectifs adoptant invariablement dans l'antéposition la forme féminine, avec consonne final fixe : « bel », « nouvel » et « vieil »[86].

L'ensemble de ces observations amène Encrevé (1988) à considérer que l'usage de liaison après l'adjectif antéposé est variable sauf pour les adjectifs « petit », « gros » et « grand » déjà étudiés par Laks (1975) et pour l'adjectif « bon ».

[85] Cependant, Delattre (1947) avait déjà fait observer que les adjectifs féminins pluriels antéposés ne manifestent pas obligatoirement la liaison.

[86] Nous nous attacherons à approfondir la question de la place de l'adjectif épithète à l'oral dans la deuxième partie de ce travail (cf. Partie 2.Chapitre 2.2.2.3).

> « […] on observe des flottements sur ce point dans l'usage, hormis pour les quelques adjectifs dont l'antéposition est quasi figée, ceux qu'on relève chez les adolescents de Villejuif, auxquels il faut ajouter 'bon' ». (Encrevé 1988, p. 49)

Ces conclusions seront ensuite reprises et développées par les travaux qui suivent (Durand et Lyche 2008; Mallet 2008; Bybee 2005) fondés sur des enquêtes plus systématiques et sur grande échelle qui conduiront à une révision définitive de ce type de liaison.

3.2.4. Durand et Lyche (2008)

L'étude menée par Durand et Lyche (2008), basée sur des données issues de l'analyse de dix points d'enquêtes différents du corpus PFC (Durand, Laks et Lyche 2009) a conduit à une ultérieure révision de la classification proposée par Delattre (1947). Les changements portent d'une part sur la classe des liaisons variables et d'autre part sur celle des liaisons erratiques.

3.2.4.1 La variabilité de la liaison après la construction impersonnelle « c'est »

Durand et Lyche (2008) semblent indiquer, contrairement à la description de Delattre (1947), que seules 30% des liaisons après la construction impersonnelle « c'est » sont réalisées. Les auteurs précisent que cette variation ne peut pas être attribuée à des variables d'ordre diastratique étant donné que « the same speaker shifting from one form to the next within the same conversation » (Durand et Lyche 2008).

3.2.4.2 Données supplémentaires pour l'étude de la liaison après adjectifs épithètes antéposés

Les données issues de l'analyse du corpus PFC, confirment l'hypothèse seulement évoquée par Encrevé (1988), d'après laquelle le contexte de liaison entre adjectif et nom manifeste une liaison variable. Cependant, les auteurs soulignent que la plupart des exemples correspondent aux séquences « grand émoi » et « grand honneur » issues de la lecture du texte du corpus PFC. En effet, à la suite des observations d'Encrevé (1988), les résultats de cette étude montrent l'influence exercée par les variables diaphasiques et diamésiques sur la variabilité des adjectifs susceptibles de se trouver en position antéposée au nom. Les auteurs font observer que le nombre d'adjectifs antéposés en parole semi-spontanée est très limité et encore plus restreint en contexte de liaison.

> « The two conversations fare poorly in providing abundant data on prenominal adjectives for the simple reason that *petit* overwhelmingly represents this particular context and that other adjectives occur only sporadically in a liaison environment […]. Prenominal adjectives appear in large numbers in the base, but only rarely in a liaison environment, and when they do, the

> liaison is not categorical. The adjective *gros* will serve to illustrate this point: we record 139 occurrences of *gros* in the base, but only 8 in a liaison context, equally distributed between liaison and no liaison ». (Durand et Lyche 2008)

Autrement dit, des adjectifs tels que « gros », « dernier » et « ancien» auraient tendance à apparaître fréquemment en position antéposée devant des mots à initiale consonantique mais très rarement dans des sites de liaison potentielle[87].

> « Most interestingly, although so-called elementary adjectives (in the terminology of traditional grammar) do occur regularly as expected, in a prenominal position, they rarely do so in a liaison environment. In addition to *gros*, already mentioned, we observe the same phenomenon with *ancien(s)*, *dernier(s)*, etc., frequent with slightly less than 100 occurrences each, but never in a potential liaison site. In other words, speakers seem to talk without difficulty about *un gros type*, *un gros chien*, but not about *un gros homme*, *un gros âne* ». (Durand et Lyche 2008)

L'ensemble de ces observations amènent Durand et Lyche (2008) à formuler l'hypothèse selon laquelle les locuteurs auraient tendance dans ce contexte à éviter de produire des séquences où l'antéposition de l'adjectif pourrait donner lieu à des liaisons peu fréquentes. Le manque de familiarité avec ces contextes rares représenterait une source de difficulté pour la réalisation de la liaison pouvant donner lieu à des productions agrammaticales[88].

> « The PFC base not only throws doubt on the categorical character of liaison in pronominal adjectives, but it suggests that speakers systematically avoid a situation where they will be compelled to make a decision concerning the presence or not of a liaison ». (Durand et Lyche (2008)

3.2.4.3 Données supplémentaires pour l'étude de la variabilité de la liaison après prépositions monosyllabiques

Durand et Lyche (2008), fournissent des données supplémentaires pour l'étude des comportements variables de la liaison après prépositions monosyllabiques, confirmant de la sorte la variabilité de ce contexte mais précisant que contrairement au reste de la classe, la préposition « en » manifeste une liaison catégorique.

> « Our own observations [...] offer certain precisions: en induces a nearly categorical liaison, and out of 1124 relevant occurrences, we note two examples only without liaison: *en// un quart d'heure*, *manqué beaucoup de confiance en// elle*. All other prepositions vary in their usage, although liaison instances are overwhelming: there is no liaison after *dans* in 5% of the occurrences, including fixed expressions like *dans// un sens* ». (Durand et Lyche 2008, p. 13)

[87] Nous reviendrons sur les caractéristiques distributionnelles de ces adjectifs dans la deuxième partie de ce travail (cf. Partie 2.Chapitre 2.2.2.3).

[88] C'est le cas par exemple des réalisations orthographiques de l'adjectif « grand ». « […] six speakers do not make the liaison and two pronounce a [d] instead of the expected [t]. Four of the six cases concern *grand émoi*, suggesting that the lack of familiarity with the construction impacts on the absence of liaison ».

De plus, selon ces résultats, la préposition « chez » présente une plus grande variation quand elle se trouve devant un mot lexical (ex. « chez un copain », « chez un patron ») ou d'un nom (ex. « chez Anne ») propre alors que la liaison serait toujours produite quand « chez » est suivi d'un pronom (ex. « chez elle ») (sans doute à cause de la dépendance prosodique qui lie ce dernier à la préposition).

3.2.4.4 La liaison entre noms pluriels et verbe comme contexte de liaison erratique

De plus, Durand et Lyche (2008) excluent de la classe des liaisons variables le type de liaison constitué par des noms pluriels suivis d'un verbe, originairement décrit par Delattre (1947). En effet, les résultats issus des données de lecture du corpus PFC (ex. « Quelques fanatiques auraient même entamé ») montrent l'absence totale de liaisons réalisées dans ce contexte. Cette position sera aussi partagée par Bonami, Boyé et Tseng (2005, p. 93) qui rappellent qu'il est « généralement admis que la liaison est impossible entre un sujet syntagmatique et le verbe ('les enfants arrivent') ».

3.2.5. Mallet (2008)

L'étude de Mallet (2008) issue de l'analyse des données du corpus PFC confirme de manière générale les révisions opérées par les travaux précédents qui apportent plus de détails sur la variabilité des comportements au sein des différents types de liaison. Pour des raisons techniques, liées au traitement du corpus, Mallet (2008) explique n'avoir considéré que cinq contextes de liaison: (a) après conjonctions monosyllabiques sauf « et » ; (b) après prépositions monosyllabiques ; (c) après prépositions polysyllabiques ; (d) après adverbes monosyllabiques ; (e) après les temps simples des verbes « être » et « avoir ».

D'après les résultats de cette étude, la liaison après les conjonctions monosyllabiques (« mais », « puis ») est variable mais le plus souvent non réalisée (taux de réalisation 0,35%).

Cependant, Mallet (2008) observe la présence de réalisations dans des expressions telles que « mais il est venu », « puis il est venu ».

Les résultats des liaisons après les prépositions monosyllabiques (« chez », « dans », « sans », « sous ») confirment les conclusions des études précédentes à savoir la présence d'un taux élevé de liaison (91,9 %) dont la réalisation reste pourtant variable.

En revanche la liaison après prépositions polysyllabiques (« après », « avant », « depuis » « devant » « pendant ») n'est jamais réalisée dans le corpus PFC.

Quant aux liaisons après adverbes monosyllabiques Mallet (2008) observe un grand écart de réalisations entre les adverbes (1,4% pour « pas », 42% pour « bien », 83% pour « tout »).

Bien qu'elle considère ces contextes comme relevant de la catégorie des liaisons variables, elle souligne la présence d'une forte variabilité des comportements de la liaison au sein de ce type de liaison. Cette hétérogénéité serait liée d'une part au nombre d'occurrences de chaque adverbe qui semble difficilement comparable (880 pour « pas » et 145 pour « très ») et d'autre part au fonctionnement syntaxique très différent qui caractérise ces adverbes « conduisant à des contextes de liaison qui ne manifestent pas forcément la même cohésion syntaxique » (Mallet 2008, p. 282)[89].

Pour terminer, la liaison après les formes fléchies des verbes « être » et « avoir » aux temps simples manifeste respectivement des taux de réalisation de 27% (pour le verbe « être ») et 3,5%.(dans le cas du verbe « avoir »). Après les formes du verbe « être » les taux de réalisation varient entre 8% et 45%. Les taux les plus faibles sont attestés après les formes de l'imparfait et les taux les plus importants après les formes du présent de l'indicatif (sauf dans le cas de la forme fléchie « sont »)[90].

Quant aux formes fléchies du verbe « avoir », les pourcentages de réalisation de la liaison sont tous très bas (compris entre 0% et 10%) à l'exception de la forme « avons » (20,8%).

3.2.6. Durand et al. (2011)

Durand et al. (2011), à la suite des études menées sur le corpus PFC (Durand, Laks et Lyche 2009) propose un bilan de la classification des types de liaisons catégoriques communs à tout l'espace francophone.

> « Pour l'heure, PFC permet de confirmer sur une grande échelle que la liaison, tout au moins dans l'espace francophone européen, est catégorique dans les contextes suivants :
>
> Contextes catégoriques typiques
>
> (i) Det + X à voyelle initiale en SN (un, les, des, mon, son, ton, mes, tes, ses) : les enfants, les autres enfants, un enfant, mon enfant, etc.
>
> (ii) Proclitiques (ils, elles, on, nous, vous, en) : on [n]en [n]avait parlé, il y en [n]a, etc.
>
> (iii) Enclitiques : De quoi parle-t-[t]on ? Comment dit-[t]on ? Encore faut-[t]il travailler, etc.
>
> (iv) Composés et syntagmes figés : tout-[t]à-fait, toujours est-[t]il, pot-[t]au-feu etc. ». (Durand et al. 2011, p. 115-116)

De plus, en revenant sur l'analyse de la liaison après adjectif épithète antéposé, ils précisent que ce type de liaison ne constitue que 0,64% de l'ensemble des liaisons réalisées. Les auteurs évoquent, à l'instar de Durand et Lyche (2008), l'instabilité de la réalisation de la

[89] Cet aspect sera abordé plus en détail dans la deuxième partie de ce travail (cf. Partie 2.Chapitre 2.2.2.2.5).

[90] À la fin du XIXème siècle Clédat (1980) partageait déjà les mêmes conclusions.

« Si l'on met à part les phrases interrogatives, le cas prévu dans notre règle IV (*la liaison après les verbes impératifs suivis d'un pronom ex. allez-y*), le verbe êtr' suivi d'un adjectif, et les deux auxiliaires suivis d'un participe, on peut dire que, dans le langage courant, les liaisons se font rarement ». (Clédat 1980, p. 56)

liaison dans ce contexte pour lequel des formes de productions orthographiques de la CL sont attestées.

> « Pour ‘grand émoi’, on note 208 liaisons en [t] correctes par rapport à la norme, mais aussi 41 non liaisons, 4 liaisons en [n] (‘grand[n]émoi’) et 3 liaisons en [d] (‘grand[d]émoi’) ! La séquence ‘grand honneur’ provoque, quant à elle, 241 liaisons enchaînées en [t], mais aussi 14 exemples sans liaison et, à nouveau, 1 liaison en [n] (‘grand[n]honneur’) et 1 liaison en [d] (‘grand[d]honneur’) ! Ces exemples montrent que, contrairement à ce qui a pu être affirmé, les locuteurs ne savent pas nécessairement comment traiter phonologiquement un adjectif pourtant aussi courant que ‘grand’ si la séquence liaisonnante ne leur est pas familière ». (Durand et al. 2011, p. 116)

4. Conclusion : hétérogénéité de la liaison et difficultés de prédiction des réalisations

À partir de l’étude classificatoire d’Encrevé (1988), les travaux qui ont suivi ont toujours mis en évidence la forte hétérogénéité qui caractérise le comportement de la liaison au sein d’un même type de liaison. Nous avons déjà évoqué les exemples de la liaison après les adjectifs épithètes antéposés (cf. Paragraphe 3.2.4.2) ou encore après les adverbes monosyllabiques (cf. Paragraphe 3.2.4.3). Dans ces types, la liaison n’est pas produite après tous les M1 relevant de la même catégorie morphosyntaxique. Il ne s’agit ici que de deux exemples de la variabilité qui, comme nous le verrons dans la deuxième partie de ce travail (cf. Partie 2.Chapitre 2.2.2.2), affecte tous les types de liaison variable.

Il en résulte que l’identification des types de liaison constitue, comme tout critère classificatoire, un moyen économique de représenter une réalité extrêmement complexe et hétérogène. En effet comme le souligne Mallet (2008) les classifications traditionnelles sont dépourvues de toute valeur explicative et ne permettent pas de rendre compte des différentes variables d’analyse de la liaison (cf. Chapitre 2).

> « Ces propositions [*les classifications*] n’ont […] aucune valeur explicative ; elles ne permettent pas non plus d’étudier conjointement les différentes dimensions mises en jeu dans la liaison, ni de déterminer quelle position occupe la dimension syntaxique dans une hiérarchisation de contraintes, si tant est qu’il existerait une hiérarchie entre les différentes dimensions ». (Mallet 2008, p. 64)

L’extrême variabilité de la liaison rend donc difficile un traitement classificatoire et surtout la prédiction de la réalisation de la liaison. Boula de Mareuil et al. (2003) ont évalué la capacité de prédiction de la réalisation de la liaison dans un corpus de lectures oralisées. Les résultats de cette étude confirment la validité des règles établies par la classification d’Encrevé

(1988) bien que les auteurs soulignent que les taux de réalisation de la liaison varient sensiblement en fonction des mots et des bigrammes des mots.

> « The measures obtained confirm most of a priori linguistic predictions, and allow a ranking of the rules proposed in the literature. Liaison rates vary significantly depending on word and word bigram identities. This confirms that liaison is strongly linked to both syntactic and lexical levels ». (Boula de Mareüil, Adda-Decker et Gendner 2003, p. 1332)

Il en va de même pour les résultats du travail mené par Fougeron et al. (2001) à partir d'un corpus de parole spontanée. Cette recherche montre qu'une analyse statistique des facteurs (syntaxiques, lexicaux et stylistiques) de liaison ne permet pas de prédire automatiquement « avec une précision suffisante » la probabilité de réalisation d'une liaison dans un type donné. La fréquence de réalisation de la liaison est en effet liée aux items lexicaux constituant les différents types de liaison.

> « […] il apparaît qu'une approche statistique globale telle qu'elle est exercée ici permet de mettre en évidence la contribution de certains facteurs sur la réalisation des liaisons. Cependant, elle ne permet pas de prédire avec une précision suffisante la probabilité d'occurrence d'une liaison dans un contexte donné. En effet, la fréquence de réalisation de la liaison semble fortement dépendre des items lexicaux impliqués. […] notre étude montre qu'une modélisation des liaisons basée sur ce seul critère [*syntaxique*] n'est pas adéquate au vu de la variabilité dans la réalisation des liaisons au sein même d'une même catégorie grammaticale ». (Fougeron et al. 2001)

Ces observations suggèrent que pour traiter de la liaison, il est nécessaire d'analyser très précisément chaque type mais également chaque « token » à savoir chaque occurrence de liaison. Comme le soulignent Durand et al. (2011) à la suite de Morin et Kaye (1982), le contexte syntaxique de réalisation de la liaison doit être défini catégorie par catégorie mais aussi mot à mot.

En résumé, bien que la définition des règles classificatoires basées sur les types de liaison constitue un moyen économique pour une description globale de la liaison, elle s'avère toutefois moins efficace pour la prédiction de la réalisation variable de la liaison. Ces conclusions montrent la nécessité de recourir à une analyse qui prend également en compte les occurrences de liaison (à savoir les « tokens ») afin de parvenir à une caractérisation plus précise de la variabilité de la liaison, ce qui sera l'objet de la deuxième partie de ce travail (cf. Partie 2.Chapitre 2). Avant de développer une étude empirique des usages de la liaison, il nous paraît pertinent de passer en revue les modèles phonologiques ainsi que les sources sur lesquelles ils s'appuient.

Chapitre 5. Traitement et statut de la consonne de liaison : approches linguistiques

Dans le chapitre précédent nous avons vu comment plusieurs descriptions et classements des types de liaison ont été proposés par des orthépistes et des phonéticiens dès le début du XXème siècle. En revanche, le développement de premières théorisations de la liaison sera beaucoup plus tardif et coïncidera avec la deuxième moitié du XXème siècle. De plus, bien que les descriptions et les traitements classificatoires de la liaison aient été renouvelés grâce à l'apport des études sur corpus, tous les modèles phonologiques n'ont pas eu recours à ces données empiriques.

Plusieurs statuts des CL ont été théorisés selon les différents modèles qui se sont succédé au fil des années et que nous nous attacherons à décrire dans ce chapitre. Le défi principal de toute modélisation de la liaison réside dans la difficulté à rendre compte, à travers une formalisation, du statut particulier des CL, ce qui oblige à distinguer ces dernières des consonnes finales fixes.

1. Consonnes de liaison et consonnes fixes : deux statuts et deux traitements ?

Comme nous l'avons évoqué dans le premier chapitre (cf. Partie 1.Chapitre 1.4), la définition de la CL nous oblige forcément à clarifier sa relation par rapport aux consonnes finales fixes. L'opposition n'est donc pas purement terminologique et a souvent été définie sur la base d'arguments phonétiques (Morin 1986; Morin 1992; Morin 2003; Tranel 1990) ; (Bonami et Boyé 2003; Bonami, Boyé et Tseng 2005), lexicaux (Fouché 1959; Léon 2007) ou morphosyntaxiques (Tranel 1990; Côté 2005).

Cependant, dans plusieurs modélisations (issues en particulier des premiers modèles génératifs), les CL reçoivent le même traitement que les consonnes fixes puisqu'elles sont analysées comme des consonnes finales resyllabées à l'initiale du mot suivant (M2). Néanmoins, bien que, comme nous l'avons souligné dans le premier chapitre (cf. Partie 1.Chapitre 1.2), la liaison et l'enchaînement constituent deux phénomènes de *sandhi* externe, ils ne peuvent pas recevoir le même traitement. La liaison, en tant que cas particulier de *sandhi*, possède des spécificités qui la caractérisent et qui demandent un traitement capable de rendre compte de cette différence. Cette démarche ne permet pas seulement de justifier l'existence de

liaisons non-enchaînées (ce qui était l'objectif d'Encrevé)[91], mais tout simplement de rendre compte de la variabilité de la réalisation de liaison. En effet, l'intérêt d'un traitement séparé de ces deux phénomènes réside dans la possibilité de distinguer la production ou la non production des consonnes finales fixes, qui a été l'objet des travaux détaillés des phonéticiens (Fouché 1959, p. 376-409; Léon 2007, p. 74-116), de la réalisation variable des CL.

Cet aspect constitue à notre avis le nœud central dans la théorisation de la liaison qui a pourtant été souvent négligé par plusieurs modèles de la phonologie générative. Cela a inévitablement conduit ces approches à un traitement unitaire des consonnes fixes et des CL et à la mise en place de divers réajustements (Milner 1967) nécessaires pour justifier la variabilité de réalisation de la liaison. En effet si la production des consonnes finales fixes, comme le dit le terme, peut être décrite comme un phénomène stable et statique qui se prête à un traitement par règles, la réalisation de la liaison est, quant à elle, fortement variable et soumise à divers conditionnements (phonologiques, syntaxiques, prosodiques, lexicaux et stylistiques). Cela revient à dire (comme nous aurons l'occasion de l'observer à plusieurs reprises dans ce chapitre) qu'une analyse purement syntaxique ou phonologique ne peut que se heurter à la difficulté de différencier les comportements des CL de ceux de toute autre consonne finale.

2. Statut des consonnes de liaison

L'évolution des modèles théoriques de description de la liaison a été caractérisée par le besoin (plus au moins explicite) de différencier, au moins dans les termes, les CL des consonnes fixes. Cette volonté s'est souvent traduite dans l'attribution d'un statut particulier de la CL qui lui a valu une pluralité de dénominations : « consonne muette », « consonne latente », « consonne flottante » etc.

En effet, comme nous l'avons déjà souligné, la plupart des modèles de la liaison inscrivent leurs analyses dans le cadre du processus général d'enchaînement qui détermine la resyllabation de la CL à l'initiale du mot suivant (M2). Néanmoins, une telle approche oblige à une prise de position quant au statut lexical de la CL (Côté 2005) à savoir la place (l'affiliation lexicale ou phonologique) de la CL au sein de la représentation phonologique.

[91] Et donc de séparer nettement la production de la CL de sa resyllabation (et donc son enchaînement). Une telle approche demande le développement d'un modèle phonologique *ad hoc* (Durand et al. 2011) qui attribue un statut spécial (flottant) au CL (cf. Partie 1.Chapitre 5.2.1.3) qui résulte être lexicalement marqué (Tranel 1995a, p. 801).

Nous reprenons ici la description proposée par Côté (2005; 2010) qui distingue cinq différents statuts de la CL décrits par les modèles d'analyse de la liaison couvrant « toute la distance entre le M1 et le M2 » (cf. Figure 2).

Principales options	**Sous-types**	**Représentation lexicale**
(1) CONSONNES FINALES : Les CL sont des consonnes finales du M1	a. ANALYSE PAR TRONCATION : Les CL sont des consonnes stables qui chutent dans des contextes de non liaison	/døz/ /ami/
	b. ANALYSE PAR SUPPLETION : Les CL sont des consonnes stables appartenant à des allomorphes distincts	/dø, døz/ /ami/
	c. ANALYSE AUTOSEGMENTALE : Les CL sont flottantes par rapport au squelette ou à la syllabe, avec une représentation distincte de celle des consonnes stables	/dø(z)/ /ami/
(2) CONSONNES EPENTHETIQUES : Les CL sont insérées par épenthèse		/dø/ /ami/
(3) CONSONNES INITIALES : Les CL sont des consonnes initiales du M2		/dø/ /zami/
(4) CONSONNES MORPHEMIQUES : Les CL correspondent à des morphèmes affixaux	a. ANALYSE PREFIXALE : Les CL sont des préfixes du M2	/dø/ /z + ami/
	b. ANALYSE SUFFIXALE : Les CL sont des suffixes du M2	/dø + z/ /ami/
(5) CONSONNES INTEGRÉES À DES CONSTRUCTIONS : Les CL font partie de constructions plus larges que le mot et (partiellement) lexicalisées		/dø z ami/

Figure 2 : Les statuts lexicaux des consonnes de liaison (CL) (Côté 2010, p. 1280)

Selon ces approches les CL sont analysées en tant que : consonnes finales (1) entièrement rattachées au M1 ; consonnes initiales (3) entièrement rattachées au M2 ; préfixes (4a) ou suffixes (4b) partiellement autonomes par rapport à leur mot d'attache ; consonnes épenthétiques (2) et donc indépendantes des mots de liaison ; et celles faisant partie de constructions plus larges que le mot (5), et dont elles dépendent.

2.1. Consonnes finales

Selon ce premier statut, le plus traditionnel parmi ceux que nous venons d'évoquer, la CL représente une consonne finale qui est rattachée au M1. Comme Côté (2005 ; 2010) le fait observer, cette approche s'est manifestée à travers trois types d'analyses de la CL (par

troncation, supplétion et flottance) qui définissent de manière différente le rapport entre celles-ci et les consonnes fixes.

2.1.1. Analyse par troncation

Les analyses basées sur la troncation (cf. Paragraphe 2.1.1) considèrent la CL comme une consonne stable. Le traitement de la liaison se propose ainsi non pas d'examiner les contextes où la liaison est produite mais de rendre compte de ceux où elle n'est pas réalisée à travers des processus d'effacement. Ces mécanismes sont basés sur une règle générale de troncation des consonnes finales en contexte de non liaison. Cela revient à dire que dans ces analyses les CL et les consonnes fixes reçoivent un traitement unique où des réajustements de nature différente interviennent pour justifier la réalisation ou non-réalisation de telle ou telle consonne fixe. Celles-ci ne seraient pas sujettes à la règle de troncation soit parce qu'elles sont suivies d'un *schwa*, soit parce qu'elles constituent des exceptions lexicales.

De plus, la non-séparation dans cette approche entre la description de la réalisation des CL et leur enchaînement conduira, comme le souligne Encrevé (1988), sous certaines formes (Schane 1965; Dell 1970), à des paradoxes tels que l'impossibilité de rendre compte à travers cette analyse des cas d'enchaînement des consonnes fixes.

Comme nous le verrons dans les pages qui suivent (cf. Paragraphe 3.1.1), la règle de troncation a été originairement proposée par Schane (1965).

2.1.2. Analyse par supplétion : allomorphie

Cette analyse considère la CL comme allomorphique. Deux représentations lexicales distinctes sont proposées : l'une correspondant à l'allomorphe de liaison (ex. /pətit/) ; et l'autre à l'allomorphe de non liaison (ex. /pəti/).

Contrairement à l'approche précédente, cette analyse permet de rendre compte (grâce à l'allomorphie) de la variabilité de réalisation des CL par rapport aux consonnes fixes.

Ces analyses supplétives ont été proposées dans le cadre de différents modèles théoriques : modèle génératif (Klausenburger 1984), théorie de l'Optimalité (Tranel 2000). Certains auteurs (Tranel 2000; Côté 2005) considèrent également la supplétion pour un groupe restreint d'adjectifs prénominaux (« bon » ; «ancien », « lointain » ; « plein » ; « prochain » ; « vain » et « vilain »), caractérisés par une modification de la voyelle dans le contexte de liaison et donc une alternance du type [õ] / [ɔn] ou [ɛ̃] / [ɛn].

> « First, *bon* is not an isolated case; there are number of similar adjectives which may occur in prenominal position with the same phonological properties (e.g. *ancien* 'former', *lointain* 'distant', *plein* 'full', *prochain* 'next', *vain* 'vain', *vilain* 'nasty'). By contrast, *sot* is in this respect

> fairly isolated ('idiot Stupid'* is the only other similar case, and it does not actually fit very well in prenominal position) ». (Tranel 1990, p. 189)

Ceux-ci posséderaient donc deux formes lexicales distinctes : d'une part, la forme de non-liaison, ayant une voyelle nasale [bɔ̃] ; et d'autre part, la forme de liaison, contenant une voyelle orale [bɔn]. Ce sous-groupe d'adjectifs serait donc traité par une analyse supplétive déjà traditionnellement employée pour les adjectifs « beau / bel », « nouveau / nouvel », « vieux / vieil », « ce / cet », « fou / fol » et « mou / mol » (Bonami et Boyé 2003; Bonami et Boyé 2005).

L'adoption d'un tel traitement pour l'étude de ces adjectifs revient à nier implicitement le processus de (dé)nasalisation vocalique. Autrement dit, si cette approche supplétive peut être également étendue à l'analyse de l'alternance entre la forme du masculin [bɔ̃] et la forme du féminin [bɔn], on postule l'existence d'une forme sous-jacente des voyelles nasales rejetant l'hypothèse de leur dérivation des voyelles orales.

> « Let us assume that both *bon* and *son* indeed have nasal vowels in their basic lexical entries, but that there is no minor rule of vowel denasalization in liaison contexts. […] The overall trend toward liaison with nasal, rather than oral, vowels can still be easily explained, this time by the expected elimination of suppletive forms from lexical entries. As to the variation observed in the speech of some speakers between liaison with an oral vowel and liaison with a nasal vowel (Encrevé 1988: 212), it can be straightforwardly interpreted as an intermediate stage characterized by an inconsistent use of the suppletive forms ». (Tranel 1990, p. 192)

Dans ce sens la variation (Encrevé 1988; Tranel 1990; Sampson 2001) dans l'utilisation des formes nasales ou orales de ces adjectifs dans des contextes de liaison pourrait être expliquée comme le résultat de la sélection variable entre les deux allomorphes[92].

D'autres travaux (Bonami et Boyé 2003; Bonami et Boyé 2005) ont élargi cette analyse supplétive à tous les adjectifs prénominaux postulant d'une part un paradigme à cinq cases (Bonami et Boyé 2003) et d'autre part un système à deux thèmes soumis à des règles implicationnelles (Bonami et Boyé 2005) (cf. Partie 2.Chapitre 2.2.2.3.1).

2.1.3. Analyses autosegmentales : la notion de flottance

Contrairement aux analyses précédentes, les CL ont dans les approches segmentales un statut instable qualifié de « flottant » par rapport au squelette (flottement squelettal) (Tranel 1990) ou à la syllabe (flottement syllabique) (Clements et Keyser 1983) ou à tous les deux à la

[92] Cependant, ce dernier aspect comme le souligne Côté (2005, p. 76) atteste une tendance à la régularisation du paradigme en faveur d'une voyelle nasale dans tous les contextes, ce qui semble suggérer une forme de liaison épenthétique plutôt que l'élimination de l'allomorphe spécifique de liaison de nature supplétive.

fois (Encrevé 1988). Leur réalisation se produit grâce à l'ancrage dans la structure prosodique, si une position d'attaque est disponible (à savoir si la condition syllabique est remplie).

> « Le segment flottant correspond à l'analyse selon laquelle la CL appartient lexicalement au mot mais doit avoir une représentation distincte de celle des consonnes toujours réalisées dans la prononciation : flottante, la CL dépend pour sa prononciation de l'intervention, ou non, d'une opération d'ancrage dans le squelette, ainsi que d'une opération d'interprétation syllabique de la position dans laquelle elle s'ancre ». (Encrevé 1988, p. 170)

Cette approche, tout en renvoyant à une seule représentation lexicale sous-jacente, contrairement à l'analyse par troncation, fournit un cadre pour la description de la variation de la liaison. Néanmoins une telle analyse attribue, comme le souligne Laks (2005b) à la suite de Tranel (1995a), un statut lexical marqué aux CL, les excluant de l'organisation syllabique ou gabaritique (sauf dans le cas où la condition syllabique serait remplie et la liaison serait donc réalisée) ce qui complexifie la représentation lexicale du mot.

2.2. Consonnes épenthétiques

Dans le cadre de cette analyse, la CL est considérée comme lexicalement indépendante du M1 et du M2 (Klausenburger 1974; Tranel 1981; Morin et Kaye 1982; Côté 2005)[93]. En réaction à l'analyse par troncation, cette approche attribue aux CL un statut lexical autonome par rapport à celui des autres consonnes[94]. En particulier, Tranel (1981) défend cette position en s'appuyant sur des données empiriques. De plus, comme nous l'avons déjà observé dans le chapitre précédent (cf. Partie 1.Chapitre 4.2.4), certains types de fausses liaisons (« erreurs à distance ») apportent des arguments en faveur de cette position[95]. Un autre argument semblerait venir également des résultats des études acoustiques menées auprès des sujets aphasiques (Kilani-Schoch 1983). En effet, Kilani-Schoch (1983) observe chez des patients la récurrence de substitutions des CL par la consonne [d] (ex. [ledetyd] pour « les études » et [tʀɛdamizɑ̃] pour « très amusant ») pour lesquelles une explication « en terme de contamination phonétique » est insuffisante puisqu'elle ne permet pas de couvrir tous les cas. L'auteur souligne également la présence de liaisons non attendues (ex. [dezpɔɛm] pour « des poèmes ») qui peuvent être

[93] Dans le cas du traitement des adjectifs singuliers, Morin et Kaye (1982) proposent de considérer les CL comme des consonnes épenthétiques alors que les CL appartenant aux adjectifs pluriels sont analysées comme des consonnes initiales (cf. Partie 1.Chapitre 5.2.3).

[94] Nous nous attacherons à décrire les détails de cette analyse dans le paragraphe (cf. Partie 1.Chapitre 5.2.2).

[95] Nous verrons dans les pages qui suivent (cf. Partie 1.Chapitre 5.2.2) comment des formes de réalisation attestées chez les enfants natifs ainsi que les résultats des analyses acoustiques semblent suggérer un statut autonome des CL.

interprétées comme « des applications exagérées de la règle (épenthétique) de liaison » (Kilani-Schoch 1983, p. 458).

2.3. Consonnes initiales

Cette approche est sans doute la moins défendue dans la période moderne par les études portant sur le traitement des CL et généralement limitée à l'analyse de quelques aspects spécifiques tels que la liaison avec les enclitiques (Morin 1979a ; Morin 1979b ; Côté 2005). En ce qui concerne l'analyse des enclitiques dans les constructions impératives ou avec inversion (ex. « vas-y » ; « dit-il »), les travaux menés par Morin (1979a ; 1979b ; 1986) et Côté (2005) suggèrent de considérer ceux-ci comme lexicalisés aux consonnes précédentes et donnant lieu à des représentations lexicales des enclitiques telles que /zi/ et /til/. La CL est donc traitée comme consonne initiale du M2 et ne représente pas à proprement parler une CL. Cette position semble être supportée par deux arguments venant de deux tendances du français laurentien observées par Côté (2005, p. 75) et qui montrent l'existence de caractéristiques phonétiques qui rapprochent les CL devant enclitiques plus des consonnes initiales que des autres CL. D'une part, dans les dialectes franco-ontariens les CL [t] dans des contextes d'inversion (ex. « dit-il ») sont sujettes à aspiration, tout comme les occlusives sourdes à l'initiale du mot (ex. « petit thème »). Ce n'est en revanche pas le cas des CL dans d'autres contextes de liaison (ex. « petit elfe »). D'autre part, les CL [t] appartenant à des constructions inversées sont elles aussi soumises obligatoirement, tout comme les mêmes consonnes à l'intérieur des mots (ex. « tu » [tˢy]), à un processus d'assibilation devant /i/, /y/, /j/ et /ɥ/qui reste en revanche variable aux frontières des mots (ex. « lutte inouïe » [lytinui] ou [lytˢinui]).

Ternes (1977) est le seul à élargir cette analyse à tous les types de liaison en proposant de considérer la CL comme morphologiquement intégrée au M2 à initiale vocalique (et non en tant que préfixe comme dans l'analyse préfixale). Le M2 connaîtrait donc plusieurs variantes (ex. « arbre » [arbʀ], [narbʀ], [tarbʀ] et [zarbʀ]) sélectionnées sur la base du contexte précédent à savoir du M1.

2.4. Consonnes morphémiques

2.4.1. Analyse préfixale

Cette analyse propose de réinterpréter les CL non pas comme intégrées lexicalement au M2 mais en tant que préfixes. Klausenburger (2001), Morin et Kaye (1982)[96] et Morin (2003)

[96] Dans le cas des adjectifs pluriels.

considèrent les CL en tant que préfixes des adjectifs prénominaux. La CL de l'adjectif « gros » dans l'énoncé [grozaʀbʀ] est donc traitée comme préfixe du M2.

Morin a également appliqué cette approche à des domaines plus larges tels que le marquage du pluriel (Gougenheim 1938; Morin et Kaye 1982; Morin 1986).

2.4.2. Analyse suffixale

Ce traitement est restreint à l'étude des adjectifs prénominaux (ex. « mauvais » [move]) où la CL est analysée par Morin (1992) comme un suffixe ayant une fonction de marque casuelle correspondant à une représentation du type /move + z/. Ce contexte de liaison se prête particulièrement à cette approche étant donné le nombre très limité d'adjectifs susceptibles de se trouver (du moins à l'oral) en position antéposée par rapport au nom (cf. Chapitre 4.3.2.4.2). Néanmoins, Morin (2003) reconnaît que ce type de modèle ne permet pas de distinguer le comportement des CL de celui des consonnes fixes dans le cas de non-enchaînement en contexte de rupture intonative.

> « One weakness of this analysis, however, is that it required *ad hoc* manipulations of the inflectional ending to have it surface as the initial consonant of a following noun whenever a pause would intervene between the two […] ». (Morin 2003, p. 395)

2.5. Consonnes intégrées à des constructions

Cette approche développée par Bybee (2001a ; 2001b ; 2005) s'inscrit dans la tradition des Grammaires de Construction (Goldberg 1995) et de la Grammaire Cognitive (Langacker 1987) et propose de traiter les CL comme indépendantes du M1 et M2 faisant partie de constructions plus larges. Celles-ci possèdent des degrés de cohésion qui varient en fonction de la fréquence de cooccurence (cf. Partie 1.Chapitre 2.2.4.1) qui détermine le niveau de lexicalisation de la séquence. Selon cette analyse donc, les séquences de mots employées fréquemment possèdent un fort degré de cohésion syntaxique et sont directement stockées et représentées dans la mémoire. Ces séquences fréquentes acquièrent ainsi une plus grande force lexicale qui facilite leur accessibilité et donc la réalisation de la liaison. La production de la liaison est le résultat direct de la mémorisation des séquences fréquentes sous forme de constructions.

> « Dans notre approche de la liaison, les unités de base sont les constructions. Puisqu'elles contiennent souvent du matériel lexical et grammatical très spécifique, nous attribuons la présence d'une consonne de liaison à la construction elle-même. Plus précisément, ces constructions sont des séquences récurrentes de morphèmes ou de mots qui sont reliés entre eux par des relations sémantiques, fonctionnelles ou statistiques particulières lorsqu'ils sont utilisés ensemble, relations qu'ils n'entretiennent pas forcément dans d'autres contextes. Les

> constructions sont plus ou moins conventionnelles, en fonction du degré auquel l'usage répété les a établies dans une langue particulière ». (Bybee 2005, p. 26)

Ces constructions prennent place tout au long d'un continuum dont on trouve à l'une des extrémités des expressions figées (ex. « c'est-à-dire »), tandis que les constructions ayant un schéma plus ouvert (ex. « préposition suivie d'un nom ») se situent plus près du centre. À l'autre extrémité du continuum, on trouve des constructions qui ont deux emplacements occupés par des items issus de classes ouvertes (ex. « nom pluriel suivi d'un adjectif »).

2.6. Un statut hétérogène

Il convient de préciser, comme nous le verrons dans les pages qui suivent, qu'à ces différents statuts de CL ne correspondent pas de modèles uniques de traitement de la liaison. En effet, comme le souligne Côté (2010), à l'exception des modèles génératifs basés sur l'étude de variables purement phonologiques, les traitements plus récents ont opté pour des analyses multidimensionnelles qui reconnaissent l'existence de CL ayant des fonctionnements et des statuts différents.

> « Alors que les approches phonologiques génératives ont le plus souvent conçu la liaison comme une processus unifié, les discussions plus récentes tendent à s'éloigner de cette conception unitaire et à œuvrer dans un cadre multidimensionnel. Toutes les consonnes de liaison n'ont pas à être soumises au même mécanisme et plusieurs reconnaissent l'existence de plusieurs catégories de consonnes de liaison: épenthétiques, supplétives, initiales et possiblement préfixales chez Côté (2005); finales, initiales et morphophonologiques dans Pagliano & Laks (2006); voir aussi Durand & Lyche (2008) ». (Côté 2010, p. 1280)

3. Traitements des consonnes de liaison

3.1. Les modèles de la phonologie générative linéaire

L'ensemble des études génératives sur la liaison s'appuient sur les travaux normatifs (Fouché 1959 ; Grevisse 1936) qui renvoient à une « connaissance personnelle de la phonétique française » du locuteur natif (Warnant 1987). En effet, comme le souligne Encrevé (1988), bien que ces auteurs aient eu la possibilité de faire référence à des travaux à caractère plus descriptif (Delattre 1947 ; Delattre 1955), ils paraissent leur préférer les exemples à valeur prescriptive (Fouché 1959 ; Grevisse 1936)[97] n'ayant jamais travaillé sur l'oral mais sur l'écrit littéraire. Ce passage par l'écrit, comme le note Encrevé (1988, p. 89), est parfaitement cohérent avec

[97] De plus, comme nous l'avons vu dans les chapitres précédents, Fouché (1959) affirme se rapporter à « au français qui se prononce dans la lecture à voix basse d'un texte ordinaire, sans prétention à la lecture à la littérature, ou, si l'on veut, dans une conversation sérieuse et soignée ».

l'approche orthoépique et didactique qui caractérise les ouvrages normatifs et qui a donc probablement attiré l'intérêt des linguistes génératifs, pour la plupart non francistes et non natifs. En effet, ce retour au « normativisme descriptif » (Morin et Paret 1983) s'expliquerait selon Morin (2000, p. 104) par la clarté descriptive de la norme du français qui assurerait une base empirique pour la formulation des généralisations théoriques, aspect explicitement reconnu par Anderson (1982).

> « […] the essentially conservative standard language defined approximately by such classical traditional treatments (as Fouché, Grammont, and Grevisse) ensures a sufficiently broad theoretical context ». (Anderson 1982, p. 535)

Ces travaux, que Laks (2008) qualifie de « linguistique de l'*exemplum* », s'inscrivent dans la tradition chomskienne qui fond ses analyses sur « des intuitions d'un locuteur-auditeur idéal et abstrait, appartenant à une communauté linguistique homogène qui apprend la langue instantanément et n'est affecté par aucune des limitations de la performance » (Chomsky 1965).

Comme le montre Laks (2008), c'est le refus même de l'observation des pratiques langagières comme objet d'étude qui aboutit au recours exclusif à l'usage de l'intuition du natif.

> « Chomsky lui-même, et toute la grammaire générative avec lui, sont très clairs sur ce point. Ce qui justifie le recours unique aux intuitions et aux jugements de grammaticalité c'est la conception même de la grammaire qui est la sienne: dès que l'on sépare définitivement le langage externe manifesté (*E-language*) du langage interne actualisant la faculté innée de langage (*I-langage*), le recours aux données externes n'a plus d'intérêt ». (Laks 2008)

En effet dans la perspective générative l'utilisation du *datum* à savoir le corpus linguistique n'a pas de valeur objective et « représente la réalité de celui qui l'a construit » (Scheer 2004)[98]. Certes tout corpus ne représente qu'un échantillon des interactions langagières possibles mais ces données constituent la base empirique pour toute formulation des modèles phonologiques dans une approche *bottom-up*. De plus, le recours à l'intuition du locuteur natif ou du linguiste lui-même a été l'objet de plusieurs critiques d'ordre empirique et méthodologique (Le Prieult 2006). Cette pratique se base en effet sur des jugements de grammaticalité qui relèvent d'une évaluation subjective propre de l'intuition d'un locuteur ou d'un groupe restreint d'individus[99] et qui se traduiront, dans la plupart des cas, dans la formulation de règles.

[98] Le corpus dans la phonologie générativiste est conçu dans le but précis de valider les hypothèses définies auparavant par le linguiste (Scheer 2004) en suivant une approche *top-down*.

[99] Le Prieult (2006) illustre bien le passage entre un jugement normatif basé sur des « valeurs historiques dominantes » à une notion d'usage purement idéologique, propre de la linguistique générative, basée sur une définition restreinte (bien qu'abondamment commentée) de l'intuition du locuteur natif chomskien (Le Prieult, 2006, p. 120).

3.1.1. Des consonnes finales latentes : application de la règle de troncation (Schane 1965)

Le recours aux exemples élaborés par les auteurs normatifs et à une tradition d'écrit oralisé est évident dans les travaux de Schane (1965), le premier à proposer en 1965 une théorisation du phénomène de la liaison posant « des bases phonologiques très proches de l'écriture orthographique » (Encrevé 1988, p. 89). En effet, Schane (1965), comme nous l'avons vu dans le paragraphe précédent (cf. Paragraphe 2.1), traite les CL comme des consonnes finales stables faisant partie de la représentation phonologique du M1. Cette approche oblige l'auteur à se limiter à l'étude des contextes de non liaison décrits à l'aide de la règle de troncation (*French Truncation Rule*) (cf. Figure 3). En effet cette analyse rend compte des cas où la liaison n'a pas lieu en supprimant les CL (contenues dans la représentation phonologique des mots) dans les contextes de non-réalisation de la liaison.

> « […] la liaison n'est traitée dans sa grammaire [*de Schane*] que dans les cas où elle *n'a pas lieu* : l'orthographe présente en effet des consonnes de liaison que Schane écrit donc dans la base, et la règle qu'il pose consiste à *supprimer* ces consonnes dans les cas de non-liaison : ce qui revient à faire passer de cette écrite à la prononciation effective ». (Encrevé 1988, p. 89)

Ce traitement, comme le souligne Malmberg (1974) rappelle celui adopté par la linguistique historique pour rendre compte des formes de la langue moderne.

> « On aura remarqué déjà dans ces quelques exemples de Schane que les règles transformationnelles ressemblent étrangement aux règles diachroniques dont se servait la description historique traditionnelle pour 'expliquer' les formes de la langue moderne, seulement en sens inverse ». (Malmberg 1974, p. 222)

De la même façon Schane essaie de décrire la liaison en sens inverse : de l'orthographe des mots aux formes orales.

De plus, Schane se donne pour objectif, à travers l'élaboration de cette règle, de formaliser simultanément le phénomène de liaison (ou mieux de non-réalisation de la liaison) et d'élision (qui entraîne la chute d'une voyelle finale devant un mot à initiale vocalique) en le présentant comme le résultat d'un même processus de troncation.

> « […] a final vowel is deleted or truncated before another word beginning with a vowel, whereas a final consonant is deleted before another word beginning with a consonant. In order to be neutral between the terms elision and absence of liaison, we shall often refer to this one and the same process as truncation […]. Then elision for vowels and absence of liaison for consonants can be considered as the same process : deletion of a segment in word final position ». (Schane 1968, p. 2)

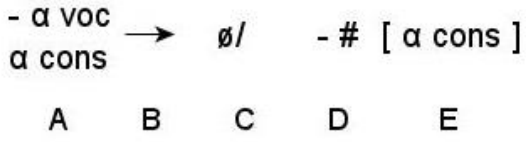


Figure 3: Règle de Troncation (Schane 1965)

La règle correspond à la formulation suivante : toute voyelle ou consonne (A) devient (B) nulle (c'est-à-dire qu'elle tombe) (C) si elle se trouve devant une frontière de mot (#) c'est à dire si elle est en finale de mot (D) et si elle est suivie d'un segment phonologique ayant la même nature (consonne ou voyelle) (E). Autrement dit la chute de ces éléments est soumise non seulement aux conditionnements phonologiques et syntaxiques que nous avons déjà évoqués (cf. Partie 1.Chapitre 1.2.3) mais est également influencée par des règles basées sur la répartition des segments phonologiques en classes naturelles (voyelles, semi-voyelles, liquides et consonnes).

Il en résulte que la règle de troncation rend compte d'une part du phénomène d'élision et donc de la chute des voyelles devant voyelles ou semi-voyelles. D'autre part, ce même formalisme ne peut que représenter les cas de non-réalisation de la liaison où la CL (consonne finale stable intégrée à la représentation phonologique) tombe si elle est suivie d'un mot commençant par consonne ou liquide.

Cela revient à dire, comme Encrevé (1988) l'a fait observer, que la règle de troncation formulée par Schane (1965), animé par la préoccupation de rendre compte simultanément de deux phénomènes distincts (élision et liaison), ne couvre en réalité que les cas de non-réalisation de la liaison négligeant totalement le traitement de la variabilité de la liaison, nœud principal de l'étude de ce phénomène.

> « Schane ne peut, en effet, lier les deux phénomènes, toujours distingués dans la grammaire française, de la non-liaison et de l'élision, qu'à condition de ne traiter que les cas où la non-liaison est obligatoire comme l'est l'élision : les cas où les mots suivants commencent par une consonne, ce qui permet le rapprochement avec l'élision d'une voyelle devant un mot commençant par une voyelle mais laisse en dehors tous les cas de non-liaison devant voyelle, soit tous les cas où la liaison facultative n'est pas réalisée ». (Encrevé 1988, p. 90)

Schane (1968) essaye de traiter l'analyse des contextes de liaison variable ayant recours à l'application d'une deuxième règle, la règle d'effacement final (*Deletion Rule*), qui puisse rendre compte (de manière très marginale) des conditionnements syntaxiques qui régissent la réalisation de la liaison et qui donc « delete a word final consonant : 1. obligatorily : a. in phrase final position; b. in a singular noun; 2. optionally, in a plural noun » (Schane 1968, p. 13).

Cependant, comme le souligne Encrevé (1988), cette dimension de variation syntaxique ne peut pas être formalisée dans le cadre du modèle de la phonologie linéaire du *Sound Pattern of English* (SPE) (Chomsky et Halle 1968).

> « […] cette Règle de délétion ne peut pas être formalisée dans le cadre de *SPE*, qui ne permet pas d'exprimer le fait que le même processus se réalise obligatoirement dans certains contextes et facultativement dans d'autres contextes. Elle est donc indéfendable dans ce cadre ». (Encrevé 1988, p. 92)

Le recours à des sources normatives (principalement basées sur des variétés d'écrit oralisé) et au cadre de SPE, conduit Schane (1968) à élaborer une formalisation qui ne peut représenter que des conditionnements phonologiques et morphologiques excluant toute autre variable.

3.1.1.1 Schane (1974)

En 1974 Schane (1974) abandonne les règles de troncation et d'effacement final proposée entre 1965 et 1968 en faveur d'une seule règle de suppression (cf. Figure 4) capable de rendre compte des cas de non-réalisation de la liaison devant toute consonne (C) ou une frontière du mot (#).

$$C \rightarrow \emptyset \; / \; - \; ([\,-\,seg\,])\,\{\,C, \#\,\}$$

Figure 4 : Règle de Suppression (Schane 1974)

Bien que Schane semble dans cette approche séparer les phénomènes d'élision et de liaison, il ne parvient toujours pas à résoudre les inconvénients de son analyse, à savoir le manque de la dimension de variation. De plus, dans les analyses élaborées par Schane (1965 ; 1968 ; 1974) et par les autres travaux issus de la phonologie générative linéaire, comme nous l'avons déjà évoqué (cf. Paragraphe 2.1), les CL ont un statut de consonnes fixes qui les rapproche de toute autre consonne finale. Cela revient à dire que non seulement cette règle, comme d'ailleurs les précédentes, ne représente que les contextes de non-réalisation de la liaison mais qu'elle ne rend même pas compte des comportements variables des consonnes fixes finales. Pour conclure, la description des seuls contextes de non liaison exclut *a priori* toute analyse de l'enchaînement à la fois des consonnes CL et des consonnes fixes.

3.1.1.2 Schane (1978)

Dans des travaux suivants influencés par d'autres études, Schane (1978) proposera une règle spécifique (cf. Figure 5) pour le traitement de la liaison selon laquelle une CL est effacée

(toujours par la représentation sous-jacente du mot) quand la frontière du mot (#) coïncide avec la frontière syllabique ($).

$$C \rightarrow \emptyset \; / \; - \; \# \; \$$$

Figure 5 : Règle de Suppression de la CL (Schane 1978)

Cette analyse, selon Schane (1978) permet d'effacer toutes les CL ne se trouvant pas en contexte de liaison étant donné que celui-ci serait « le seul où le placement de la frontière de syllabe ne coïncide pas avec la frontière de mot ». Une telle approche, comme le souligne Encrevé (1988), revient à nier l'enchaînement en dehors de liaison, alors que comme nous l'avons déjà souligné, celle-ci représente un cas particulier du processus d'enchaînement (cf. Partie 1.Chapitre 1.2).

3.1.2. Des consonnes finales enchaînées : application de la règle de métathèse (Dell 1970)

Contrairement aux travaux de Schane, la resyllabation des CL sera au centre de l'étude de Dell (1970) qui rejette l'analyse par troncation élaborée par cet auteur proposant deux traitements distincts pour l'étude de l'élision et de la liaison. Celle-ci sera décrite par Dell (1970) sous forme d'une « Règle de Liaison » (« LIAS ») (cf. Figure 6) qui enchaîne obligatoirement la CL (ayant pourtant toujours un statut de consonne finale stable) à la voyelle initiale (marquée par un trait [+syll]) du mot suivant.

> « La règle LIAIS indique que, lorsque dans une séquence de mots M1#M2, le premier est terminé par une consonne et que le second commence par une voyelle, la consonne de M1 passe à l'initiale de M2 ». (Dell 1970)

$$\text{LIAIS (obl) : } [\text{ - syll }] \; \# \; [\text{ + syll }] \qquad 1 \quad 2 \quad 3 \quad = \quad 2 \quad 1 \quad 3$$

Figure 6 : Règle de Liaison (Dell 1970)

Cette règle, improprement qualifiée par Dell (1970) comme une règle de métathèse[100], comme le remarque Encrevé (1988, p. 26), semble plus décrire le processus de resyllabation de toute consonne finale (CL ou consonnes fixes) que le phénomène spécifique de liaison.

[100] Cette dénomination vient du fait que, comme Encrevé (1988, p. 99) le fait observer, Dell (1970) adopte le format transformationnel proposé par *SPE* (Chomsky et Halle 1968) pour décrire les métathèses. D'où la confusion qui amène Dell (1970) à traiter le phénomène de resyllabation des CL (à savoir une modification d'appartenance syllabique de la CL) comme un phénomène de métathèse (où un échange de position a lieu entre deux segments ayant la même nature ce qui exclut *a priori* le contexte de liaison).

> « […] LIAIS, qui déplace une frontière de mot et du même coup produit une resyllabation, n'est pas, telle qu'elle est formulée, une Règle de liaison mais une Règle d'enchaînement, et c'est la seule que Dell propose, confondant en fait le maintien de CL et sa resyllabation qu'il traite comme un seul phénomène ». (Encrevé 1988, p. 26)

De plus, à l'instar de Schane, Dell (1970) est obligé de postuler l'existence d'une deuxième règle, la « Règle de Troncation » (cf. Figure 7) qui efface tout segment obstruant en fin de mot, à savoir toute consonne ne se trouvant pas en contexte de liaison. Cette règle devrait permettre de rendre compte des contextes de non-réalisation de la liaison.

$$\text{TR (obl): } [\text{ - son }] \rightarrow \text{ø} \ / \ \text{—} [\text{ - son }]_{0}\#$$

Figure 7 : Règle de Troncation (Dell 1970)

Cependant, étant donné que les CL possèdent dans les approches génératives linéaires un statut qui ne les distingue pas des consonnes fixes finales, cette règle de troncation risquerait également d'effacer toute consonne finale et ne permettrait pas de rendre compte de cas où celles-ci sont toujours prononcées (ex. « chef », « sec », « sept » etc.). Pour cette raison, Dell (1970) propose d'attribuer à la représentation sous-jacente de ces mots un *schwa* final (/ʃɛfə/ ; /sɛkə/ et /sɛtə/) protégeant leurs consonnes finales de l'application de la Règle de Troncation. Toutefois, comme Encrevé (1988, p. 26) le fait remarquer, ce réajustement empêche à ces consonnes fixes d'être enchaînées, étant tout simplement exclues du contexte d'application de la Règle de Liaison. Cet aspect démontre bien comme la Règle de Liaison n'est en réalité qu'une description de l'enchaînement qui finit par bloquer la resyllabation des consonnes fixes finales.

L'inconvénient de cette représentation est donc double. D'une part elle ne distingue pas le statut des CL des autres consonnes et, d'autre part, dans le cadre de la description du phénomène de la liaison, elle ne sépare pas la réalisation de la CL de sa resyllabation. Ces deux aspects conduiront au « résultat paradoxal d'interdire l'enchaînement en dehors de la liaison, c'est-à-dire l'enchaînement proprement dit » (Encrevé 1988, p. 26).

3.1.2.1 Dell (1973)

Dans une étude successive Dell (1973) propose une nouvelle règle de Troncation proche de celle élaborée par Schane (1968). Dell (1973), contrairement à l'analyse développée en 1970, intègre dans sa description la dimension syntaxique stipulant l'existence d'une relation entre les dimensions phonologiques et syntaxiques selon laquelle lorsque la condition phonologique est remplie, intervient la condition syntaxique.

3.1.3. Des consonnes épenthétiques : application de la règle d'épenthèse

C'est dans le cadre de la phonologie générative naturelle que se développera une conception épenthétique de la liaison visant à limiter l'abstraction des représentations sous-jacentes. Dans ces approches la CL sort donc de la représentation sous-jacente du M1 pour acquérir une nature indépendante.

Encrevé (1988, p. 119) souligne que contrairement à la phonologie générative « standard » du SPE (Chomsky et Halle 1968), les travaux issus de la phonologie générative naturelle (en particulier les études de Tranel) ont en général une préférence pour les données issues de l'oral. Tranel semblerait attribuer une grande importance à l'étude attentive des données, parfois issues aussi des « fausses liaisons », qu'il utilise pour argumenter ses hypothèses.

3.1.3.1 Klausenburger (1974)

Le premier à proposer une analyse épenthétique fut Klausenburger (1974) qui définit une règle de liaison où l'insertion de la CL épenthétique est conditionnée par la présence de la voyelle suivante. Cette approche sera critiquée entre autres par Encrevé (1988, p. 105) qui mettra en évidence l'abstraction ainsi que l'indépendance par rapport au M1 et M2 de la CL qui finit par posséder « le même statut que n'importe quelle autre consonne ». De plus, Encrevé (1988) et Pagliano (2003) soulignent que le processus formel par le biais duquel la grammaire parvient à sélectionner et insérer la forme correcte de CL épenthétique ne semble être jamais décrit par Klausenburger (1974).

3.1.3.2 Tranel (1974)

En revanche, le processus d'insertion épenthétique sera illustré par Tranel (1974) dans l'analyse du comportement des liaison en [n].

Rejetant les règles de troncation des consonnes finales, il propose une règle d'insertion pour les liaisons en [n] qui défend, contre la phonologie générative « standard », l'existence de voyelles nasales sous-jacentes pour les mots qui ne présentent pas d'alternance Ṽ / VN.

3.1.3.3 Tranel (1976 ; 1981)

Tranel (1976) modifie la règle d'insertion de [n] (cf. Figure 8), en défendant la présence de voyelles nasales sous-jacentes même dans les cas de mots ayant une alternance Ṽ / VN en surface (ex. [bɔn] / [bɔ̃]) pour lesquels il formalise une règle de dénasalisation de la voyelle.



Figure 8 : Règle de dénasalisation de la voyelle (Tranel 1976)

Cette analyse en termes d'insertion sera par la suite élargie à d'autres contextes et d'autres CL (Tranel 1981) comme les liaisons en [z] marquant le pluriel, des liaisons verbales en [z], [t] et [ʀ] et aussi à toute CL appartenant à des mots invariables.

Dans le cadre de cette nouvelle analyse, il reconnaît une place cruciale à la cohésion syntaxique dans la description des processus phonologiques conduisant à l'insertion des CL. Cependant, d'après Encrevé (1988), cette approche, comme les autres modèles génératifs linéaires sont incapables « de reconnaitre le rôle des structures non segmentales et notamment, s'agissant de liaison, syllabiques » (Encrevé 1988, p. 119).

3.2. Les modèles de la phonologie multilinéaire : des consonnes instables

En réaction au modèle de la phonologie générative linéaire du *SPE* (Chomsky et Halle 1968), dans les années 70, un nouveau cadre formel (métrique, autosegmental et tridimensionnel) se met en place autour de la publication respectivement des travaux de Liberman (Liberman et Prince 1977; Liberman 1979), Goldsmith (1979) et Kahn (1980). Ceux-ci poseront les bases de la phonologie multilinéaire. L'apport de ces travaux dans l'étude de la liaison consiste principalement dans le fait d'avoir dissocié l'analyse métrique de celle des constituants squelettaux et syllabiques. La représentation phonologique se compose donc de plusieurs lignes autosegmentales parallèles (suites alignées d'unités homogènes telles que des accents, des phonèmes, des syllabes etc.) mises en relation par un certain nombre de conventions universelles et de règles[101]. Le traitement de la liaison ne se réalise donc plus seulement au niveau des segments phonologiques, à savoir des consonnes finales, mais sur la base de constituants sous-jacents plus larges tels que les syllabes (Dell, Hirst et Vergnaud 1984).

Contrairement aux modèles de la phonologie générative linéaire, certains des auteurs appartenant à ce courant abandonnent les sources normatives en faveur de travaux structuralistes et descriptifs comme celui de Delattre (1966) ou des données issues des enquêtes (Encrevé 1988). En effet, comme nous le verrons dans les prochaines pages, ces travaux ne

[101] Parmi celles-ci nous pouvons citer la contrainte de bonne formation formulée par Goldsmith (1976) selon laquelle les lignes verticales qui associent les unités des différentes lignes (horizontales) autosegmentales ne peuvent pas se croiser (*Line Crossing Constraint*).

cherchent plus à définir les cas où la CL chute mais plutôt ceux où elle se produit, pour s'ouvrir vers une description de la variation du phénomène.

3.2.1. Consonnes flottantes

Ce cadre permettra d'abandonner les analyses par troncation en faveur d'une approche visant à décrire les contextes où la liaison se produit. En effet, le cadre autosegmental que nous avons brièvement évoqué permet d'apporter une solution formelle à la description de la variation de la liaison. Les CL non réalisées ne seront donc plus décrites comme des consonnes finales qui chutent mais comme des « segments flottants » non associés au squelette (flottement squelettal) (Wetzels 1987; Tranel 1990) ou à la syllabe (flottement syllabique) (Clements et Keyser 1983) ou aux deux à la fois (Encrevé 1988) [102]. En revanche, leur réalisation aurait lieu grâce à l'ancrage de ces segments dans la structure prosodique, si une position d'attaque est disponible (à savoir si la condition syllabique est remplie).

3.2.2. Des consonnes flottantes par rapport au squelette

3.2.2.1 Wetzels (1987), Tranel (1990)

Wetzels (1987), Tranel (1990) proposent de considérer les CL comme des éléments flottants par rapport au squelette. En particulier, Tranel (1990), dans le cadre d'une analyse par supplétion, traite les CL des mots non-supplétifs (en l'occurrence des adjectifs ex. « gros ») comme des segments flottants dont la réalisation enchaînée peut être considérée comme le résultat de l'ancrage de la CL flottante au squelette de la syllabe initiale du M2.

> « I assume that the fundamental characteristic of linking consonants in non-suppletive cases is that they are floating melodies without lexical skeletal slots. […] Liaison with *enchaînement*, the normal case of liaison, can be viewed as the conjunction of the insertion of a skeletal slot for the floating consonant and the assignment of this slot as an onset to the nucleusinitial syllable of the second word ». (Tranel 1990, p. 182‑183)

3.2.3. Des consonnes flottantes par rapport à la syllabe

3.2.3.1 Clements et Keyser (1983)

Dans le cadre de cette analyse les CL sont présentes dans la représentation sous-jacente d'un mot, mais contrairement à d'autres segments, elles ne sont pas associées à un nœud syllabique (σ). Clements et Keyser (1983) proposent une règle de liaison qui, sous des conditions syntaxiques appropriées, lie un élément consonantique à savoir la CL (C) à un nœud syllabique (σ) dominant un élément vocalique (V) suivant (cf. Figure 9).

[102] Pour un bilan voir Tranel (1995a ; 1995b).

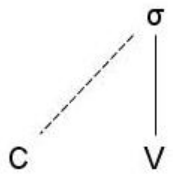


Figure 9 : Flottance syllabique de la CL

En d'autres termes les CL constituent des segments « extrasyllabiques » (Clements et Keyser 1983, p. 77)[103]. Par exemple le mot « petit » serait représenté de la manière illustrée par la figure ci-dessous (cf. Figure 10).

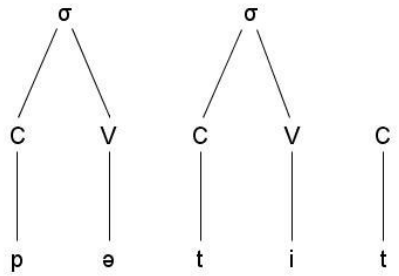


Figure 10 : Extrasyllabicité de la CL du mot « petit » (Clements et Keyser 1983)

Encrevé (1988, p. 121) critiquera l'emploi de la notion d'extrasyllabicité dans cette approche, soulignant que dans le modèle trilinéaire proposé par ces auteurs les CL « ne sont pas exactement des auto-segments flottants » puisqu'elles sont reliées à une position de la ligne centrale (CV-*tier*), et que c'est seulement celle-ci à ne pas être associée à la ligne syllabique (σ). La non-flottance des CL par rapport à la ligne centrale (*CV-tier*) oblige Clements et Keyser à proposer une Règle de troncation capable de supprimer les consonnes que la Règle de liaison n'a pas reliées à la ligne syllabique. De plus, Encrevé (1988) fait observer que, comme dans le modèle de Dell (1970), la liaison est confondue avec le phénomène plus général d'enchaînement excluant implicitement de l'analyse les cas de liaison non-enchaînée ainsi que l'enchaînement de toute autre consonne fixe.

3.2.4. Des consonnes flottantes par rapport au squelette et à la syllabe

3.2.4.1 Encrevé (1988)

L'analyse élaborée par Encrevé (1988), contrairement aux travaux précédents, propose une solution formelle pour distinguer la production de la CL de son enchaînement. Un traitement séparé de ces deux aspects oblige Encrevé à postuler une flottance à la fois

[103] Cette notion d'extrasyllabicité sera ensuite reprise par Booij (1984) qui proposera une Règle de Liaison en dehors du modèle trilinéaire où les CL non reliées ne seront pas supprimées par une Règle de Troncation mais simplement sur la base d'une « convention » qui établit la suppression de tout segment non dominé par un nœud syllabique.

squelettale et syllabique de la CL. Dans le cadre de la phonologie autosegmentale, le statut spécifique de la CL est représenté sur trois plans différents: (a) sur la ligne autosegmentale des segments, la CL correspond à un « segment flottant » sans association avec le squelette ; (b) dans le plan autosegmental syllabique sur la ligne des constituants de la rime, une « coda flottante » sans association avec le squelette correspond au segment flottant à savoir la CL ; (c) sur la ligne du squelette, une position sans aucune association est disponible pour l'ancrage à la fois de la CL et d'un constituant syllabique (Encrevé 1988, p. 169).

Autrement dit, la CL flottante, dépourvue de toute association avec la ligne auto-segmentale des segments (a) et des syllabes (b), ne peut être réalisée et enchaînée que si une position sur la ligne du squelette est disponible pour l'ancrage à la fois de la CL et d'un constituant syllabique (c) (« condition phonétique »)[104].

> « Le segment flottant correspond à l'analyse selon laquelle la CL appartient lexicalement au mot mais doit avoir une représentation distincte de celle des consonnes toujours réalisées dans la prononciation : flottante, la CL dépend pour sa prononciation de l'intervention, ou non, d'une opération d'ancrage dans le squelette, ainsi que d'une opération d'interprétation syllabique de la position dans laquelle elle s'ancre ». (Encrevé 1988, p. 170)

En revanche, dans le cas des liaisons non-enchaînées, la CL ne reçoit pas d'ancrage syllabique mais uniquement squelettal. La distinction opérée par Encrevé entre ancrage squelettal et ancrage syllabique permet comme nous l'avons dit de rendre compte séparément du phénomène d'enchaînement et de non-enchaînement des liaisons attestées dans son corpus, ainsi que de la variation de la liaison[105].

Cependant, la double flottance postulée par Encrevé rend possible l'ancrage de la CL flottante correspondant à une « coda flottante » (cf. ci-dessus) (sur la la ligne des constituants de la rime à savoir noyau (N) et coda (C)) à tout constituant accessible sur la ligne des constituants syllabiques (attaque (A) ou de rime (R)) (point* cf. Figure 11). Cet inconvénient oblige Encrevé (1988) à poser une convention (« convention universelle d'association des codas flottantes ») qui établit qu' « une coda flottante ne peut s'associer sur la ligne des constituants syllabiques A, R qu'avec l'unité R dominant le noyau qui la précède sur la ligne N, C » évitant une configuration mal formée (*) (Encrevé 1988, p. 171).

[104] Sur la base d'une « condition phonétique » propre au français, la CL ne peut s'ancrer au squelette du M1 que si « le mot suivant possède à l'initiale de la couche syllabique un élément A *flottant*, c'est-à-dire non lié au squelette, à l'inverse d'une attaque consonantique ». Le recours à une notion de flottance squelettale et syllabique permet à Encrevé d'éviter d'établir « une règle spéciale supprimant son association avec le squelette lorsque la liaison est enchainée » (Encrevé 1988, p. 171)

[105] En effet, la condition de double flottance (squelettale et syllabique) de la CL permettrait de rendre compte des cas de non-réalisation de la liaison.

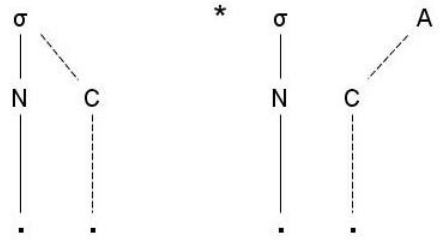


Figure 11 : Convention universelle d'association des codas flottantes (Encrevé 1988)

Autrement dit cette convention établit que l'association à la ligne des constituants syllabique (A et R) ne peut se réaliser que par lien direct du segment flottant avec un constituant attaque (A), et « non par l'intermédiare d'une coda flottante qui se 'tranformerait' en attaque » (Encrevé 1988, p. 172). Ce statut spécial permet à Encrevé de rendre compte de la liaison non-enchaînée tout en distinguant la CL des autres consonnes finales en évitant le recours à des règles de réajustement. En effet, contrairement aux CL, les consonnes fixes « sont ancrées dans le squelette dès la représentation lexicale, qu'elles soient en position initiale, interne ou finale dans le mot » (Encrevé 1988, p. 173).

Cependant, comme le souligne Laks (2005b) à la suite de Tranel (1995a), cette solution de double flottance des CL proposée par Encrevé assigne un statut lexical marqué en complexifiant la représentation lexicale du mot. De plus, si d'une part ce statut de double flottance permet comme nous l'avons vu de séparer le phénomène de la liaison du phénomène plus général de l'enchaînement (et aussi de distinguer les CL des consonnes fixes), il présente d'autre part l'inconvénient de laisser libre l'ancrage de la CL[106].

3.3. Des consonnes non intrinsèques : les modèles basés sur la théorie de l'optimalité

Le cadre de la théorie de l'optimalité (*Optimality Theory*, dorénavant OT) (Prince et Smolensky 1993), théorie générative non dérivationnelle en rupture avec l'analyse par règle des modèles précédents, constitue un nouveau cadre d'étude du phénomène de la liaison. La notion de règle sera donc abandonnée en faveur d'un ensemble de contraintes hiérarchisées tout en conservant la distinction entre représentations sous-jacentes (inputs) et représentations de

[106] « Qu'est-ce alors qui explique que les consonnes de liaison, sans ancrage et/ou sans syllabation lexicalement définie, continuent d'être finales dans la représentation du mot et viennent s'ancrer uniquement à l'initiale suivante sous liaison ? […] Totalement flottante, une consonne pourrait, en principe, s'ancrer en toute position accessible, par exemple en position interne au premier ou au second mot qui présenterait une suite voyelle-voyelle et donc une place consonantique vide ». (Laks 2005b, p. 106)

surface (outputs). Étant donné un input et un ensemble de contraintes universelles hiérarchisées de manière différente dans chaque langue (CON), un module (GEN) permet de générer un ensemble (théoriquement infini) de candidats et de violations de contraintes. Ensuite un module (EVAL) évalue l'ensemble des candidats selon une hiérarchie de contraintes universelles (établies à partir de CON) de la contrainte la plus forte à la plus faible. Ce processus d'évaluation élimine les candidats les moins harmoniques (ceux qui présentent les plus de violations de la contrainte) et fait revenir le candidat optimal ou les candidats optimaux, c'est-à-dire le ou les output(s).

Une telle approche comparative régie par des contraintes permet d'abandonner un système basé sur des règles, jugé comme purement descriptif, incapable de fournir une explication aux phénomènes phonologiques (Tranel 2000, p. 44). En particulier, Tranel (2000, p. 44) rejette l'analyse par règles qu'il estime être non explicative et « le plus souvent spécifique à une langue donnée et même à un phénomène donné dans une langue donnée ». En réaction à ce traitement, il propose de traiter la liaison et l'élision, non pas comme les résultats du même processus comme dans l'analyse de Schane (1968), mais comme des phénomènes spécifiques du français qui peuvent être décrits comme les effets principaux de la contrainte universelle d'évitement d'*hiatus*[107].

> « […] le concept fondamental d'évitement d'*hiatus* (la contrainte *W) permet d'échapper à l'étroitesse de vue inhérente à des règles qui s'appliquent dans une langue donnée (ici le français) et décrivent des phénomènes spécifiques (ici l'élision et la liaison). […] L'élision et la liaison peuvent donc être conçues non plus comme des processus idiosyncrasiques qui font partie de la grammaire du français, mais comme des effets dus à la contrainte universelle *W ». (Tranel 2000, p. 45)

Cependant cette contrainte universelle est en compétition avec d'autres types de contraintes que nous examinerons dans les pages qui suivent.

3.3.1. Tranel (2000)

Bien que Tranel (2000) ne définisse pas le statut lexical des CL il opère un distinction entre les consonnes fixes qu'il qualifie d' « intrinsèques »[108] et les CL considérées comme des consonnes lexicalement disponibles, mais non intrinsèques.

[107] Cependant Tranel (2000) précise qu'il exclut de cette analyse « les cas de la liaison plurielle et de la liaison verbale, où l'apparition d'une consonne de liaison peut se produire sans qu'il y ait de nécessité de résolution d'*hiatus* (durs-anneaux, court-il ?). La liaison a ici un caractère morpho-sémantique qui est absent ailleurs » (Tranel 2000, p. 45).

[108] Les consonnes fixes ou intrinsèques répondent à la contrainte de fidélité dominante MAX(C) qui pénalise les effacements et exige que tout élément C dans un input donné ait un correspondant C dans l'output. Autrement dit MAX établit qu'un input soit « maximalement » représenté en surface et interdit donc l'effacement des consonnes fixes.

> « […] qu'il s'agit, selon les analyses, ou bien d'une consonne finale flottante, ou bien d'une consonne empruntée au féminin, ou bien encore d'une consonne de déclinaison, en tout cas d'une consonne lexicalement disponible, mais non intrinsèque ». (Tranel 2000, p. 51)

L'apparition de cette CL en surface, à savoir dans l'input, est conditionnée par l'interaction de contraintes différentes telles que les contraintes DEP (« dépendance des éléments de l'output dans l'input ») qui pénalisent en général les épenthèses et donc l'insertion des CL (DEP (C)) mais qui peut être violée lorsqu'il s'agit d'éviter un *hiatus* (*VV). Dans le cadre de la grammaire OT la contrainte universelle d'évitement d'*hiatus* (*VV) est donc dominante sur DEP (C).

Comme le souligne Eychenne (2011), cette analyse présente quelques incohérences théoriques. En effet, si la CL doit être présente et disponible lexicalement dans l'input (cf. ci-dessus) il est difficile d'imaginer que son apparition dans l'output viole une contrainte DEP qui exige que tout élément de l'output ait un correspondant en input. Cette contradiction relève probablement d'un manque de prise de position quant au statut lexical des CL et surtout d'une volonté, en opposition aux autres modèles génératifs, de n'appliquer aucune règle au niveau des représentations sous-jacentes (Tranel 2000, p. 47).

3.4. Les modèles de la phonologie exemplariste : des consonnes appartenant à des constructions

Les modèles de la phonologie exemplariste, issus de la tradition des Grammaires de Construction (Goldberg 1995) et de la Grammaire Cognitive (Langacker 1987), conçoivent la phonologie comme une connaissance procédurale (Bybee 2001b). En effet, contrairement aux approches génératives, les représentations ne sont pas stables mais sujettes aux changements résultant de l'usage. Celui-ci joue donc un rôle central dans la modélisation des représentations phonologiques étant donné que leur stockage est tributaire de la fréquence d'usage des formes de surface. Les représentations sous-jacentes (conçues plutôt comme des généralisations) ne sont pas séparées des réalisations mais émergent au contraire directement de ces dernières.

Dans le cadre de ces approches les CL sont considérées comme indépendantes du M1 et M2 et font partie de constructions plus larges à savoir des *chunks* (traduit en français par le terme « tronçon »). Ces unités de base constituent des séquences fréquentes de morphèmes ou de mots qui sont reliées « par des relations sémantiques, fonctionnelles ou statistiques particulières lorsqu'ils sont utilisés ensemble, relations qu'ils n'entretiennent pas forcément dans d'autres contextes » (Bybee 2005, p. 26).

Bybee (2005) précise que le figement (à savoir le degré de cohésion), ainsi que le stockage de ces tronçons, varient en fonction de la fréquence de cooccurrence de la séquence (Bybee

2005, p. 26). Celle-ci relève aussi bien de l'automatisation liée à la répétition d'une séquence fréquente que de la catégorisation des mots qui occupent des positions récurrentes au sein des constructions.

> « Les constructions sont plus ou moins conventionnelles, en fonction du degré auquel l'usage répété les a établies dans une langue particulière. Les mécanismes d'établissement des constructions sont (i) l'automatisation par répétition de « tronçons » de matériau linguistique (ii) la catégorisation des items qui apparaissent dans certaines positions de ces « tronçons » plus grands. Puisque la fréquence d'emploi est un facteur majeur dans la formation des constructions, il n'est pas obligatoire qu'elles aient une signification non-déductible des segments qui les composent : il peut s'agir simplement de « tronçons » de discours employés fréquemment ». (Bybee 2005, p. 26)

Ces constructions correspondent aux unités de base à l'intérieur desquelles les CL sont stockées. L'hypothèse de lexicalisation des CL, comme le souligne Bybee (2005), serait appuyée par les réalisations de liaisons non-enchaînées ou précédées d'une pause qui suggérerait que l'analyse de la liaison s'inscrit en dehors des unités traditionnelles de syllabe, mot phonologique ou groupe rythmique.

> « Cette hypothèse de la lexicalisation ou de l'intégration de la liaison à la morphologie est étayée par le fait - signalé dans différentes études, mais plus explicitement dans Encrevé (1988), Morin et Kaye (1982) et Post (2000), ainsi que dans Âgren (1973) et Green et Hintze (1988) - que les consonnes de liaisons peuvent apparaître aussi bien avant qu'après une pause, avec ou sans resyllabation sur la voyelle de droite. Autrement dit, alors que la liaison dépendait à l'origine de la resyllabation, elle n'est plus contrainte aujourd'hui par les frontières des groupes créés par des pauses ou par celles d'un mot phonologique ». (Bybee 2005, p. 26)

La réalisation de la liaison serait donc liée à la fréquence d'usage et plus précisément à la fréquence de cooccurrence de la séquence dans laquelle elle est produite. Celle-ci détermine le degré de cohésion[109] de la séquence ainsi que son stockage sous forme de constructions. Les séquences de mots employées fréquemment possèdent donc un fort degré de cohésion syntaxique et sont directement stockées et représentées dans la mémoire. Ces séquences fréquentes acquièrent une plus grande force lexicale qui facilite leur accessibilité et donc la réalisation de la liaison. L'usage modifie ainsi les représentations lexicales favorisant le développement de généralisations basées sur des schémas plus ou moins ouverts (selon l'emplacement occupé par des items issus de classes ouvertes et selon la fréquence de répétition de la forme de surface).

[109] Le degré de cohésion des séquences est donc le résultat direct de la fréquence de cooccurrence d'une séquence, ce qui montre comment dans une telle approche syntaxe, phonologie et stockage lexical interagissent les uns avec les autres.

L'opposition traditionnelle entre liaison catégorique et liaison variable est ainsi ré-analysée sur la base de la notion de la fréquence d'usage conduisant d'une part à la lexicalisation des constructions grammaticales telle que le contexte « déterminant - nom » ou des expressions figées et d'autre part à la définition de schéma plus ouvert (cf. Figure 12) comme dans le cas des séquences « nom - adjectif » (où deux emplacements sont occupés par des items issus de classes ouvertes).

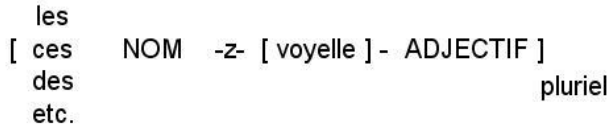


Figure 12 : Construction à schéma ouvert de la réalisation de la liaison dans le type « nom-adjectif » (Bybee 2005)

Quant à la variabilité des réalisation de la liaison, Bybee (2005) propose de la traiter comme le résultat de l'émergence variable de schémas conditionnée encore une fois par la fréquence d'usage des formes de surface correspondantes. Dans ce cadre, la production variable de la liaison des séquences « nom - adjectif » serait donc liée à la présence de deux schémas, l'une correspondant aux contextes de liaison (cf. Figure 12) (moins fréquent puisque limité aux adjectifs à initiale vocalique) et l'autre de non-réalisation de la liaison (cf. Figure 13) plus fréquent (parce qu'il s'applique à davantage de mots) et donc plus facilement accessible.

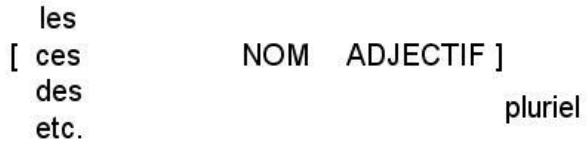


Figure 13 : Construction à schéma ouvert de la non-réalisation de liaison dans le type « nom-adjectif » (Bybee 2005)

Dans la lignée de ces études, Laks (2005b) propose de considérer les liaisons postposées à une catégorie principale (cf. Partie 1.Chapitre 2.2.3.2) et sans valeur fonctionnelle comme « des constructions synthétiques stockées comme telles ».

3.5. Les recherches en acquisition L1

Si les travaux théoriques sur la liaison possèdent une longue tradition, ce n'est qu'au cours de ces dernières années que des études systématiques (Chevrot et Fayol 2000 ; Chevrot et Fayol 2001 ; Wauquier-Gravelines et Braud 2005 ; Dugua 2006; Chevrot, Chabanal et Dugua 2007 ; Wauquier 2009 ; Chevrot et al. 2013 ; Wauquier et Shoemaker 2013) ont été menées dans l'acquisition de la L1. Ces travaux ont permis en particulier de recueillir chez les enfants natifs,

à travers la mise en place de protocoles expérimentaux, différents types d'erreurs pouvant constituer des indices pour l'étude du statut des CL : (a) substitution d'une CL en contexte de liaison (ex. [œ̃zuʀs] pour « un ours ») ; (b) non-réalisation d'une CL en contexte de liaison catégorique (ex. [œ̃uʀs] pour « un ours ») ; (c) insertion d'une CL non attendue (ex. « papa ours » [papanuʀs]) ; (d) substitution d'une consonne initiale stable (ex. [œ̃nɛbʀ] pour « un zèbre ») ; (e) non-réalisation d'une consonne initiale stable (ex. [œ̃ɛbʀ] pour « un zèbre »).

Si l'ensemble des travaux s'accordent sur la description des données définissant une chronologie commune d'évolution des erreurs et d'acquisition, les positions divergent quant à l'interprétation des résultats et à l'attribution du statut des CL. Deux principaux scénarii développementaux ont été notamment élaborés. D'une part, un scénario d'acquisition appelé « exemplariste » (Chevrot et Fayol 2000 ; Chevrot et Fayol 2001 ; Dugua 2006 ; Chevrot, Chabanal et Dugua 2007 ; Chevrot et al. 2013) fondé sur une conception lexicale de l'acquisition de la liaison basée sur la théorie des exemplaires (Pierrehumbert 2001) et des Grammaires de Construction (Goldberg 1995). D'autre part, un scénario développemental basé sur des principes universels d'acquisition faisant partie d'une grammaire en construction chez l'enfant (Wauquier-Gravelines et Braud, 2005 ; Wauquier 2009 ; Wauquier et Shoemaker 2013 ; Splendido 2014 ; Ruvoletto 2014 ; Ruvoletto et Wauquier 2015).

Pour ce qui est du scénario « exemplariste » développé par Chevrot et al., les auteurs identifient trois étapes principales basées sur les propositions de Morin (2003). Lors de la première étape (vers l'âge de 2-3 ans), l'enfant essaierait de résoudre la disjonction entre frontière lexicale et frontière de syllabe créée par la liaison, en mémorisant plusieurs exemplaires de chaque unité lexicale (/narbr/, /zarbr/, /tarbr/ dans le cas du mot « arbre »).

En effet, dans la tranche 2-3 ans et dans la tranche 3-4 ans, l'enfant aurait tendance à privilégier une segmentation CV (Dugua 2002) et donc à entendre chaque M2 précédé de différentes CL : par exemple, « arbre » précédé de /n/ dans « un arbre », de /z/ dans « les arbres », de /t/ dans « petit arbre ». Ces formes seraient attestées dans les cas d'erreurs par substitution de la CL, très fréquentes à ce stade (5%) selon l'étude de Chevrot et al. (2009). Dans un deuxième stade, vers l'âge de 3-4 ans, à partir du jeu d'exemplaires mémorisés, l'enfant apprendrait les connexions entre chaque M1 avec la variante adéquate du M2. L'apprentissage de la liaison, dans une première phase, reviendrait à apprendre les associations entre l'exemplaire du M2 (ex. /zarbr/) et le M1 précédent (« les »).

La mémorisation de ces réseaux d'associations se mettrait en place par exposition aux séquences M1-M2 bien formées rencontrées dans l'environnement langagier. Cela contribuerait à la diminution des erreurs de substitution. Dans une troisième étape, une nouvelle typologie

d'erreurs émerge chez les enfants de 4-5 ans, basée sur la substitution des consonnes stables initiales des M2 par une CL (ex. [œ̃nɛbʀ] pour « une zèbre »). D'après Chevrot et al. ce type d'erreur indiquerait que l'enfant surgénéralise le fonctionnement de la liaison à des items lexicaux qu'il n'a jamais entendus. À ce stade l'enfant serait capable de développer des schémas abstraits (Bybee 2001b), à partir du jeu d'exemplaires appris au cours des deux premières étapes. Ces séquences mémorisées permettraient à l'enfant de mettre en relation un M1 spécifique et une classe de variantes de M2 telle que « les + zX » (liant le déterminant défini pluriel à l'ensemble des variantes commençant par le phonème /z/).

Contrairement au modèle « exemplariste », pour Wauquier-Gravelines et Braud (2005), les enfants francophones L1 acquerraient la liaison, non pas par exposition aux séquences Ml-M2 rencontrées dans leur environnement langagier, mais plutôt par généralisation grammaticale à partir de l'âge de 4 ans, correspondant à la période de disparition des erreurs dans les contextes catégoriques. Selon ces auteurs, l'enfant aurait dans un premier stade développemental une représentation globale des séquences de liaison, correspondant à la notion de gabarit (Vihman et Croft 2007 cité par Wauquier-Gravelines et Braud 2005) (une unité minimale précoce à partir de laquelle les généralisations phonologiques se développent). Une forme de « proto-déterminant » ferait partie de cette unité globale. Le second stade se situerait au moment de l'explosion lexicale, et correspond à la segmentation des gabarits : c'est alors que l'enfant détermine la frontière entre déterminant et nom. Cette segmentation s'actualise en favorisant la structure CV (structure qui respecte le principe du *Maximal Onset* à savoir « attaque maximale »), rattachant la position consonantique de la CL à l'attaque du M2. Toutefois à ce stade le contenu phonétique de cette attaque consonantique du M2 est encore vide et non défini. Cela expliquerait pourquoi l'enfant remplit les attaques de syllabe du M2 par des CL erronées. D'après les auteurs, ces remplissages erratiques seraient à l'origine des erreurs de substitution observées dans le modèle « exemplariste ». Cependant, le choix de la consonne erratique ne serait pas aléatoire mais lié à différents facteurs tels que la probabilité d'emploi de la consonne en contexte, l'harmonie avec une consonne interne au nom (« un éléphant » produit [lelefɑ̃]) et l'utilisation d'une semi-voyelle par défaut (« les oiseaux » prononcé [lejajo]).

Au troisième stade, vers l'âge de 4 ans, dans les contextes de la liaison catégorique entre article et nom, le développement de généralisations d'ordre morphologique permettrait à l'enfant de traiter la CL comme un morphème ayant une forme proche de celle du féminin. Grâce à ce processus, appelé *bootstrapping* morphologique, l'enfant aboutirait à une

représentation phonologique où la CL, sous forme d'un segment flottant, serait encodée à la finale du M1[110].

En ce qui concerne l'acquisition des liaisons variables (qui commence vers 6-7 ans et se finalise vers l'âge de 7-8 ans), les deux modèles semblent converger vers une approche plus « lexicaliste » où les associations sont optionnelles et se font au cas par cas. Dans ces contextes, Wauquier-Gravelines et Braud (2005) ne font pas référence à une acquisition phonologique mais plutôt à l'acquisition d'une maîtrise de compétences pragmatiques et stylistiques qui se développe contexte par contexte. Celle-ci serait favorisée également par l'acquisition de la lecture et par l'exposition aux formes orthographiques des mots, absentes à l'âge d'acquisition de la liaison catégorique chez l'enfant préscolaire.

À la suite de ces résultats et des différentes interprétations données par les deux modèles développementaux, il est possible d'établir des hypothèses quant à l'évolution du statut de la CL chez les enfants natifs. Comme Côté (2005, p. 69) le fait observer, si les erreurs produites dans les deux premiers stades d'acquisition (à savoir les substitutions des CL) semblent suggérer que la CL est rattachée au M2, l'augmentation des non-réalisations des CL d'une part, et de remplacements et omissions des consonnes stables dans le stade suivant, d'autre part, pourrait indiquer un détachement progressif de la CL du M2 et une éventuelle acquisition d'un statut épenthétique des CL.

> « Une réanalyse des CL comme épenthétiques rend compte de ces deux évolutions. Les CL se détacheraient progressivement du mot de droite et seraient plutôt insérées. L'omission des CL résulterait de la non-application du processus d'épenthèse. Le détachement se renforcerait au fur et à mesure que l'enfant progresse dans la liaison, d'où une probabilité accrue d'omissions. Les erreurs sur les consonnes initiales stables correspondraient à une surapplication du processus de détachement : l'enfant détache les /n, z, t/ initiaux dans les mots qui possèdent une variante à initiale vocalique mais également au début de quelques mots qui ne présentent pas d'alternances (p. ex. zèbre) ». (Côté 2005, p. 69)

En revanche le maintien du rattachement au M2 est défendu par Dugua (2006) qui suggère d'interpréter les erreurs de non-réalisation des CL comme le résultat d'un mécanisme d'inhibition des exemplaires à initiale consonantique qui entraînerait un renforcement de la variante à initiale vocalique. Des processus d'analogie pourraient, selon Côté (2005), expliquer l'apparition des erreurs d'omission et de substitution des consonnes stables où ces formes seraient créées sur le modèle d'autres items à initiale vocalique (ex. « ours »).

[110] Plus récemment, Encrevé, Scheer et Wauquier (2015) ont suggéré d'interpréter les occurrences de liaison non-enchaînée attestées par l'étude de Splendido (2014) (cf. Partie 1.Chapitre 2.2.1.1) comme des traces d'une étape intermédiaire précédant le *bootstrapping* morphologique (stade 3) (cf. ci-dessus).

Le rattachement des CL au M1 comme consonnes finales est également envisageable[111] même si, comme le souligne Côté (2005), on devrait observer dans ce cas des erreurs de prononciation des CL dans des contextes de non-liaison (ex. devant des mots à initiale consonantique) ou des confusions entre CL et consonnes fixes alors qu'aucune de ces formes n'est attestée dans les travaux évoqués ci-dessus.

3.6. Conclusion

Les différentes positions concernant le statut de la CL que nous avons évoquées dans les pages précédentes renvoient à un débat théorique plus large qui anime également les recherches psycholinguistiques. En effet la notion de statut lexical de la CL oblige à prendre en compte la question de la représentation des mots comportant une liaison dans le lexique mental du locuteur. Cette problématique est au cœur des recherches psycholinguistiques.

C'est au sein de ce débat, comme le relève Racine (2008, p. 66), que la phonologie et la psycholinguistique se rejoignent et que l'étude phonologique de liaison s'enrichit de l'apport des recherches psycholinguistiques. Néanmoins la rencontre entre ces deux disciplines dans le domaine d'étude de la liaison est assez récente[112], et est le plus souvent envisagée dans le cadre d'une contribution théorique plus que d'une formulation d'une approche commune. Il nous paraît donc intéressant dans le prochain chapitre (cf. Partie 1.Chapitre 6) de compléter l'étude des traitements théoriques de la liaison par un bref survol des approches psycholinguistiques d'analyse de la liaison.

[111] Dans ce cas les non-réalisations des CL dans des contextes catégoriques pourraient être expliquées comme des erreurs relevant d'un stade transitoire où la CL ne sont rattachées ni au M1 ni au M2.

[112] Voir Soum-Favaro, Coquillon et Chevrot 2014.

Chapitre 6. Représentations de la liaison : approches psycholinguistiques

Dans ce chapitre nous présenterons un bref bilan non exhaustif des approches psychologuistiques d'analyse de la liaison. Nous nous attacherons tout d'abord à décrire quelques modèles de reconnaissance des mots parlés, écrits et de la production orale. Nous décrirons par la suite les hypothèses proposées pour rendre compte des représentations de la liaison et du traitement cognitif au sein des processus de perception et de production préalablement décrits. Cela nous amènera à étudier le rôle joué par l'orthographe sur les représentations de la liaison et les effets inter-tâches correspondantes.

1. Modèles perceptifs et productifs : repères théoriques

1.1. Modèles de reconnaissance des mots parlés

Plusieurs modèles ont été proposés pour rendre compte de la reconnaissance des mots au sein de l'énoncé à partir de l'analyse du signal de la parole. Comme le souligne Segui (2003, p. 224) deux questions principales méritent d'être examinées dans l'élaboration d'un modèle de reconnaissance des mots : (a) quelle est la nature des informations exploitées par le système dans l'identification des mots ? ; (b) dans quelle mesure les connaissances lexicales interviennent-elles pendant la réception du signal sonore afin de faciliter la reconnaissance des mots ?

Deux différentes approches ont été élaborées afin de répondre à ces questions et de tenter de rendre compte du processus de segmentation et de reconnaissance des frontières lexicales. D'une part, les approches « lexicales » (ex. modèle Cohorte cf. Paragraphe 1.1.1) (cf. Bagou 2006 ; Racine 2008) qui conçoivent la reconnaissance d'un mot comme un processus séquentiel s'appuyant sur les indices acoustiques. D'autre part, des approches « sous-lexicales » (ex. modèle TRACE cf. Paragraphe 1.1.2) (cf. Bagou 2006; Racine 2008) qui considèrent que la reconnaissance des mots parlés implique une compétition lexicale entre plusieurs candidats lexicaux. Celle-ci n'est pas uniquement conditionnée par des facteurs de nature acoustique mais également par des informations contextuelles, rythmiques et phonotactiques[113].

[113] Néanmoins le débat ne concerne pas seulement les mécanismes qui guident la segmentation et la reconnaissance des frontières des mots. Les positions divergent également quant à la nature de la relation existant entre les unités sensorielles et l'information stockée dans le lexique (hypothèse pré-lexicale et hypothèse directe) et aussi bien quant aux modalités de représentation encodées dans le lexique mental (hypothèse abstractionniste et exemplariste) (cf. Partie 1.Chapitre 6.2). Concernant la première problématique, deux modèles concurrents

Deux différents modèles (modèle Cohorte et modèle TRACE) seront brièvement présentés dans les pages qui suivent afin d'illustrer les deux approches « lexicale » et « sous-lexicale » que nous avons évoquées ci-dessus.

1.1.1. Modèle Cohorte

Le modèle « lexical » (cf. Bagou 2006 ; Racine 2008) considère dans sa première version (Marslen-Wilson et Welsh 1978) que la reconnaissance d'un mot a lieu à partir du moment (dit « point d'identification » ou « point d'unicité ») où le mot cible est le seul candidat compatible avec le signal acoustique perçu. Le traitement progressif du signal sonore entrant permet d'éliminer de l'ensemble des candidats possibles (« cohorte ») ceux qui ne correspondent plus au signal acoustique. Comme le font observer Spinelli et Ferrand (2005), dans ce modèle « l'activation des candidats se fait 'en tout ou rien' : soit un mot fait partie de la cohorte, soit il en est exclu ». La reconnaissance des mots parlés est, dans cette première approche, fortement tributaire des informations acoustiques permettant à elles seules la détermination de la cohorte initiale. Cependant ce modèle postule également le recours à des données d'ordre contextuel dans la mesure où elles contribuent à la réduction progressive des candidats préalablement activés. L'importance accordée aux aspects acoustiques, comme le soulignent Spinelli et Ferrand (2005, p. 72), revient à attribuer un rôle crucial à l'ordre temporel de réalisation et donc aux phonèmes initiaux dans la mesure où ils permettent de sélectionner la cohorte initiale.

Dans une version modifiée du modèle présentée par Marslen-Wilson (1987) l'hypothèse d'activation catégorielle des candidats est abandonnée en faveur d'une approche « sous-lexicale » basée sur plusieurs niveaux d'activation. Cela a pour conséquence que la reconnaissance des candidats ne s'appuie plus sur une correspondance précise entre le signal acoustique (traits phonétiques) et la représentation lexicale, ce qui permet l'identification des items « mal prononcés »[114]. Cette deuxième version du modèle Cohorte permet également d'intégrer dans le traitement les effets de fréquence des candidats selon lesquels les mots de haute fréquence possèderaient un niveau d'activation plus élevé que les mots rares.

s'opposent sur ce point : les modèles pré-lexicaux et les modèles directs. Si l'hypothèse pré-lexicale envisage l'existence d'une représentation intermédiaire (sous forme de syllabes ou phonèmes) entre l'input sensoriel continue et les représentations discrètes stockées dans le lexique mental, aucune hypothèse n'est en revanche formulée quant à une étape entre le signal et les représentations par les modèles directs (voir Spinelli et Ferrand 2005 ; Racine 2008).

[114] Voir (Spinelli et Ferrand 2005, p. 72).

1.1.2. Modèle TRACE

Le modèle TRACE est un modèle « sous-lexical » (cf. Bagou 2006 ; Racine 2008) à activation interactive dans lequel le traitement est réalisé par un ensemble d'unités issues de différents niveaux : des traits acoustiques, des unités phonémiques et des unités des mots. Celles-ci sont reliées entre elles à l'intérieur d'un réseau caractérisé par des connexions facilitatrices entre niveaux et inhibitrices au sein d'un même niveau. D'après ce modèle les traits phonétiques du premier niveau permettent de détecter les phonèmes du deuxième niveau. Ceux-ci sont également activés en fonction des mots au troisième niveau celui des unités lexicales. Lorsque le mot est activé, un mécanisme de rétroaction lexicale se met en place en renforçant les unités déjà activées et en donnant lieu à des inhibitions intra-niveaux des autres phonèmes et mots et ainsi à la réduction des hypothèses lexicales. La puissance d'inhibition dépend de la force des connexions des unités.

La reconnaissance du mot a lieu lorsqu'il est significativement plus activé que tous les autres mots à savoir lorsque sa probabilité de reconnaissance dépasse le seuil défini par la règle de Luce (1959 cité par Spinelli et Ferrand 2005, p. 78). Ce seuil est calculé en divisant la force de réponse d'une unité particulière par la somme des forces de réponse de toutes les unités pour un temps donné.

Contrairement au modèle Cohorte la réduction progressive des unités lexicales n'est pas opérée de façon strictement linéaire mais présuppose un alignement exhaustif selon lequel le locuteur n'est pas conscient des frontières des mots. Le modèle TRACE postule un système de réduplication du lexique qui permet de vérifier toutes les hypothèses lexicales activées par la perception d'un nouveau phonème et de mettre en place les mécanismes d'inhibition qui permettront d'aboutir au seuil de reconnaissance.

1.1.3. Limites

Comme le font observer Spinelli et Ferrand (2005, p. 85) « aucun des modèles spécifiquement dédiés à la reconnaissance des mots parlés ne prend en compte l'influence de l'orthographe dans le traitement de la parole ». Cet aspect, sur lequel nous reviendrons dans les prochains paragraphes (cf. Paragraphe 3), est encore plus frappant si nous considérons la place accordée au phonème dans les modèles de reconnaissance des mots écrits (cf. Paragraphe 1.2)[115].

[115] Dans les pages qui suivent (cf. Partie 1.Chapitre 6.3.4) nous présenterons le modèle interactif bimodal (Grainger et Ferrand 1996; Ziegler, Muneaux et Grainger 2003) qui reconnaît l'existence d'interactions entre les représentations phonologiques et orthographiques de manière à rendre compte de l'influence de l'orthographe tant en perception qu'en production.

De plus, la plupart de ces modèles ont recours aux phonèmes comme unités intermédiaires du processus d'appariement entre signal acoustique et représentations lexicales abstraites. Cependant étant donné les phénomènes de coarticulation auxquels les phonèmes sont sujets plusieurs approches ont proposé la syllabe comme unité d'accès au traitement de la parole (Mehler et al. 1981 ; Cutler et al. 1986)[116].

En outre ces modèles ne permettent pas de rendre compte des effets d' « illusion » perceptive attestés par les études de Dupoux et al. (1998 ; 1999 ; 2001) (cf. note 128).

1.2. Modèles de reconnaissance des mots écrits

Quant aux modèles proposés pour le traitement de la lecture à haute voix nous ferons référence au travail de Detey (2005) qui cite deux types d'approches. L'une basée sur un modèle de type Cohorte développé par Johnson (1992 cité par Detey 2005) et l'autre s'appuyant sur la notion de « cascade » (Ziegler, Perry et Coltheart 2000 cité par Detey 2005).

1.2.1. Modèle Cohorte

En s'appuyant sur les travaux menés par Marslen-Wilson, Johnson (1992 cité par Detey 2005) développe un modèle Cohorte pour la reconnaissance des mots écrits basé sur trois principes :

- « Words conceal their letters » (« word-priority effect »);
- « Lexical processing is facilitated by increasing word frequency »;
- « Lexical processing is influenced […] by the number of words that are visually similar to the target ».

Ce modèle est articulé autour de deux composants principaux :

- l'encodage visuel de l'information orthographique ;

[116] Par opposition aux modèles de perception catégorielle établi par la première version du modèle Cohorte (Marslen-Wilson et Welsh 1978), des modèles de perception multimodale seront proposés afin d'intégrer les informations de sources multiples d'information dans le processus de reconnaissance de la parole. En particulier, Massaro (1987) proposera un modèle multimodal de perception de parole connu sous le nom de modèle Flou (*Fuzzy logical model of perception* dorénavant FLMP).

Cette approche abandonne un traitement catégoriel du signal par intermédiaire des unités discrètes telles que les phonèmes en faveur d'une logique qui repose sur une série continue de valeurs de vérité. La comparaison entre les différentes valeurs de vérité assignées aux sources d'information permet d'exclure les hypothèses lexicales et de parvenir à la reconnaissance du mot. Celle-ci est réalisée en trois étapes : (a) évaluation des traits : une valeur de vérité est assignée à chaque trait à l'intérieur de la syllabe en fonction du degré d'appariement entre le trait perçu et celui correspondant au sein du prototype mémorisé ; (b) intégration des traits : dans la deuxième étape les traits sont comparés avec les prototypes stockés dans la mémoire afin d'évaluer le degré d'appariement d'un trait spécifique à l'intérieur de la syllabe ; (c) classification des schémas : les prototypes potentiels sont évalués afin d'aboutir à la reconnaissance du mot.

- la relation entre l'information issue de cet encodage et l'activation ou la désactivation des représentations lexico-sémantiques.

1.2.2. Modèle à deux voies en cascades (Dual-Route Cascaded Model, DRC)

Le modèle *Dual-Route Cascaded Model* (dorénavant DRC) de reconnaissance visuelle lexicale et de lecture à haute voix a été proposé par Ziegler, Perry et Coltheart (2000 cité par Detey 2005). Le modèle DRC s'articule en deux « voies » ou « routes » :

- route lexicale (voie d'adressage) basée sur deux principes opérés de façon automatique. L'un présuppose une activation interactive bidirectionnelle entre différents niveaux (axe vertical) et l'autre d'inhibition au sein d'un même niveau (axe horizontal) ;
- route non-lexicale (voie d'assemblage) : à l'intérieur de cette voie un processus sériel de conversion phonographémique (*Grapheme-Phoneme Correspondences*, GPC) permet de transformer les lettres ou groupes de lettres en phonèmes. Ce mécanisme est régi par deux autres principes. Le premier dit de « constance morphémique » postule que malgré tout changement de prononciation déterminé par le contexte, l'orthographe des morphèmes reste stable. Le deuxième principe permet d'évaluer les irrégularités orthographiques sur la base du désaccord par rapport aux règles de correspondances phonographémiques (GPC) d'une langue donnée[117].

1.3. Modèles de production orale

Les principaux modèles de production de la parole ont été établis par Levelt (1989) et par Dell (1986).

1.3.1. Levelt

Le modèle de Levelt (1989) s'appuie sur les travaux de Garrett (1975) et s'articule en plusieurs modules de production de l'oral :

- le « conceptualisateur » permet l'élaboration du contenu du message préverbal sous forme de représentations sémantiques et pragmatiques. Ce processus est le résultat d'une part d'une phase de macroplanification (sélection des données pertinentes à la

[117] Pour plus de détails concernant les modèles de reconnaissance des mots écrits et parlés voir Detey (2005, pp. 275-279).

situation de communication) et d'autre part de microplanification (sélection des formes linguistiques disponible en fonction de la situation de communication) ;

- le « formulateur » correspond à l'encodage grammatical et phonologique du message. Le premier sélectionne les éléments lexicaux disponibles dans le lexique mental et les organise au sein de la structure syntaxique et grammaticale du message. Pour cette raison ce modèle postule un lien direct avec les composants du lexique à savoir sa signification, ses propriétés syntaxiques (lemme), ses caractéristiques morphologiques (forme morpho-phonologique de l'item lexical). En revanche l'encodage phonologique définit les formes phonologiques des entrées lexicales sélectionnées ;
- l' « articulateur » permet finalement d'exécuter les choix faits par l'encodeur phonologique ;
- pour terminer, un système de compréhension du discours met en place des procédures de contrôle de la production orale permettant de corriger les erreurs et de réaliser un *feedback* sur le message.

1.3.2. Dell

Le modèle proposé par Dell (1986) est de type connexionniste et s'appuie sur les mêmes étapes de traitement proposées par Levelt (1989) : traitement sémantique, syntaxique, morphologique et phonologique. Cependant, contrairement au modèle de Levelt (1989), Dell (1986) postule un système de restructuration des représentations à chaque étape du processus de production orale. De plus, si dans le modèle de Levelt (1989) la mise en place des étapes de traitement est linéaire, d'après Dell (1986) les différentes étapes peuvent avoir lieu simultanément et s'influencer réciproquement au sein d'un réseau régi par des connexions d'activation et inhibition.

1.4. Bilan

Le bref survol des modèles que nous avons effectué dans les paragraphes précédents nous a permis d'identifier selon différentes approches (catégorielle et connexionniste) les étapes sous-tendant les processus cognitifs de la reconnaissance orale et écrite ainsi que de la production orale. Bien que ce bilan ne soit pas exhaustif, il nous permet de mieux situer les hypothèses et les résultats des travaux psycholinguistiques menés sur la liaison (cf. Paragraphe 1.5 ; Paragraphe 1.6) et également d'évaluer le niveau (lexical ou sub-lexical) d'influence de l'orthographe dans les processus de perception et de production de la liaison (cf. Paragraphe 3).

1.5. Reconnaissance des séquences de liaison

Différentes études psycholinguistiques ont été consacrées à la reconnaissance des mots à l'intérieur des séquences de liaison.

Les travaux conduits par Gaskell et al. (2002) ont permis de vérifier un éventuel coût de traitement dans la reconnaissance des mots resyllabés par la liaison à travers une série d'expériences d'amorçage intermodal en comparant le temps de réponse de la condition de liaison (ex. « un généreux italien ») avec des conditions d'enchaînement (ex. « un virtuose italien ») et d'alignement syllabique (ex. « un chapeau italien »). Aucune différence entre les conditions décrites ci-dessus n'a été observée, ce qui suggère que l'ambiguïté transitoire ainsi que les changements phonologiques déterminés par la liaison ne ralentissent pas la reconnaissance des mots de liaison. Des travaux successifs verront dans le recours au contexte lexical (Spinelli, Mcqueen et Cutler 2002) et à des indices acoustiques associés à la CL (Spinelli, Mcqueen et Cutler 2003) des éléments informatifs capables d'assurer l'accès aux représentations des mots de liaison et donc de justifier l'absence de coût de traitement dans ces contextes.

L'étude de Spinelli et al. (2002) propose en particulier aux sujets de réaliser une tâche de décision lexicale sur des mots cibles (ex. « avion ») qui étaient précédés par des amorces contenant une séquence de liaison variable (ex. « grand avion ») ou un type de liaison « illégale » (ex. « vrai [t] avion »). Les cibles ont été reconnues plus rapidement dans le premier type d'amorce (« grand avion ») que dans les séquences de liaison « illégale » (ex. « vrai [t] avion ») ce qui semblerait soutenir l'hypothèse selon laquelle les informations du contexte lexical serviraient à résoudre l'ambiguïté transitoire engendrée par la liaison.

Quant au recours aux indices acoustiques pour la désambiguïsation des contextes de liaison, l'étude de Wauquier-Gravelines (1996) montre que les locuteurs natifs n'exploitent pas les différences acoustiques (ex. durée totale et d'occlusion des CL) existant entre les CL et les consonnes fixes initiales des M2 et que la reconnaissance des CL présente un coût de traitement plus important par rapport à celle des consonnes initiales. Les travaux menés par Grosjean et Yersin-Besson (1996) semblent nier la présence de différences de durée entre les CL et les consonnes fixes initiales. En revanche, la recherche conduite par Spinelli et al. (2003) indique que les CL sont en moyenne de 15% plus brèves que les consonnes initiales. De plus à l'aide d'un paradigme de *cross splicing*[118] les auteurs ont montré que les indices acoustiques aussi

[118] Les auteurs ont inverti la partie de signal acoustique correspondant à « tavion » produit dans les séquences déjà utilisées dans Spinelli et al. (2002) afin de créer des séquences hybrides telles que « grand avion »

bien que les informations du contexte lexical sont exploités dans le processus de désambiguïsation des contextes de liaison. Cependant les indices acoustiques ne sont pas suffisants pour assurer la réduction des candidats dans une décision lexicale sur des mots cibles (ex. « oignon » ; « rognon » qui étaient respectivement précédés par des amorces contenant les séquences « c'est le dernier oignon » et « c'est le dernier rognon»). L'étude récente menée par Tremblay et Spinelli (2014), semble aller dans la même direction en montrant que les indices acoustiques ne parviennent pas à bloquer les effets de compétitions lexicales lorsque les mots sont à initiale vocalique, à savoir en contexte de liaison. Ces indices acoustiques seraient donc exploitables dans des cas d'absence d'un « compétiteur lexical explicite » (Tremblay et Spinelli 2014, p. 110).

Tremblay et Spinelli (2014) s'interrogent également sur l'influence de ces indices phonético-acoustiques sur les représentations de la liaison stockées dans le lexique mental du locuteur. En accord avec les hypothèses abstractionnistes ces indices acoustiques influenceraient les représentations sub-lexicales en activant des allophones positionnels (Tremblay et Spinelli 2014, p. 111). En revanche sur la base des hypothèses exemplaristes les entrées lexicales seraient stockées en mémoire sous forme de plusieurs variantes caractérisées par des indices acoustico-phonétiques détaillés, chacune constituant « un signal phonétique idosyncrasique ».

1.6. Production orale des séquences de liaison

D'après les modèles établis pour la production orale (Dell 1986 ; Levelt 1989), les entrées lexicales seraient stockées dans le lexique mental sous forme de représentations phonologiques abstraites et seraient encodées phonologiquement avant d'être intégrées dans la structure métrique de l'énoncé. À ce stade la séquence phonologique peut faire l'objet de modifications et d'ajustements en fonction du contexte et c'est à ce moment que se situe le traitement cognitif en vue de la production ou de la non-réalisation de la liaison.

Nous présentons ici l'état de l'art proposé par Burki et Laganaro (2014) concernant les hypothèses (abstractionnistes et exemplaristes) et les résultats des études psycholinguistiques (Bagou et Laganaro 2009 ; Laganaro et al. 2010) consacrées à l'analyse du traitement et des représentations cognitives de production de la liaison.

Burki et Laganaro (2010) distinguent deux différentes hypothèses abstractionnistes. D'une part celle qui considère que la forme mémorisée correspond à la forme sans la CL. Dans

= indice acoustique de la consonne initiale + indice lexical contexte de liaison et « vrai [t] avion » = indice acoustique CL+indice lexical contexte de non liaison.

le cadre de la production orale, selon cette hypothèse, la CL serait ajoutée pendant l'étape d'encodage phonologique en fonction du contexte (à savoir si le mot suivant est à initiale vocalique). La CL ne fait pas partie de la représentation phonologique et donc la réalisation de la liaison implique un coût de traitement pour la production de la liaison.

D'autre part une deuxième hypothèse abstractionniste serait celle selon laquelle la représentation stockée en mémoire contiendrait la CL. Une telle approche postule donc que la non-réalisation de la liaison serait le résultat d'un mécanisme d'effacement ayant lieu pendant l'étape d'encodage phonologique. Contrairement à l'hypothèse précédente, la suppression de la CL et donc la non-réalisation de la liaison impliquerait un coût de traitement par rapport à sa réalisation.

En revanche, d'après les hypothèses exemplaristes les deux différentes formes (avec et sans CL) stockées dans le lexique mental permettraient de rendre compte respectivement des réalisations et des non-réalisations de la liaison. De plus, comme nous l'avons déjà souligné (cf. Partie 1.Chapitre 5.3.4), contrairement aux approches précédentes, l'hypothèse exemplariste attribue un rôle déterminant à la fréquence des formes (avec et sans CL) pour un mot donné. Plusieurs études (pour un bilan voir Burki et Laganaro 2010) ont démontré que la rapidité de production est influencée par la fréquence d'usage d'une variante donnée.

En effet, plusieurs arguments apportés par les études psycholinguistiques menées sur d'autres phénomènes phonologiques (ex. le *schwa*) (Burki et Gaskell 2012) semblent confirmer l'hypothèse exemplariste. L'étude menée par Burki et Gaskell (2012) sur des pseudo-homophones avec et sans *schwa* en position post-accentuelle (ex. « sement » et « sment ») a montré que les mots anglais de base correspondant (ex. *cement*) ayant deux variantes (avec et sans *schwa*) possèdent deux représentation distinctes dans le lexique mental du locuteur. Les résultats des travaux menés par Burki et Frauenfelder (2012) et basés sur un paradigme de nouveaux mots ont également montré que parmi ceux-ci les mots pouvant se prononcer avec ou sans *schwa* (ex. [ʃənyk] ou [ʃnyk]) présentent un coût de traitement plus élevé que ceux n'ayant qu'une seule variante. Les auteurs ont également relevé un effet de fréquence selon laquelle la variante la plus fréquente serait produite le plus souvent. Les résultats de ces études semblent suggérer que deux représentations distinctes sont stockées dans le lexique mental du locuteur dans le cas d'entrées lexicales présentant deux variantes de réalisation en compétition pendant le processus de production.

Contrairement aux travaux conduits en perception, un nombre restreint de travaux psycolinguistiques ont été jusqu'à présent consacrés à l'étude de représentations de la liaison dans la production. Burki et Laganaro (2010) mentionnent en particulier une étude

psycholinguistique (Bagou et Laganaro 2009) et une deuxième de nature neuro-psycholinguistique (Laganaro et al. 2010) conduite auprès de sujets aphasiques.

Les résultats de Bagou et Laganaro (2009) indiquent que les séquences (« grand ami », « grand tapis ») contenant des adjectifs pouvant donner lieu à deux variantes de prononciation (avec ou sans CL ex. « grand » [grɑ̃] et [grɑ̃t]) présentent un coût, en termes de latence de production[119], plus important par rapport aux séquences constituées par des adjectifs n'ayant qu'une seule prononciation (ex. « demi »). Ces résultats sembleraient apporter des arguments en faveur de l'hypothèse exemplariste d'après laquelle la réalisation ou la non-réalisation de la liaison sous-tend une sélection entre deux différentes représentations de la même entrée lexicale stockées dans la mémoire lexicale.

Les données issues de l'étude menée par Burki et al. (soumis) à l'aide d'une tâche de dénomination d'image[120] vont dans la même direction : les résultats de ce travail montrent en effet que la latence de production du déterminant « un » devant des séquences constituées par un adjectif suivi d'un nom est moins importante lorsque ces derniers (adjectif et nom) possèdent tous les deux une attaque aussi bien consonantique (« un demi » + nom à initiale consonantique) que vocalique (« un immense » + nom à initiale vocalique) (*phonological consistent sequences*) que lorsqu'ils ont une attaque différente (« un demi » + nom à initiale vocalique ou « un immense » + nom à initiale consonantique) (*phonological inconsistent sequences*). Les mêmes tendances ont été observées dans l'étude de Burki, Laganaro et Alario (2014) sur la variation phonologique des déterminants « mon » et « ma » (expérience 1), « la » (« l'» devant adjectif à initiale vocalique) (expérience 2) et « la » (« l' » devant adjectif à initiale vocalique) et « le » (« l'» devant adjectif à initiale vocalique) (expérience 3) devant des séquences constituées par un adjectif suivi d'un nom à initiale consonantique ou vocalique[121].

Ces résultats révèlent un effet de consistance phonologique selon lequel on observe l'influence d'un contexte « éloigné » (*distant context* à savoir le nom) sur la production du

[119] Nous citons ici la définition de latence de production proposée par Alario (2008). « […] le temps qu'il faut à un locuteur pour retrouver un mot en mémoire et commencer à l'articuler. Un même locuteur pourra produire un mot de façon plus ou moins rapide selon les circonstances dans lesquelles ce mot est présenté. Cette vitesse variable est interprétée comme un indicateur de la difficulté de mise en œuvre de processus cognitifs, ou de récupération de représentations en mémoire. Expliquer pourquoi la production est plus ou moins rapide (i.e. plus ou moins facile) dans telle ou telle condition expérimentale peut directement contribuer à expliquer comment cette production est organisée » (Alario 2008, p. 2-3).

[120] Pour plus de détails sur la méthodologies utilisée voir le protocole de l'étude de Burki, Laganaro et Alario (2014).

[121] Ex. *phonological consistent sequences* : « ma nouvelle table » (expérience 1), « la nouvelle table » (expérience 2) et « la nouvelle balance», « le demi banc » (expérience 3) ; ex. *phonological inconsistent sequences* « mon ancienne table » (expérience 1), « l'ancienne table » (expérience 2) et « l'ancienne balance», « l'immense banc ») (expérience 3).

déterminant qui est phonologiquement contrainte par le contexte le plus proche (*local context* à savoir l'adjectif). À en croire les auteurs, cela suggère que la production de la liaison n'est pas soumise à l'application de règles phonologiques post-lexicales mais s'appuie sur des variantes différentes (avec et sans liaison) des déterminants stockés dans la mémoire du locuteur.

Du point de vue neuro-psycholinguistique dans les travaux menés par Laganaro et al. (2011), les auteurs ont demandé à des patients aphasiques de produire les mêmes séquences soumises dans l'étude précédente. Un taux important de non-réalisations a été observé dans les séquences de liaison potentielle. Si ces résultats permettent d'exclure l'hypothèse abstractionniste postulant une forme unique de représentation contenant une CL[122], ils ne permettent pas de trancher entre les deux hypothèses restantes.

Une deuxième étude conduite par Pellet Cheneval et al. (2013) (toujours sur des sujets aphasiques) montre une corrélation entre les taux de non-réalisation de la CL et la capacité de planification phonologique des phénomènes inter-mots. Comme le font observer Burki et Laganaro (2014), ces données contrastent avec les résultats de l'étude de Bagou et al. (2010) et des études menées sur le *schwa* (et donc avec une hypothèse exemplariste) et sembleraient suggérer l'existence d'une forme unique de la représentation de la liaison (sans CL) à laquelle les sujets aphasiques ne seraient pas en mesure d'associer une CL pendant la phase d'encodage phonologique. Cependant une autre interprétation pourrait à notre avis rendre compte de l'ensemble de ces résultats. En effet si, à la suite des approches constructionnistes d'analyse de la liaison (cf. Partie 1.Chapitre 5.3.4), nous considérons la CL comme faisant partie de constructions plus larges que le mot et (partiellement) lexicalisées, nous pourrions fournir une explication à ces données apparemment contradictoires. Tout en conservant une hypothèse exemplariste la présence de non-réalisations observée chez les sujets aphasiques, associée à leur capacité réduite de planification inter-mots, pourrait être l'effet d'une difficulté à sélectionner la variante correcte au sein de la séquence M1-M2 qu'ils ne parviennent pas à anticiper et pour laquelle aucune construction n'est accessible. Il s'agit bien évidemment d'une hypothèse qui nous semble toutefois envisageable, étant donné l'ensemble des données issues des études psycholinguistiques et neuro-psycholinguistiques.

[122] En effet, si la CL était encodée dans la forme stockée dans le lexique mental, on aurait dû obtenir des productions de liaisons dans des contextes non attendus (ex. devant M2 à initiale consonantique).

2. Représentations de la liaison dans le lexique mental

Les travaux consacrés à l'analyse des représentations et des processus cognitifs de la liaison sont relativement récents. Des études psycholinguistiques et neuro-psycholinguistiques ont été menées en particulier en perception (Wauquier-Gravelines 1996 ; Grosjean et Yersin-Besson 1996 ; Spinelli, Mcqueen et Cutler 2002 ; Spinelli, Cutler et Mcqueen 2003 ; Tremblay et Spinelli 2014) et en production orale (Bagou et Laganaro 2009; Laganaro et al. 2010)[123]. Comme nous l'avons déjà évoqué dans les paragraphes précédents, deux hypothèses principales ont été formulées dans l'analyse du traitement et des représentations cognitives de la liaison. Les hypothèses abstractionnistes d'une part, et les hypothèses exemplaristes de l'autre.

2.1.1. Hypothèses abstractionnistes

De manière générale les modèles psycholinguistiques s'appuyant sur des hypothèses abstractionnistes postulent qu'à chaque mot correspond une représentation phonologique abstraite. Les hypothèses abstractionnistes de représentation de la liaison considèrent donc que la forme de liaison enregistrée dans la mémoire lexicale est unique. Autrement dit indépendamment de la réalisation ou de la non-réalisation de la liaison variable une seule forme est stockée dans la mémoire du locuteur natif. Comme nous l'avons déjà observé, Burki et Laganaro (2010) distinguent deux différentes hypothèses abstractionnistes. D'une part celle qui considère que la forme mémorisée correspond à la forme sans la CL. D'autre part celle selon laquelle la représentation stockée en mémoire contient la CL.

En revanche, l'approche abstractionniste proposée par Nguyen et al. (2007) (dans le cadre du modèle autosegmental) attribue à la notion de double flottance (cf. Partie 1.Chapitre 5.2.1.3) une valeur à la fois perceptive et cognitive. En effet, en s'appuyant sur les résultats issus de l'étude de Wauquier-Gravelines (1996), ces auteurs suggèrent que le coût plus important du traitement perceptif des CL (ex. /t/ dans « grand éléphant ») par rapport à celui des consonnes initiales (ex. /t/ dans « grand téléphone ») pourrait s'expliquer par le statut flottant de la CL.

Malgré la diversité des positions, toutes les approches abstractionnistes postulent l'existence d'une seule représentation lexicale de la liaison résultat d'un processus de catégorisation permettant de faire abstraction de la variabilité phonétique. Néanmoins, Nguyen et al. (2007) précisent qu'une telle démarche ne permet pas de rendre compte de la sensibilité

[123] Voir Burki et Laganaro (2014).

aux informations phonétiques fines démontrée par les études perceptives sur la liaison (cf. Paragraphe 1.5).

2.1.2. Hypothèses exemplaristes

En revanche, selon les hypothèses exemplaristes chaque mot possède des représentations multiples. Les approches exemplaristes postulent ainsi que la représentation de la liaison n'est pas unique puisque deux différentes formes (avec et sans CL) sont stockées dans le lexique mental. De plus, contrairement aux approches précédentes, l'hypothèse exemplariste attribue un rôle crucial à la fréquence des formes (avec et sans CL) pour un mot donné. Le coût de traitement varie ainsi en fonction de la fréquence d'usage des deux représentations.

Du point de vue de la production, Laks (2005b) dans le cadre de la phonologie auto-segmentale, propose à la suite de Bybee (2005) de considérer la notion de fréquence d'usage comme critère explicatif pour rendre compte de l'émergence de la variante de liaison ou de non-liaison. En effet la fréquence d'usage des variantes selon Bybee (2001b) influence la modalité de représentation des mots dans le lexique mental puisque chaque mot est stocké selon des catégories qui regroupent l'ensemble des variations dont le locuteur / auditeur a fait l'expérience. Dans ce sens, Racine (2008) souligne que l'approche proposée par Bybee (2001b) ne peut pas être considérée comme totalement exemplariste puisqu'elle fait appel à la notion de catégorisation afin de regrouper l'ensemble des exemplaires associés à un mot.

À la suite de McQueen, Cutler et Norris (2006), Racine (2008), avance qu'aucune des hypothèses extrêmes, totalement abstractionniste ou totalement exemplariste, ne permet de rendre compte des phénomène phonologiques tels que la liaison ou l'effacement du *schwa*[124].

3. Représentation de la liaison : entre phonologie et orthographe

Dans les paragraphes précédents nous avons présenté les traitements cognitifs associés à la reconnaissance des mots et des séquences écrits et oraux comme des processus basés uniquement sur un type d'information caractéristique de la modalité des données de l'input (écrit ou oral). Si l'existence d'une interaction entre phonologie et orthographe fait l'objet d'un consensus, les positions divergent quant au degré d'autonomie entre ces deux systèmes ainsi que sur les modalités et le niveau d'activation de ces interactions. Dans les pages qui suivent

[124] En raison de ces observations des hypothèses hybrides comme celle de Bybee (2001a) ont été proposées afin de conjuguer les deux approches.

nous présenterons les deux différentes hypothèses s'opposant dans la littérature. La première postule une autonomie entre orthographe et phonologie (*orthographic autonomy hypothesis*) sans pourtant rejeter l'idée d'une interaction optionnelle entre les deux systèmes. En revanche, la deuxième défend un recours automatique et obligatoire à la médiation phonologique dans la reconnaissance des mots écrits et plus en général une forte interaction entre phonologie et orthographe (*phonological mediation hypothesis*).

L'ambition de ce travail n'est pas de trancher entre les deux approches mais tout simplement d'analyser l'influence (qu'elle soit obligatoire ou optionnelle) de l'orthographe sur la reconnaissance des mots parlés (cf. Paragraphe 3.2.1) et sur la production orale (cf. Paragraphe 3.2.2).

3.1. Indépendance entre système oral (phonologie) et écrit (orthographe) (orthographic autonomy hypothesis)

La première approche (cf. Rapp, Benzing et Caramazza 1997) (cf. Figure 14) postule une autonomie entre le traitement orthographique et le traitement phonologique tout en reconnaissant une possible interaction entre les deux systèmes.

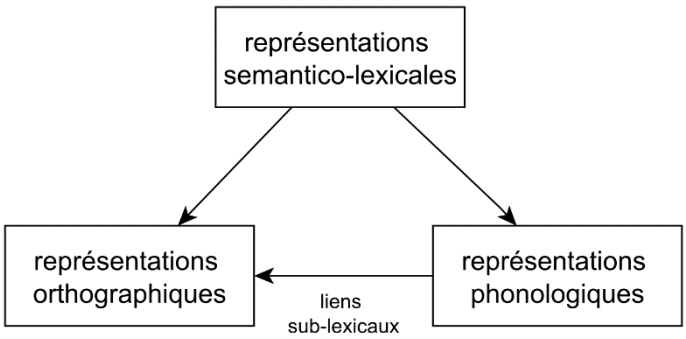


Figure 14 : Hypothèse de l'Indépendence entre système phonologique et orthographique (*orthographic autonomy hypothesis*) (tiré de Rapp, Benzing et Caramazza 1997)

Les connaissances orthographiques permettraient d'accéder directement au lexique mental sans qu'une médiation du système phonologique ne soit obligatoirement requise.

> « According to […] the orthographic autonomy hypothesis […], orthographic knowledge can make contact with the more abstract aspects of lexical knowledge (such as meaning and syntax) directly, without phonological mediation ». (Rapp, Benzing et Caramazza 1997, p. 75)

Cependant cette approche n'exclut pas la possibilité d'une médiation phonologique optionnelle et donc d'une éventuelle interaction entre orthographe et phonologie que ce soit au niveau lexical ou au niveau sub-lexical (cf. Figure 14).

> « It is important to note that we are considering the issue of whether or not lexical (or sublexical) phonology is required for written language processing. Thus, evidence in support of a direct relationship between meaning and orthography does not rule out the possibility of optional phonological mediation either by means of lexical connections between orthography and phonology or via sublexical processes ». (Rapp, Benzing et Caramazza 1997, p. 73)

De nombreux travaux neuropsychologiques (Zesiger et Partz 1994 ; Nespoulous et Dordain 1994 cité par Detey 2005, p. 466) apportent des arguments en faveur de cette hypothèse. En particulier celle-ci semble être confirmée par les résultats de plusieurs recherches neuro-psycolinguistiques menées auprès de sujets présentant des troubles du langage qui suggèrent une séparation entre les processus cognitifs de traitement de la parole de ceux de l'écrit. En s'appuyant sur le modèle à deux voies de la lecture (cf. Paragraphe 3.4), Zesiger et Partz (1994) étudient des cas pathologiques montrant une séparation entre l'altération du processus d'adressage et d'assemblage. Il en va de même pour les résultats de l'étude menée par Nespoulous et Dordain (1994 cité par Detey 2005, p. 466) auprès de patients aphasiques souffrant de lésions dans le cortex primaire (moteur, auditif et visuel) et donc de troubles du langage[125]. Différents types de séparation ont été observés auprès de ces patients. Ils peuvent : (a) comprendre une phrase qu'ils ne sont pas capables de produire (distinction compréhension / production) ; (b) répéter une phrase qu'ils ne peuvent pas comprendre (distinction répétition / compréhension) ; (c) énoncer oralement une phrase qu'ils ne peuvent pas écrire (distinction production orale / écrite); (d) lire / relire un énoncé qu'ils ont eux-mêmes écrit (distinction production / compréhension écrite) ; (e) mieux répéter que lire à voix haute, et vice-versa pour d'autres patients (cas de double dissociation) (distinction répétition / lecture à haute voix).

Dans une étude plus récente, Kouider et Dupoux (2001) s'interrogent sur l'autonomie des deux systèmes phonologique et orthographique à l'aide d'une tâche d'amorçage intermodale (visuel-oral). Ces auteurs essaient en particulier de vérifier si lors de l'exposition à des mots écrits les sujets activent de façon automatique et obligatoire les représentations phonologiques du lexique ou s'ils ont recours à un système de contrôle central (*central executive*) et donc à une activation consciente.

> « […] what is the relationship between the two processing systems in the literate adult? Is the reading processing system dependent from the auditory system, or has it acquired the status of an autonomous processing module […]. In order to distinguish between these hypotheses, it is important to control for the involvement of the central executive in the tasks at hand ». (Kouider et Dupoux 2001, p. 36-37)

[125] Anarthie pure, agraphie pure, surdité verbale et cécité verbale pure (voir Nespoulous 2002 cité par Detey 2005).

Les résultats de cette étude montrent que dans une tâche d'amorçage intermodale (visuel-oral), l'effet d'amorce n'a été enregistré que dans les cas où l'amorce est consciemment perçue. Ces données suggèrent que dans une situation de traitement cognitif inconscient des mots écrits, les sujets n'activent pas de façon automatique et obligatoire le lexique phonologique.

3.2. Implication inter-tâches entre système oral (phonologie) et écrit (orthographique) (phonological mediation hypothesis)

Quant à la deuxième hypothèse (*phonological mediation hypothesis*) (cf. Figure 15) l'accès au lexique mental serait soumis à une médiation obligatoire et automatique du système phonologique.

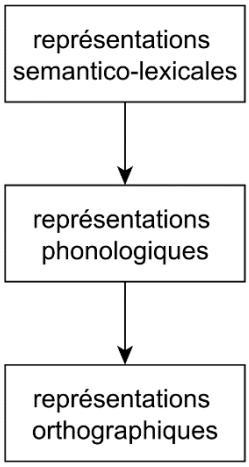


Figure 15 : Hypothèse de la médiation phonologique (*phonological mediation hypothesis*) (tiré de Rapp, Benzing et Caramazza 1997, p. 75)

> « According to the obligatory phonological mediation hypothesis, written language is entirely parasitic upon spoken language. As a consequence, in reading one must go from a written stimulus to a phonological representation before one can gain access to the meaning of a word. Similarly, in writing one must retrieve the spoken form of a word in order to gain access to the graphemic form ». (cf. Rapp, Benzing et Caramazza 1997, p. 73)

En effet, contrairement à l'approche précédente, plusieurs études psycholinguistiques suggèrent que non seulement il existe des interactions mutuelles entre le système phonologique et orthographique mais que celles-ci sont directes et automatiques (non optionnelles) et présentes au niveau inter-tâche, tant au niveau lexical que sub-lexical. Cela revient à dire que d'après ces travaux « la voie phonologique est impliquée dans des tâches pouvant en principe être effectuées sans elle » (Detey 2005, p. 468). Comme le font observer Spinelli et Ferrand (2005) entre autres, langage écrit (orthographe) et oral (phonologie) interagissent entre eux dans le mesure où « la lecture silencieuse fait appel à l'information phonologique des mots » et

« l'information orthographique intervienne dans la reconnaissance des mots parlés » (Spinelli et Ferrand 2005, p. 183). Les études ayant recours à des tâches de détection bimodale permettent notamment, et contrairement aux travaux consacrés à la reconnaissance visuelle lexicale, de vérifier tant l'influence des phonèmes sur les graphèmes que l'inverse (Detey 2005, p. 470)[126].

3.2.1. Influence de l'orthographe dans la reconnaissance des mots parlés

L'hypothèse de l'influence de l'orthographe dans la reconnaissance des mots parlés présuppose que les représentations orthographiques sont activées lors du traitement auditif de la langue parlée. Cette activation implique d'après Spinelli et Ferrand (2005, p. 207) que « la connaissance de l'orthographe conduirait à l'élaboration de procédures servant à l'analyse consciente des mots ». Plusieurs études psycholinguistiques (Hallé et al. 1998 ; Dupoux et al. 1998 ; Dupoux et al. 1999 ; Hallé, Chéreau et Segui 2000 ; Racine, Bürki et Spinelli 2014) font état de cette influence.

En particulier, l'étude menée par Hallé et al. (2000) a montré que dans une tâche de détection de phonèmes les sujets percevaient davantage le phonème /b/ dans un mot comme « absurde » que dans « capsule » bien que « absurde » soit prononcé [apsyrd] suite à une assimilation. L'influence de l'orthographe dans le traitement auditif de ces items est d'autant plus évidente si on considère que la détection du phonème /b/ dans le mot « absurde » ne se manifeste pas dès le début mais n'a lieu que lorsque l'information lexicale est accessible.

Racine et Grosjean (2005) et Racine (2008) ont également montré, par le biais d'une tâche de décision lexicale, que les variantes avec effacement du *schwa* [lakasʀɔl] de la séquence « la casserole » bien qu'elle soit plus fréquente que celle sans effacement [lakasəʀɔl], n'est pas reconnue plus rapidement que la forme sans effacement. En revanche dans le cas où les deux formes, avec et sans effacement, ont une fréquence d'occurrence similaire (ex. [ʃmɛ̃] et [ʃəmɛ̃] « chemin ») ou que la forme avec effacement a une fréquence d'occurrence basse (ex. [tuʀnsɔl] « tournesol »), la forme sans effacement est reconnue plus rapidement. D'après les auteurs, ces résultats montreraient que la fréquence d'occurrence des variantes ainsi que les informations graphiques sont représentées dans le lexique mental du locuteur. L'influence de l'orthographe permet en effet d'expliquer la raison pour laquelle des variantes avec effacement du *schwa*

[126] Dans le cadre de ce travail nous nous attacherons à étudier uniquement ce deuxième type d'interaction : Pour un bilan de l'influence de la phonologie sur la reconnaissance des mots voir (Frost 1998), Detey (2005), Spinelli et Ferrand (2005).

[lakasʀɔl] ayant une fréquence plus importante par rapport aux variantes sans effacement [lakasəʀɔl] ne sont pas reconnues plus rapidement[127].

Ces conclusions sont confirmées par les travaux de Racine et al. (2014) et montrent, à l'aide d'une tâche de détection menée auprès de deux groupes d'enfants, lecteurs et pré-lecteurs, que l'orthographe joue un rôle crucial dans la reconnaissance des variantes d'effacement du *schwa*. Racine et al. (2014) montrent que dans des cas d'effacement obligatoire du *schwa*, contrairement aux enfants qui ont appris à lire, les enfants pré-lecteurs présentent des taux de réaction plus importants pour la reconnaissance des variantes avec *schwa* ([bʁasəlɛ] « bracelet ») que des variantes où cet élément est effacé ([bʁaslɛ]). Ces résultats suggèrent qu'en l'absence d'une connaissance de l'orthographe, les enfants pré-lecteurs font uniquement référence à la variante dont l'usage est plus fréquent ce qui ralentit leur capacité de reconnaître la variante avec *schwa* ([bʁasəlɛ]) plus proche de la forme orthographique « bracelet ». En revanche dans des contextes d'effacement variable ([ʁənaʁ] « renard ») aucune différence n'a été enregistrée entre les enfants pré-lecteurs et lecteurs dont les représentations seraient influencées, dans les deux cas, par la fréquence de la variante en question[128].

3.2.2. Influence de l'orthographe dans la production orale

L'influence de l'orthographe sur la production de la parole a été peu abordée par la littérature (Damian et Bowers 2003 ; Alario et al. 2007). Damian et Bowers (2003) ont expérimenté l'hypothèse selon laquelle l'orthographe peut conditionner la production de mots isolés. Les auteurs ont donc montré, à l'aide d'un test d'amorçage mené auprès de sujets anglais que les items commençant par le même phonème (ex. /k/) mais correspondant à des épellations différentes (« coffee» ; « kennel ») étaient reconnus moins rapidement que les items partageant la même épellation (« coffee» ; « camel »).

> « Importantly, however, sets in which response words share their initial segment, but differ in their spelling, do not receive any benefit. This finding suggests an effect of orthographic codes in speech production ». (Damian et Bowers 2003, p. 124).

Cependant ces données ne semblent pas confirmées par l'étude de Roelofs (2006) auprès de sujets néerlandais ayant utilisé le même paradigme expérimental. Si Roelofs (2006) attribue ces résultats contradictoires aux différences entre les systèmes orthographiques de l'anglais et

[127] Des résultats similaires ont été obtenus par d'autres travaux portant sur des phénomènes de variation phonologique issus d'autres langues (pour un bilan voir Racine 2008).

[128] D'autres arguments indirects viennent de l'influence pré-lexicale des connaissances lexicales sur la phonologie attestée par des recherches (Dupoux et al. 1998 ; Dupoux et al. 1999) menées auprès des apprenants japonais de français langue étrangère. Ces auteurs ont montré, à travers quatre études, que le traitement perceptif des apprenants japonais du français était influencé par les propriétés phonotactiques et le système syllabique de leur L1. Cela les amenait à percevoir des voyelles « illusoires » au sein des groupes consonantiques.

du néerlandais, Alario et al. (2007) font plutôt référence à un biais méthodologique. En effet le paradigme expérimental utilisé par les études de Damian et Bowers (2003) et Roelofs (2006) demande au sujet de mémoriser des paires des mots à reproduire pendant le test d'amorçage lorsque le mot cible s'affiche à l'écran. Il est donc possible que les sujets aient exploité les indices orthographiques afin de mémoriser les items et que ces stratégies se reflètent dans la production.

> « In the initial learning phase of the experiment, participants have to memorize cue-target associations. It is entirely possible that the orthographic consistency effect has its locus in word-pair learning rather than in word-production, if participants use an orthographic code to improve memorization. This is a plausible hypothesis given that orthographic codes can be implicitly activated even when words are presented only in their spoken form ». (Alario et al. 2007, p. 466‑467)

Afin d'éviter cet éventuel effet de mémorisation, Alario et al. (2007) tout en adoptant le même paradigme que Damian et Bowers (2003) et Roelofs (2006), utilisent une tâche de dénomination d'image. L'absence d'effets significatifs de l'orthographe sur la production des mots ayant le même phonème initial mais une épellation différente suggère que les résultats obtenus par Damian and Bowers (2003) pourraient être liés à un biais méthodologique.

> « This result suggests that, even in a language where the mapping between phonology and spelling is at least as inconsistent as English, orthography does not affect production. The absence of orthographic effect in picture naming suggests that Damian and Bowers' (2003) results might be due to the word-association learning phase they employed to elicit word production ». (Alario et al. 2007, p. 470)

3.3. Bilan

Si le rôle de l'orthographe dans la production des natifs ne semble pas aujourd'hui faire l'objet d'un consensus, de solides arguments ont cependant été apportés en faveur d'une influence orthographique dans des tâches de reconnaissance des mots (cf. Paragraphe 3.2.1). Néanmoins les modèles perceptifs que nous avons brièvement évoqués au début de ce chapitre (cf. Paragraphe 1.1), comme le font observer Racine et al. (2014), ne permettent pas de rendre compte de la présence d'interactions existant entre phonologie et orthographe. Ces limites nous ont donc amené à choisir le modèle interactif bimodal proposé par Grainger et Ferrand (1996) comme cadre théorique de référence pour notre étude.

3.4. Modèle interactif bimodal

Contrairement aux modèles précédemment évoqués (Cohorte, TRACE), le modèle interactif bimodal (Grainger et Ferrand 1996 ; Ziegler, Muneaux et Grainger 2003) parvient à représenter les connexions entre phonologie et orthographe en perception et production tant au niveau lexical que sub-lexical (cf. Figure 16). Dans le cadre des modèles interactifs les connexions établies entre les différents niveaux ont une nature facilitatrice.

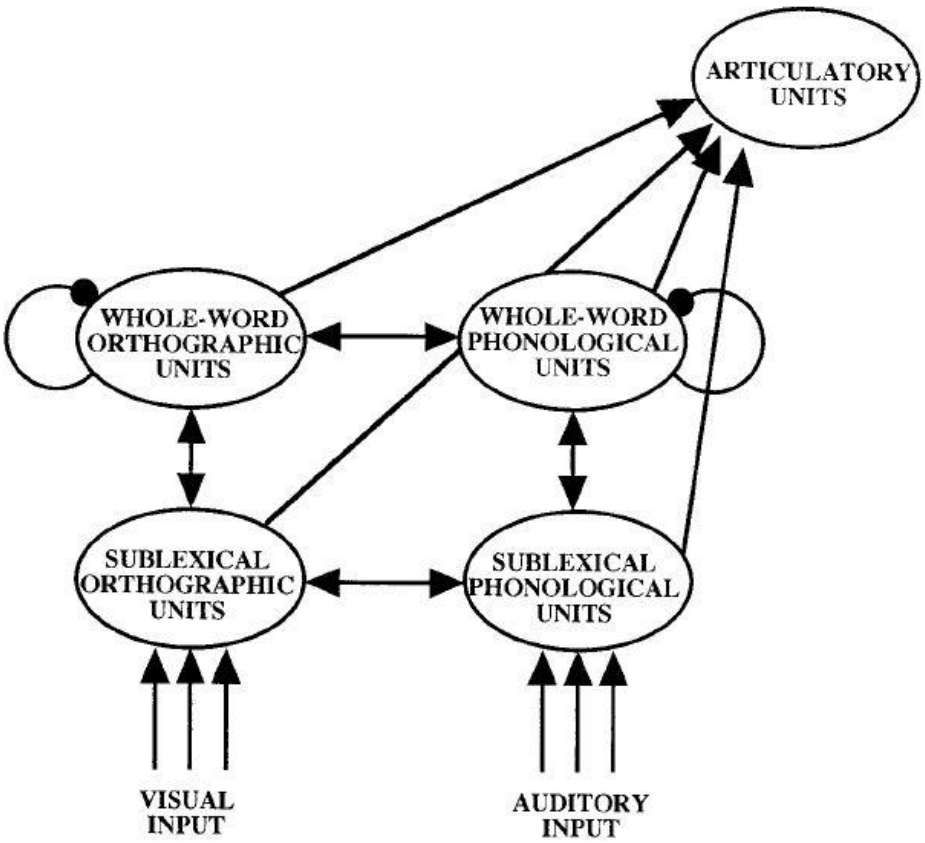


Figure 16 : Modèle interactif bimodal (Grainger et Ferrand 1996 ; Ziegler, Muneaux et Grainger 2003)

Ainsi les unités du niveau phonologique sublexical (*sublexical phonological units*) permettent d'activer d'autres unités relevant du niveau orthographique (ex. *sublexical orthographic units*). Une telle approche, comme nous l'avons observé, permet de rendre compte du rôle joué par le lexique mental orthographique (*sublexical orthographic units)* dans les processus de reconnaissance et production.

> « Within this theoretical framework it is hypothesized that responses in both the visual lexical decision and the perceptual identification tasks are based on activity in the orthographic lexicon (but see Grainger & Ferrand 1994, for cases where decisions may be based on activity in the phonological lexicon). […] The bidirectional connections between orthographic and phonological units in the model allow it to handle the early effects of phonology observed in

> both the visual lexical decision and the perceptual identification tasks […] (and also to explain orthographic influences in auditory tasks, e.g., Dijkstra, Frauenfelder, & Schreuder, 1993; Seidenberg & Tanenhaus, 1979) ». (Grainger et Ferrand 1996, p. 426)

De la même façon ce modèle pourrait permettre de rendre compte des effets de l'orthographe sur la production de la parole chez les natifs (cf. Paragraphe 3) ainsi que chez les non-natifs (cf. Partie 4.Chapitre 3.1.3.1)[129].

> « On the other hand, it is hypothesized that responses in the word naming task are governed by activity in articulatory output units, which receive excitatory input from all compatible lexical and sublexical units in the network ». (Grainger et Ferrand 1996, p. 426)

4. Conclusion

À l'issue de ce bref survol des travaux consacrés à l'analyse des processus cognitifs qui guident la reconnaissance et la production de la liaison il paraît évident que l'apport des études psycholinguistiques est crucial pour la compréhension des représentations cognitives de la liaison et donc également du statut phonologique des CL. De plus une approche capable de conjuguer les contributions issues des recherches psycholinguistiques et les données issues de la linguistique de corpus pourrait aboutir à une meilleure compréhension des modalités et des niveaux d'activation des interactions existant entre la phonologie et l'orthographe.

Afin de mieux analyser le statut des CL ainsi que sa nature hétérogène, nous nous pencherons désormais sur une étude empirique dans laquelle nous développerons une analyse fréquentielle des usages de la liaison tels qu'ils sont observés dans le corpus PFC (cf. Partie 2.Chapitre 2).

[129] Nous reviendrons sur cet aspect dans la deuxième et quatrième partie de ce travail.

Partie 2. ANALYSE FREQUENTIELLE DE LA LIAISON DANS UN CORPUS DE FRANÇAIS PARLE

Chapitre 1. Le corpus « Phonologie du Français Contemporain » (PFC) : un corpus de français parlé codé en liaison et annoté en morphosyntaxe

Comme nous avons pu le constater dans la première partie de ce travail, la liaison est un phénomène complexe qui a été au centre de nombreux débats théoriques qui ont fait souvent référence à des sources normatives (cf. Partie 1.Chapitre 3.2.1) ou à des données obtenues par introspection. Cependant la fin du XXème siècle a été caractérisée par un nouvel intérêt porté aux approches variationnistes en phonologie grâce à l'émergence de nouvelles technologies issues du traitement automatique des langues (TAL) pour la constitution et l'analyse de grands corpus (Laks 2008). C'est dans ce cadre théorique que s'inscrit et se développe le projet « Phonologie du Français Contemporain » (Durand, Laks et Lyche 2009) (PFC) dans le but de recueillir un important corpus de référence du français parlé capable de rendre compte de la diversité d'usage du français parlé au sein de l'espace francophone.

Dans la première partie de ce chapitre nous proposons une brève introduction du projet PFC et de ses objectifs et méthodologies. Nous présenterons par la suite le travail de ré-annotation morphosyntaxique que nous avons mené sur le corpus et qui nous a permis de réaliser une analyse fréquentielle des productions de la liaison variable, à laquelle sera consacré le prochain chapitre.

1. Brève présentation du projet PFC

Le projet international PFC est actuellement codirigé par Marie-Hélène Côté (Université Laval), Jacques Durand (ERSS, Université de Toulouse-Le Mirail), Bernard Laks (MoDyCo, Université de Paris Ouest) et Chantal Lyche (Universités d'Oslo et de Tromsø)[130]. Il a été lancé en 1999 et a abouti aujourd'hui à la constitution d'un des plus grands corpus de français parlé issu non seulement de la France hexagonale mais également de différents pays francophones (Belgique, Suisse, Québec, Afrique francophone, Louisiane etc.). Cette enquête sociolinguistique de terrain a ainsi permis d'élaborer une base de données en mesure de documenter les principaux phénomènes phonologiques de la langue française (à savoir la liaison et le *schwa*) dans les différentes variétés du français parlé dans l'espace francophone.

[130] http://www.projet-pfc.net/

Un tel corpus peut être exploité non seulement dans le cadre des domaines de recherches linguistiques (phonétique, phonologie, syntaxe, pragmatique, sociolinguistique, analyse conversationnelle, etc.), mais aussi bien dans l'enseignement et l'apprentissage du français (langue maternelle et seconde) ou tout simplement dans la diffusion des savoirs (conservation du patrimoine linguistique francophone et présentation générale du français oral contemporain pour les non-spécialistes).

1.1. Objectifs

Parmi les objectifs que le projet PFC se propose, nous pouvons citer les suivants, contenus dans Durand, Laks et Lyche (2009 ; 2014) :

- fournir une meilleure image du français parlé dans son unité et sa diversité ;
- mettre à l'épreuve les modèles phonologiques et phonétiques sur le plan synchronique et diachronique ;
- favoriser les échanges entre les connaissances phonologiques et les outils du traitement automatique de la parole ;
- permettre la conservation d'une partie importante du patrimoine linguistique des espaces francophones du monde ;
- et encourager un renouvellement des données et des analyses pour l'enseignement du français.

1.2. Points d'enquêtes

Afin de réaliser ces objectifs les diverses équipes de recherche de PFC ont aujourd'hui couvert 38 points d'enquête dans l'espace francophone international. Le tableau ci-dessous (cf. Tableau 2) résume l'ensemble des points d'enquête ainsi que le nombre de tokens correspondants.

Code	**Point d'enquête**	**Tokens**	**Code**	**Point d'enquête**	**Tokens**
FRANCE			**AFRIQUE**		
11a	Douzens	38.693	aba	Béjaia	24.607
12a	Rodez	42.233	aca	Chlef	23.141
13a	Marseille Centre Ville	32.000	bfa	Burkina Faso	25.451
13b	Aix- Marseille	18.164	cia	Abidjan	44.347
21a	Dijon	31.233	cya	Cameroun	14.075
31a	Toulouse	48.533	maa	Bamako	76.577
38a	Grenoble	19.989	rca	Bangui	42.448

42a	Roanne	24.868	sna	Sénégal Dakar	23.479
44a	Nantes	24.180	**CANADA**		
50a	Brécey	36.009	caa	Peace River	22.455
54b	Ogéviller	26.482	cqa	Québec ville	27.967
61a	Domfrontais	68.954	cqb	Saguenay	119.348
64a	Biarritz	36.519	**BELGIQUE**		
69a	Lyon	23.093	bga	Gembloux	31.619
75c	Paris Centre Ville	54.097	bla	Liège	28.153
75x	Aveyronnais à Paris	49.711	bta	Tournai	26.611
81a	Lacaune	22.115	**SUISSE**		
85a	Vendée	24.610	sca	Neuchâtel	60.547
91a	Brunoy	24.917	sga	Genève	41.104
92a	Puteaux-Courbevoie	11.315	sva	Nyon	50.461
974	Île de la Réunion	27.939			
			TOTAL CORPUS PFC		**1.368.044**

Tableau 2 : Points d'enquête du corpus PFC

1.3. Méthodologies

1.3.1. Protocole d'enquête

Chaque enquête a été menée à partir d'un protocole commun soumis à un ensemble d'environ 12 locuteurs répartis en deux groupes égaux d'hommes et de femmes représentatifs idéalement de différentes tranches d'âge (18-30 ans, 30-45 ans, 45-60 ans et plus de 60 ans)[131] et, dans la mesure du possible, de différentes catégories socio-culturelles[132]. De plus la sélection des informateurs est réalisée « par proximité en contact selon la technique des réseaux denses illustrée notamment par Milroy 1987 » (Durand, Laks et Lyche 2009, p. 27). Elle se base sur une connaissance préalable du terrain de la part des enquêteurs qui sélectionnent uniquement les locuteurs solidement ancrés dans leur communauté de référence et / ou ayant effectué la quasi-totalité de leur scolarité dans la région. Des métadonnées sont fournies pour chaque locuteur de chaque point d'enquête à l'aide d'une fiche signalétique détaillée (cf. Annexe 3) assurant la comparabilité des données.

[131] Au moins deux tranches d'âge.

[132] Toutefois, il convient de préciser que le projet PFC ne vise « aucune représentativité explicitement construite sur la base de recensements sociaux démographiques ou autres […]. La variation sociale est pourtant bien représentée dans l'échantillon, par le nombre et la diversité des points d'enquêtes et par la consigne donnée aux enquêteurs d'équilibrer à chaque point les sexes et les âges ainsi que, autant que faire se peut, en restant dans le réseau communautaire considéré, les professions, les niveaux scolaires et les catégories sociales » (Durand, Laks et Lyche 2009, p. 29).

Le protocole d'enquête s'articule en deux volets soumis à tous les locuteurs : une lecture de documents et une conversation. Le premier comprend d'une part la lecture à haute voix d'une liste de 94 mots visant principalement à établir les inventaires phonémiques du locuteur (cf. Annexe 4). D'autre part, un texte correspondant à un bref article (cf. Annexe 4) a été fabriqué par l'équipe de direction du projet PFC[133] afin d'analyser les phénomènes caractéristiques de la phonologie du français à savoir le *schwa*, la liaison, la nasalisation, le « h » aspiré, la loi de position et la chute des liquides finales. Ce texte permet également de dégager l'inventaire phonémique du locuteur reprenant des paires minimales contenues dans la liste des mots dans un style toujours formel mais moins surveillé que celui de la tâche précédente. Ce dernier aspect offre des possibilités intéressantes pour une comparaison entre les deux types de production orale comportant un degré de contrôle des réalisations différent de la part du sujet[134]. Le deuxième volet consiste en une conversation articulée en deux parties : une entrevue formelle (entretien guidé) correspondant à une interaction face à face conduite par deux enquêteurs[135] d'une vingtaine de minutes par locuteur et une discussion informelle (conversation libre) entre pairs d'une trentaine de minutes par locuteur et enregistrée lors d'une réunion.

1.3.2. Traitement et transcription des données

L'ensemble des données ainsi recueillies, comprenant les deux tâches de lecture, 5 minutes au minimum de l'entretien guidé et 5 minutes de la discussion libre, sont transcrites en orthographe standard[136] à l'aide du logiciel Praat (Boersma et Weenink 2015) et alignées au signal sonore par groupe rythmique. Aucune modification prenant en compte la prononciation du locuteur n'a été insérée dans la transcription orthographique. Cela se justifie par une volonté d'assurer une comparabilité des résultats et la réalisation d'indexation et d'exploration des données. L'annotation phonétique n'est pas adoptée au sein de ce projet à grande échelle qui pour des raisons de taille ne peut pas assurer le degré de précision et de cohérence requis pour une transcription phonétique. Outre le désaccord inter-transcripteurs, (comme nous le verrons dans la quatrième partie de ce travail cf. Partie 4.Chapitre 1.2.4), le recours à une transcription phonétique implique inévitablement un degré d'abstraction à partir du signal et donc un filtrage

[133] Bien que ce texte présente des traits qui le caractérisent comme issu de la France hexagonale, il reste tout à fait compréhensible et réalisable par n'importe quel locuteur francophone. Il n'est marqué ni sur le plan politique ni sur le plan religieux et se présente comme un simple article de journal régional ou local.

[134] Si d'une part l'inclusion de la lecture d'un texte dans le protocole PFC limite l'accès aux sujets moins lettrés (ce qui implique une perte de représentativité), d'autre part, en ligne avec les objectifs du projet PFC, elle assure une comparaison entre l'inventaire phonologique de la liste des mots et celui enregistré pendant la lecture d'un texte.

[135] L'un des deux enquêteurs doit appartenir au réseau dense du locuteur.

[136] Tout en annotant les disfluences (hésitations, faux départs, répétitions, pause remplies etc.) (cf. Durand, Laks et Lyche 2013).

de la part de l'annotateur[137]. Une couche zéro d'annotation orthographique est donc proposée par le projet PFC tout en autorisant des travaux successifs sur des détails phonétiques ou acoustiques plus fins. En effet, la possibilité offerte par le logiciel Praat d'ajouter des tires d'annotation complémentaires permet d'annoter sur une couche séparée des phénomènes tels que la liaison et le *schwa* à l'aide de systèmes de codages spécifiques définis par le protocole du projet.

1.3.3. Codage et analyse de la liaison

L'étude de la liaison est au cœur des recherches menées par le projet PFC qui grâce au recours à un codage alphanumérique permet d'annoter de façon globale les sites de liaison potentielle sur une tire indépendante par rapport à celui de la transcription orthographique. Ce système de codage a été conçu pour des non-spécialistes visant à une analyse holistique[138] des données capable de décrire les grandes typologies de liaison (à savoir catégorique, variable et erratique cf. Partie 1.Chapitre 4.3) ainsi que les liaisons épenthétiques (liaisons non attendues et les cas de substitutions des CL). Le système de codage alphanumérique de la liaison adopté par le corpus PFC comprend deux symboles numériques obligatoires et d'autres symboles alphabétiques qui varient en fonction du type d'occurrence de la liaison. Dans tout le système de codage l'ordre des symboles est fixe.

Le premier chiffre indique le nombre de syllabes du M1 :

- 1 = une syllabe ;
- 2 = deux syllabes ou plus.

Le deuxième, en revanche, donne des informations sur les types de réalisation de la liaison (absence de liaison, enchaînement, non-enchaînement ; incertitude quant à l'enchaînement) et la présence d'une liaison épenthétique :

- 0 = absence de liaison ;
- 1 = liaison enchaînée ;
- 2 = liaison non enchaînée ;
- 3 = incertitude ;
- 4 = liaison épenthétique.

[137] Cet aspect sera approfondi dans la quatrième partie de ce travail (cf. Partie 4.Chapitre 1.2.4.1).

[138] Une telle approche d'analyse globale ne cherche donc pas à annoter des détails tels que la qualité de la voyelle précédant la CL.

Après cette suite de codes numériques, vient s'ajouter une série de codes alphabétiques en fonction du type d'occurrence de liaison en suivant toujours un ordre fixe. Un premier code, pris dans l'ensemble des symboles consonantiques du français écrit {b,…z, B,….Z}, est inséré dans les cas de réalisation de la liaison afin d'annoter la nature de la CL (conforme ou épenthétique)[139].

D'autres codes alphabétiques sont ajoutés en ordre fixe :

- H = dans les cas d'hésitation et / ou coup de glotte ;
- C = dans le cas de réalisation d'une CL inattendue devant un mot à initiale consonantique ;
- VN = pour indiquer la nature nasale de la voyelle précédente la CL /n/ ;
- VO = pour indiquer la nature orale de la voyelle précédente la CL /n/ ;

L'exemple ci-dessous correspond à un extrait d'annotations d'occurrences de liaisons contenues dans le texte fabriqué (cf. Paragraphe 1.3.1).

Transcription orthographique[140] :
« Le maire de Beaulieu, Marc Blanc, est [ɛ̃] en revanche très [trɛzɛ̃kjɛ] inquiet »
Codage liaison :
« Le maire de Beaulieu, Marc Blanc, est**10** en revanche très**11z** inquiet »
Interprétation du codage :
« Est10 en » = M1 monosyllabique (1) + absence de liaison (0) ;
« Très11z inquiet » = M1 monosylabique (1) + liaison enchaînée (1) en [z].

Exemple 1 : Énoncé annoté en liaison par le biais du codage liaison PFC

Une telle annotation permet d'effectuer des recherches rapides dans la base de données à partir des attributs décrits par le codage alphanumérique.

2. Annotations morphosyntaxiques pour l'étude de la liaison

Les possibilités d'annotation offertes par les codages liaison ne permettent pas [141] toutefois de préciser les informations concernant les catégories morphosyntaxiques d'appartenance des M1 et M2 de liaison à savoir le type de liaison (cf. Partie 1.Chapitre 4.1). Afin d'assurer une analyse non seulement phonologique mais également syntaxique et lexicale du contexte de liaison, une annotation en partie de discours (*Part of speech* dorénavant POS) a

[139] En cas d'absence de liaison, aucun symbole n'est utilisé.
[140] La transcription phonétique a été ajoutée uniquement dans le but d'illustrer le système de codage.
[141] Pour les mêmes raisons que nous avons déjà évoquées dans le cas de la transcription phonétique (cf. Partie 4.Chapitre 1.2.4.1).

été réalisée sur l'ensemble du corpus PFC. Dans un premier temps le corpus a été annoté par le biais de l'annotateur morphosyntaxique *Treetagger* (Schmid 1994) [142] principalement utilisé pour l'étiquetage morphosyntaxique des corpus écrits de plusieurs langues[143]. Ensuite, à partir de 2014, nous avons réalisé un important travail de ré-annotation du corpus à l'aide de l'annotateur *DisMo*[144] (Christodoulides, Avanzi et Goldman 2014 ; Christodoulides, Barreca et Avanzi 2015), un annotateur multi-niveaux conçu pour faire face aux spécificités des corpus oraux telles que les marqueurs de discours et les disfluences.

2.1. Première annotation morphosyntaxique à l'aide de l'étiqueteur Treetagger (Schmid 1994)

Un premier travail d'annotation morphosyntaxique et lemmatisation a été effectué à l'aide de l'étiqueteur *Treetagger* (Schmid 1994) sur l'ensemble du corpus PFC. Cet outil a fourni une première annotation des contextes morphosyntaxiques de liaison ainsi qu'une lemmatisation des M1 et M2. Bien que cette première campagne d'annotation ait permis d'identifier les types de liaison (cf. Partie 1.Chapitre 4.1) et donc les grandes typologies de liaison (catégorique, variable et erratique), elle présentait quelques inconvénients qui limitaient l'analyse syntaxique. En effet, étant un annotateur conçu pour les corpus écrits de plusieurs langues[145], *Treetagger* (Schmid 1994) n'était pas en mesure de traiter des phénomènes spécifiques de l'oral à savoir les disfluences et les marqueurs de discours. De plus les modèles statistiques utilisés[146] ainsi que le recours à une base de données lexicale du français assez pauvre donnaient lieu à une annotation automatique peu fiable des mots composés (ex. « États-Unis ») et des adverbes composés (ex. « de plus en plus »).

L'identification et la correcte annotation de ces unités poly-lexicales s'avèrent cruciales pour l'analyse syntaxique de ces contextes de liaison qui, comme nous l'avons déjà souligné (cf. Partie 1.Chapitre 4.2.2), donnent lieu à des réalisations de liaison catégorique. Le degré de précision de 96,36% de l'annotation automatique fournie par *TreeTagger* (Schmid 1994) ainsi que l'absence d'un corpus d'entraînement de français corrigé manuellement contribuaient

[142] http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger.

[143] Des fichiers paramètres pour l'annotation automatique de corpus de français parlé ont été plus récemment proposés par Benzitoun et al. (2012) afin d'adapter *TreeTagger* (utf-8) à l'annotation du français oral. Des ressources lexicales du français oral obtenues par fusion du corpus du « Projet d'Etiqueteur Robuste pour l'Ecrit et pour l'Oral » (PERCEO oral) et de « Morphalou » ont été également intégrées à *TreeTagger* afin d'améliorer les performances d'annotation des corpus de français oral.

[144] http://corpusannotation.org/

[145] Parmi celles-ci nous pouvons citer : l'allemand, l'anglais, le français, l'ancien français, l'italien, le néerlandais, l'espagnol, le bulgare, le russe, le portugais, le galicien, le chinois, le swahili, le slovène, le latin, l'estonien et le polonais.

[146] Basés sur des arbres de décision.

souvent à une annotation imprécise des sites de liaison. C'est la raison pour laquelle nous avons décidé de mettre en place une nouvelle campagne d'annotation morphosyntaxique du corpus PFC réalisée à l'aide de l'étiqueteur *DisMo*.

2.2. Un travail de ré-annotation morphosyntaxique à l'aide de l'étiqueteur DisMo

Plusieurs étiqueteurs morphosyntaxiques sont aujourd'hui disponibles en libre source pour l'annotation automatique en POS des corpus. Parmi ceux-ci nous avons déjà évoqué *TreeTagger* (Schmid 1994) et nous pouvons également citer *Stanford Tagger* (Toutanova et al. 2003) et *MElt* (Denis et Sagot 2009) Il s'agit dans la plupart des cas d'étiqueteurs conçus pour l'écrit qui ont été, dans certains cas, utilisés pour l'annotation des corpus oraux à la suite d'un prétraitement des transcriptions des données orales (Blanc et al. 2008). Certains auteurs ont aussi opté pour des solutions proposant aussi bien un apprentissage automatique à partir de données corrigées manuellement (Eshkol et al. 2010) que des fichiers de paramétrages spécifiques pour les données orales (Benzitoun, Fort et Sagot 2012). Ces derniers ont dû faire face aux problèmes posés par l'absence de signes de ponctuation et donc de frontières de phrases dans les transcriptions de l'oral et qui, en revanche, s'avèrent nécessaires pour l'utilisation des étiqueteurs conçus pour l'écrit. Une segmentation en unités de phrases a été donc établie sur la base des unités comme le groupe de souffle ou simplement le tour de parole. Ces travaux ont donc le mérite d'avoir fourni des solutions pour l'annotation de l'oral sans renoncer à intégrer dans ce traitement les caractéristiques des corpus oraux telles que les disfluences, les répétitions, les reformulations, les amorces de mots ainsi que les marqueurs discursifs. En particulier, ces derniers phénomènes ont été annotés au niveau micro-syntaxique en utilisant des étiquettes crées *ad hoc* afin d'« isoler » ces éléments et d'améliorer ainsi les performances des annotateurs conçus pour l'écrit.

À la suite de ces travaux, afin de parvenir à une description plus précise du contexte syntaxique de réalisation de la liaison dans des corpus de français oral, nous avons mené en collaboration avec George Christodoulides de l'Université Louvain la Neuve une nouvelle campagne d'annotation du corpus PFC à l'aide de l'annotateur *DisMo* (Christodoulides, Avanzi et Goldman 2014 ; Christodoulides, Barreca et Avanzi 2015). Il s'agit d'un étiqueteur morphosyntaxique directement conçu pour l'oral qui s'articule sur plusieurs niveaux d'annotation permettant d'accorder aux spécificités de l'oral une place plus importante dans l'étiquetage des données d'oral spontané. Cet outil ne demande pas de prétraitement ou d'adaptation spécifique des données de l'oral puisqu'il accepte comme input des transcriptions

de l'oral dépourvues de signes de ponctuation ou de frontières de phrases. D'autre part, l'annotation à l'aide de cet outil permet non seulement de détecter les phénomènes spécifiques de l'oral spontané mais aussi de les étiqueter dans des couches d'annotation autres que la couche morphosyntaxique. En effet, *DisMo* fournit un output qui s'articule en six niveaux d'annotation :

- unités lexicales minimales (*tok:min* à savoir *Minimal Tokens*) ;
- unités polylexicales (*tok:mwu* à savoir *Multi-word units*) ;
- étiquettes morphosyntaxiques des unités lexicales minimales (*pos:min* à savoir *POS tag of minimal tokens*) ;
- étiquettes morphosyntaxiques des unités poly-lexicales (*pos:mwu* à savoir *POS tag of multi-word units*) ;
- difluences (*disfluencies*) ;
- marqueurs de discours (*discourse*).

La première couche d'annotation (*tok:min*) est le résultat d'un prétraitement et découpage du corpus en unités lexicales minimales (*tokenisation*). Par la suite l'application de ressources linguistiques[147] permet de détecter les disfluences (*difluencies*), les marqueurs discursifs (*discourse*) ainsi que les unités poly-lexicales (*tok:mwu*) potentiels. Cette phase est suivie par une annotation morphosyntaxique préliminaire des unités lexicales minimales (*pos:min*). Pour conclure *DisMo* réalise l'annotation des difluences et l'annotation morphosyntaxique finale qui comprend également l'étiquetage des unités poly-lexicales (*pos-mwu*)[148].

Ce traitement multi-niveaux assure une étude systématique de l'ensemble des phénomènes qui caractérisent la parole spontanée à l'aide d'un seul outil permettant également d'améliorer la qualité des résultats des requêtes ayant comme attribut les catégories morphosyntaxiques.

Le travail de ré-annotation que nous venons de décrire a été réalisé en plusieurs étapes. Après une première annotation automatique, une phase de post-traitement des données a été mise en place.

Les modèles statistiques de *DisMo* ont été entraînés d'abord sur le corpus CPROM-PFC (Avanzi 2014), qui contenait lui-même des échantillons du corpus PFC. Deux annotateurs

[147] Il s'agit notamment de ressources lexicales basées sur les dictionnaires DELA (Courtois et al. 1997), GLÀFF (Sajous et al. 2013) et des dictionnaires des unités nommées crées par les auteurs.

[148] Toutefois, il convient de souligner que ces unités polylexicales (ex. « tout à fait ») dans le premier niveau d'annotation minimale sont également analysées sur la base des lexèmes minimaux (*tok:min*) qui les composent (ex. « tout », « à » et « fait »).

experts ont corrigé l'annotation de 57 mille tokens. Ces premiers modèles ont été utilisés pour annoter les entretiens semi-guidés de trois points d'enquête du corpus PFC (Paris, Aveyronnais et Lyon), soit 127 mille tokens, et un annotateur expert a corrigé manuellement ces échantillons. L'ensemble de ces données nous a permis d'entraîner des modèles statistiques finaux. Après ces campagnes d'annotation et de correction, la précision du système est de l'ordre de 97% pour le jeu d'étiquettes complet, et de 98% pour un jeu d'étiquettes simplifié.

De plus cette ré-annotation, contrairement à la précédente, nous a permis d'identifier avec une plus grande précision les unités poly-lexicales et donc d'annoter les noms composés, les adverbes composés donnant lieu à des sites de liaison catégorique (cf. Partie 1.Chapitre 4.2.2).

2.2.1. Jeu d'étiquettes morphosyntaxiques

Le jeu d'étiquettes morphosyntaxiques[149] que nous avons établi se base sur une version de celui utilisé par *Treetagger* (Schmid 1994) que nous avons enrichi de plusieurs catégories et fonctions. Le tableau ci-dessous (cf. Tableau 3) présente un bilan des jeux d'étiquettes proposés par d'autres étiqueteurs, ce qui nous permettra, dans le prochain paragraphe, de comparer, au-delà des étiquettes spécifiques, les approches adoptées par les différents annotateurs.

	Tree tagger (Schmid 1994)	**Stanford POS tagger (Toutanova et Manning, 2000)**	**French Treebank**	**MElt (Denis et Sagot, 2012)**	**Cordial adapté (Eshkol et al. 2010)**	**Tree tagger TCOF-POS (Benzitoun et al. 2012)**	**DisMo (Christodoulides et al. 2014, 2015)**
Abréviation	ABR	_	_	_	_	X:sig*	X:acr*
Adjectif	ADJ	A	A	ADJ	ADJ	ADJ	ADJ
Adjectif interrogatif	ADJ	A	_	ADJWH	ADJ	PRT:INT**	DET:int
Adverbe	ADV	ADV	ADV	ADV	ADV	ADV	ADV
Adverbe, acronyme	ADV	ADV	ADV	ADV	ADV	ADV	ADV:acr
Adverbe interrogatif	ADV	ADV	ADV	ADVWH	PIINT	ADV	ADV:int
Adverbe de négation	ADV	ADV	ADV	ADV	ADVNEG	ADV	ADV:neg
Adverbe d'intensité	ADV	ADV	ADV	ADV	ADV	ADV	ADV:deg
Conjonction de coordination	KON	C	CC	CC	CONJCOO	KON	CON:coo
Conjonction de subordination	KON	C	CS	CS	CONSUB	KON	CON:sub
Déterminant défini	DET:ART	D	D	DET	DET/DEF	DET:def	DET:def
Déterminant démonstratif	?	D	D	DET	DET/DEM	DET:dem	DET:dem
Déterminant indéfini	DET:ART	D	D	DET	DET/IND	DET:ind	DET:ind
Déterminant intérrogatif	?	D	D	DETWH	DET/INT	DET:int	DET:int
Déterminant exclamatif	?	D	D	D	DET/IND	DET:ind	DET:exc
Déterminant partitif	DET	D	D	D	DET/IND	DET:par	DET:par
Déterminant possessif	DET:POS	D	D	D	DET/POSS	DET:pos	DET:pos
Mot étranger	_	ET	ET	ET	ET	ETR	FRG
Marqueur discursif	INT	I	I	I	INT	INT	ITJ
Interjection	INT	I	I	I	INT	INT	ITJ:itj

[149] Nous remercions Julie Peuvergne pour son aide dans le travail d'adaptation du jeu d'étiquettes morphosyntaxiques pour l'annotation du corpus PFC.

Introductif	_	_	_	_	_	_	INTROD
Onomatopée	INT	I	I	I	UEUPH	INT	ITJ:ono
Nom acronyme	NOM	N	NC	NC	NC	NOM:sig	NOM:acr
Nom commun	NOM	N	NC	NC	NC	NOM	NOM: com
Nom propre	NAM	N	NP	NPP	NP	NAM	NOM:pro
Nom propre, anonymisé	NAM	N	NP	NPP	NP	NOM:sig	NOM:pro: anm
Numéral cardinal, déterminant	NUM	D	D	DET	DET	NUM	NUM:crd: det
Numéral cardinal, adjectif	NUM	D	D	DET	DET	NUM	NUM:crd: adj
Numéral cardinal, pronom	NUM	D	D	DET	DET	NUM	NUM:crd: pro
Numéral cardinal, nom	NUM	D	D	DET	DET	NUM	NUM:crd: nom
Numéral ordinal, adjectif	NUM	D	D	DET	DET	NUM	NUM:ord: adj
Numéral ordinal, pronom	NUM	D	D	DET	DET	NUM	NUM:ord: pro
Numéral ordinal, nom	NUM	D	D	DET	DET	NUM	NUM:ord: nom
Préposition	PRP	P	P	P	PREP	PRP	PRP
Préposition + Déterminant (contracté)	PRP:det	P	P	P+D	PREP	PRP:det	PRP:det
Préposition + Prénom	PRP	P	P	P+PRO	PREP	PRP	PRP
Préfixe	_	_	PREF	PREF	_	_	PFX
Ponctuation	PUN	PUNC	PONCT	PONCT	PCT	SENT	PNC
Ponctuation citation	PUN:cit	PUNC	PONCT	PONCT	PCT	SENT	PNC
Pronom démonstratif	PRO:DEM	PRO	PRO	PRO	P/DEM	PRO:dem	PRO:dem
Pronom indéfini	PRO:IND	PRO	PRO	PRO	P/IND	PRO:ind	PRO:ind
Pronom interrogatif	PRO	PRO	PRO	PROWH	P/INT	PRO:int	PRO:int
Pronom personnel, sujet	PRO:PER	PRO	PRO	CLS	PPERSUJ	PRO:cls	PRO: per:sjt
Pronom, impersonnel, sujet	PRO:PER	PRO	PRO	CLS	PPERSUJ	PRO:clsi	PRO: per:sjt
Pronom personnel, objet direct	PRO:PER	CL	Cl	CLO	PPERCOMPL	PRO:clo	PRO: per:objd
Pronom personnel, objet indirect	PRO:PER	CL	Cl	CLO	PPERCOMPL	PRO:clo	PRO:per: obji
Pronom possessif	PRO:POS	PRO	PRO	PRO	PPOSS	PRO:pos	PRO:pos
Pronom relatif	PRO:REL	CL	Cl	PROREL	PREL	PRO:rel	PRO:rel
Pronom réflexif	PRO:PER	CL	Cl	CLR	PPERCOMPL	PRO:clo	PRO:ref
Pronom personnel tonique	PRO:PER	PRO	PRO	PRO	PPERTON	PRO:ton	PRO: per:ton
Symbole	SYM	_	_	_	_	_	PNC
Verbe conditionnel	VER:cond	V	V	V	VCON	VER:cond	VER:cond
Verbe conditionnel auxiliaire	VER:cond	V	V	V	VCON	AUX:cond	VER: cond:aux
Verbe futur	VER:futu	V	V	V	VF	VER:fut	VER:fut
Verbe futur auxiliaire	VER:futu	V	V	V	VF	AUX:fut	VER: fut:aux
Verbe gérondif	VER:ppre	V	V	VPR	V	V	VER:ger
Verbe impératif	VER:impe	V	V	VIMP	VPIMP	VER:impe	VER: impe
Verbe impératif auxiliaire					VPIMP	AUX:impe	VER: impe:aux
Verbe imparfait	VER:impf	V	V	V	VI	VER:impf	VER:impf
Verbe imparfait auxiliaire	VER:impf	V	V	V	VI	AUX:impf	VER: impf:aux
Verbe infinitif	VER:infi	V	V	VINF	VINF	VER:infi	VER:inf
Verbe infinitif auxiliaire	VER:infi	V	V	VINF	VINF	AUX:infi	VER: inf:aux
Verbe participe passé	VER:pper	V	V	VPP	VPPPAS	VER:pper	VER:ppas

Verbe participe passé auxiliaire					VPPPAS	AUX:pper	VER:ppas:aux
Verbe participe présent	VER:ppre	V	V	VPR	VPPRES	VER:ppre	VER:ppre
Verbe participe présent auxiliaire					VPPRES	AUX:ppre	VER:ppre:aux
Verbe présent	VER:pres	V	V	V	VPINP	VER:pres	VER:pres
Verbe présent auxiliaire	VER:pres	V	V	V	VPAUX	AUX:pre	VER:pres:aux
Verbe présentatif	_	_	_	_	PRES	_	VER:pres:entatif
Verbe passé simple	VER:simp	V	V	V	VPPPAS	VER:simp	VER:simp
Verbe passé simple auxiliare	VER:simp	V	V	V	VP	AUX:simp	VER:simp:aux
Verbe subjonctif imparfait	VER:subi	V	V	VS	VPSUB	VER:subi	VER:subi
Verbe subjonctif imparfait auxiliare	VER:subi	V	V	VS	VPSUB	AUX:subi	VER:subi:aux
Verbe subjonctif présent	VER:subp	V	V	VS	VPSUB	VER:subp	VER:subp
Verbe subjonctif présent auxiliaire	VER:subp	V	V	VS	VPSUB	AUX:subp	VER:subp:aux
Disfluences							
Amorce de nom commun	_	_	_	_	_	NAM:trc	Voir Tableau 4
Amorce de verbe	_	_	_	_	_	NOM:trc	
Amorce de nom propre	_	_	_	_	_	VER:trc	
Amorce d'adjectif	_	_	_	_	_	ADJ:trc	
Discours							
Marqueur de discours	INT	I	I	I	MD / MDEUH / MDINT	INT	MD

Tableau 3 : Comparaison des jeux d'étiquettes utilisés par les différents annotateurs morphosyntaxiques

Comme nous l'avons souligné dans le paragraphe précédent, contrairement aux travaux précédents, nous n'avons pas essayé d'adapter le jeu d'étiquette existant pour l'écrit aux spécificités de l'oral, mais nous avons plutôt adopté la démarche proposée par *DisMo*. Cet annotateur comprend un système d'étiquettes qui précise les fonctions de ces éléments à des niveaux linguistiques différents. Autrement dit, au lieu de créer des étiquettes *ad hoc* tels que NAM:trc ; NOM:trc ; VER:trc ; ADJ:trc (Benzitoun et al. 2012) qui associent les phénomènes de disfluences (comme les amorces) aux classes morphosyntaxiques traditionnelles (en nous forçant de les traiter comme des formes réduites de ces dernières), il propose un jeu d'étiquettes qui s'articule sur plusieurs niveaux : (a) une double couche d'annotation morphosyntaxique (*pos :min / pos :mwu*) ; (b) une couche d'annotation des marqueurs discursifs (*discourse*) (c) une dernière couche consacrée à l'annotation des disfluences (*disfluences*) (cf. Paragraphe 2.2).

2.3. Conclusion

Les avantages de la campagne de ré-annotation et de correction manuelle que nous avons réalisée sont multiples. Premièrement, celle-ci nous a permis, comme nous l'avons déjà évoqué, d'améliorer la précision d'étiquetage et donc d'identification des contextes morphosyntaxiques de liaison (à savoir le type).

De plus, contrairement à l'annotation fournie par les autres étiqueteurs, *DisMo* permet de garder séparées les informations d'ordre catégoriel et fonctionnel. Par exemple, sur l'étiquette utilisée pour l'annotation du pronom personnel objet « les » (« PRO:per: objd ») dans l'exemple ci-dessous (cf. Exemple 2), l'annotation de la fonction d'objet direct (« objd ») est séparée de celle de la catégorie morphosyntaxique (à savoir « PRO:per »).

> « […] je connaissais déjà des gens […] je *les* (PRO:per:objd) rencontrais pendant les vacances donc dans l'Aveyron ». (PFC Aveyronnais à Paris; 75xmb1gg)

Exemple 2 : Annotation du pronom personnel objet « les » réalisée par le biais de l'annotateur *DisMo*

De plus, ce système d'annotation multi-niveaux nous permet d'annoter des marqueurs de discours à la fois dans la couche d'annotation morphosyntaxique (*pos:min ; pos:mwu*) selon leur catégorie traditionnelle (par exemple « VER:pres » pour « tu vois » cf. Exemple 3) et en tant que marqueur de discours dans la couche d'annotation *discourse* (MD). Cette approche nous permet de préciser la fonction pragmatique des marqueurs de discours dans une couche séparée et ainsi préserver les éventuelles ambiguïtés (Dostie 2004).

> « Moi je suis arrivé en septième enfin en CM2 […] on a été genre une dizaine ils en ont pris genre six *tu vois* (pos:mwu VER: pres; discouse: MD) un truc comme ça ». (PFC Paris centre ville; 75ccb2lg)

Exemple 3 : Annotation du marqueur discursif « tu vois » réalisée par le biais de l'annotateur *DisMo*

Quant à l'annotation des disfluences, le système proposée par *DisMo* s'avère par ailleurs plus économique, puisqu'au lieu de créer des étiquettes doubles (ex. « NAM » pour le nom propre entier ; « NAM :trc » pour son amorce correspondant), il propose une classification des disfluences basée sur leurs attributs et leurs fonctions et non sur la base de la classe morphosyntaxique de référence.

Le schéma d'annotation des disfluences proposé par *DisMo* est présenté dans le tableau ci-dessous (cf. Tableau 4). Les étiquettes sont ajoutées au niveau du lexème minimal (token

minimum de l'annotation, *tok:min*). Il est ainsi possible de préserver la catégorie morphosyntaxique originale des tokens disfluents.

Niveau 1: Les **disfluences simples** qui affectent un seul token		
FIL	Pauses remplies	c' est pour ça que j' hésite euh un peu à en parler FIL
LEN	Allongement lié à une hésitation	au cercle d'oenologie de= Bruxelles LEN
FST	Amorce lexicale	comme infirmière so/ sociale FST
WDP	Pause intra-mot	il m' a dit ça su+ _ +ffit WDP
Niveau 2: Les **répétitions** d'un ou plusieurs tokens à l'identique		
REP	Répétition	les disques et et lancer les jingles REP* REP_
		il a il a il a dit que REP:1 REP:2 REP:1 REP*:2 REP REP
		c' est pas c' est pas un système génial REP:1 REP:2 REP*:3 REP_ REP_ REP_
Niveau 3: Les disfluences **structurées**, dites d'édition		
DEL	Suppression	c' est vraiment un en tout cas la parole DEL DEL DEL DEL*
SUB	Substitution	cette personne était enfin c' est un ami de SUB* SUB:edt SUB_ SUB_
INS	Insertion	c' est vrai que Béthune euh vivre à Béthune a été INS* INS+FIL INS_ INS_ INS_
Niveau 4: Disfluences complexes qui sont une combinaison des disfluences simples et structurées (mais qui ne peuvent pas être décrites en combinant les niveaux 1 à 3)		
COM	Complexe	les ac/ les actions enfin les activités enfin professionnelles COM COM COM COM COM COM COM COM COM

Tableau 4 : Schéma dannotation des disfluences proposé par *DisMo*

L'apport de l'annotation des disfluences et des MD est d'une importance cruciale pour l'analyse des contextes de liaison. Les exemples ci-dessous (cf. Exemple 4 ; Exemple 6) permettent d'illustrer ce point.

« […] les ac/ les actions enfin les activités enfin professionnelles ».

Exemple 4 : Annotation du faux départ « les ac/ » réalisée par le biais de l'annotateur *DisMo*

« *Bon* (ADJ) à l'époque on travaillait que sur Golden, maintenant on a sorti de nouvelles variétés (bruit) comme pour la vigne d'ailleurs ». (PFC, Douzens 11aal1lg)

Exemple 5 : Annotation du marqueur discursif « bon » réalisée par le biais de l'annotateur *Treetagger*

« *Bon* (ITJ) à l'époque on travaillait que sur Golden, maintenant on a sorti de nouvelles variétés (bruit) comme pour la vigne d'ailleurs ». (PFC, Douzens 11aal1lg)

Exemple 6 : Annotation du marqueur discursif « bon » réalisée par le biais de l'annotateur *DisMo*

L'Exemple 4 montre que l'annotation des disfluences est nécessaire afin de repérer le contexte de liaison contenant des amorces lexicales ou faux départs (*false start,* étiquette *FST*).

Il en va de même pour l'Exemple 6 où l'annotation de l'unité « bon » en tant que marqueur discursif[150] permet d'exclure ce contexte de l'ensemble des sites de liaisons potentielles. En revanche l'annotation précédente fournie par *Treetagger* incluait cette séquence parmi les contextes de liaison variable proposant d'étiqueter « bon » comme adjectif (« ADJ »).

Finalement, la double couche d'annotation morphosyntaxique proposée par *DisMo* permet de réaliser une annotation par lexème et par unités poly-lexicales et donc d'annoter par exemple, des noms composés (ex. « petits-fils »), des adverbes composés (ex. « tout de suite »), mais également des unités plus complexes à traiter comme les introductifs « il y a », « c'est » et « ce sont ». Les composants de ces unités (notamment « il », « y » et « a ») sont traités séparément au niveau *pos-min* et sont ensuite annotés (à l'aide de l'étiquette « INTROD » à savoir « introducteur ») comme des unités poly-lexicales au niveau d'annotation morphosyntaxique suivant (*pos-mwu*).

« Je n'ai pas pu obtenir de poste à Lyon *tout de suite* (*pos:mwu* ADV) donc j'ai été exilé ». (PFC Lyon; 69aag1gg)

Exemple 7 : Annotation de l'adverbe composé « tout de suite » réalisée par le biais de l'annotateur *DisMo*

Comme nous l'avons déjà évoqué l'identification des unités poly-lexicales est une étape cruciale de l'annotation des types de liaison puisqu'elle permet d'isoler ces contextes au sein desquels la liaison est réalisée de façon catégorique. En effet un étiquetage incorrect de ces formes pourrait donner lieu à l'identification de contexte variable de liaison. C'est le cas de l'unité poly-lexicale « petit à petit » présentée dans l'Exemple 8 et annotée par *Treetagger* en tant que préposition suivie d'un déterminant (PRP-DET) (cf. Exemple 8). Cette annotation incorrecte néglige la nature poly-lexicale de cette unité en entraînant l'identification d'un type de liaison variable.

« *Petit à petit* (ADJ-PRP), on a essayé d'acheter une maison ». (PFC, Bangui rcamk1gg)

Exemple 8 : Annotation de l'adverbe composé « petit à petit » réalisée par le biais de l'annotateur *Treetagger*

« *Petit à petit* (ADV), on a essayé d'acheter une maison ». (PFC, Bangui rcamk1gg)

Exemple 9 : Annotation de l'adverbe composé « petit à petit » réalisée par le biais de l'annotateur *DisMo*

[150] Annoté « ITJ » dans la couche d'annotation *pos-min* alors que sa fonction pragmatique est précisée dans la couche *discourse* à l'aide de l'étiquette « MD ».

En revanche, l'annotation réalisée par *DisMo* permet d'annoter cette forme en tant qu'adverbe composé dans la couche d'annotation *pos:mwu* (cf. Exemple 9) et ainsi de la reconnaître comme un site de liaison catégorique (tout en gardant une trace de l'annotation des unités isolées au niveau morphosyntaxique précédent *pos:min*).

Chapitre 2. Analyse de la distribution de la fréquence de la liaison variable dans la parole semi-spontanée

Le travail de ré-étiquetage morphosyntaxique mené sur le corpus PFC nous a permis de disposer d'une large base de données contenant de nombreux contextes de liaisons potentielles codées et annotées en morphosyntaxe et lemme. À partir de cette ressource nous avons développé une analyse fréquentielle des réalisations des liaisons variables issues de la parole semi-spontanée (entretiens guidés et conversations libres) du corpus PFC (cf. Partie 2.Chapitre 1.1.3.1). Des études précédentes avaient déjà conduit des analyses de la distribution de la fréquence des liaisons catégoriques et variables sur un corpus de lectures (Boula de Mareüil, Adda-Decker et Gendner 2003) et sur un corpus mixte[151] (Durand et al. 2011 ; Laks, Calderone et Celata 2014). À la suite de ces travaux, à partir des données de parole semi-spontanée du corpus PFC (entretiens guidés et conversations libres), notre contribution se propose de mener une analyse fréquentielle des contextes morphosyntaxiques (types) de liaison variable dans un échantillon issu uniquement de la parole semi-spontanée et de préciser la notion de variabilité de la liaison au sein des classifications traditionnelles (Delattre 1966 ; Encrevé 1988).

1. Études précédentes

Des analyses quantitatives (Boula de Mareüil, Adda-Decker et Gendner 2003 ; Durand et al. 2011 ; Laks, Calderone et Celata 2014) sur de grands corpus de français oral ont été menées afin d'étudier la distribution de la fréquence des réalisations catégoriques et variables de la liaison. L'étude de Boula de Mareüil et al. (2003) a été conduite sur le corpus BREF[152] contenant 91.000 sites de liaisons potentielles et 40.000 occurrences de liaisons réalisées. Ce corpus a été annoté automatiquement à l'aide de l'étiqueteur *Cordial Tagger*[153] et comprend 100 heures totales de lectures oralisées des extraits du journal « Le Monde » produites par 120 locuteurs (65 femmes et 55 hommes). Parmi les sujets de cette étude, 80 locuteurs ont lu chacun environ 10.000 mots (soit environ 650 phrases), et 40 autres locuteurs en ont lu environ la moitié.

[151] Contenant à la fois les données de lecture et de parole semi-spontanée (entretiens guidés et conversations libres) du corpus PFC annoté par le biais de *Treetagger* (cf. Partie 2.Chapitre 1.2.1).

[152] Recueilli par le laboratoire LIMSI : http://catalog.elra.info/product_info.php?products_id=537&language=fr

[153] http://www.cordial.fr/Cordial_Analyseur/Presentation_Cordial_Analyseur.htm.

En revanche, l'étude menée par Laks, Calderone et Celata (2014) a été réalisée sur l'ensemble du corpus PFC comprenant la parole semi-spontanée et les lectures oralisées (cf. Partie 2.Chapitre 1.1.3.1) annoté dans sa première version à l'aide de *TreeTagger* (Schmid 1994).

Les résultats de ces travaux ont montré que la répartition statistique des fréquences d'apparition des réalisations de la liaison dans les corpus observés suit une distribution statistique de la fréquence décrite initialement par Zipf (1949) et ensuite définie par Mandelbrot (1957) (sous le nom de distribution de Mandelbrot-Zipf). Cette distribution représente un modèle de la loi de puissance[154] décrivant une relation mathématique entre deux quantités (l'une correspondant à la fréquence d'un évènement et l'autre indiquant la taille d'un évènement) où la décroissance des fréquences d'un évènement est moins importante lorsque la taille de l'évènement augmente.

1.1. Loi de Zipf

Les premières analyses conduites par Zipf (1949) ont été réalisées sur la base des listes de fréquence, à savoir des listes des types (mots) et tokens (occurrence), des textes écrits en langue anglaise[155]. Le tableau ci-dessous (cf. Tableau 5) montre un exemple construit de ce type de liste de fréquence (f).

Type	**f**	**Type**	**f**
demain	2	mon	1
pour	3	avec	1
aucun	1	sans	2
Elle	1	chien	1
nous	6	entre	1
chat	3	eau	1
homme	1		

Tableau 5 : Exemple construit de liste de fréquence

[154] Tout comme la loi de Pareto décrite en 1897 par l'économiste Vilfredo Pareto (1896). En analysant la répartition des revenus personnels des individus dans les principaux pays industrialisés, Pareto observe qu'elle suit une loi de puissance. Selon cette distribution qui sera généralisée par Joseph Juran en 1954, 80% des effets sont le produit de 20% des causes. Contrairement à la loi de Zipf (1949) qui concerne directement les résultats des fréquences absolues des mots, la loi de Pareto intéresse les probabilités.

[155] Zipf (1949) travailla initialement sur la langue anglaise écrite avant de se pencher sur d'autres langues. Parmi les ouvrages étudiés, le premier a été « Ulysses » de James Joyce.

Afin de mener une étude des fréquences Zipf (1949) adopte une classification décroissante de la fréquence des mots organisée en rangs. Celle-ci, également connue sous le nom de « profil rang / fréquence », correspond à une répartition des fréquences en classe (rang) (r) en fonction de leur fréquence (f) d'apparition. Parmi les mots appartenant au rang 1 on retrouve donc les mots les plus fréquents comme « nous » ayant 6 occurrences, au rang 2 les mots « pour » et « chat » et ainsi de suite (cf. Tableau 6).

r	f	**r**	f
1	6	**8**	1
2	3	**9**	1
3	3	**10**	2
4	2	**11**	1
5	2	**12**	1
6	1	**13**	1
7	1		

Tableau 6 : Exemple construit de classification rang-fréquence

Selon la forme la plus simple de la loi de Zipf (1949) (également appelée « loi rang-fréquence ») le produit rang-fréquence est constant et la fréquence des mots est inversement proportionnelle à leur rang. Cette répartition est d'autant plus évidente dans le spectre de fréquence (cf. Tableau 7) qui représente le nombre de types (V(f)) ayant la même fréquence (f) d'occurrence (ex. « pour » et « chat »).

f	**V(f)**
1	8
2	2
3	2
6	1

Tableau 7 : Exemple construit de spectre de fréquence

Certaines études (Yule 1924) ont montré que bien que cette distribution affecte tout particulièrement les mots relevant d'une classe fermée (ex. articles, pronoms, conjonctions et prépositions), elle est également attestée à l'intérieur d'une classe morphosyntaxique ouverte comme par exemple celle des noms.

Cette distribution a été observée dans plusieurs langues (à l'écrit comme à l'oral) telles que l'anglais (Miller et Newman 1958), le chinois (Lua 1994), l'italien (Baroni et Evert 2005 ;

Baroni et Evert 2014), l'allemand (Billmeier 1969 cité par Baroni et Evert 2005) et plusieurs interprétations ont été proposées pour l'expliquer[156].

Zipf (1949) interprétera la régularité de cette distribution statistique comme la conséquence d'une propriété universelle de l'esprit humain qui privilégie « le moindre effort ». Sur la base de ce principe, le locuteur aurait tendance à réduire son vocabulaire à un nombre restreint de mots fréquents couvrant un ensemble de significations. D'après cette conception, le langage naturel reposerait sur un équilibre entre économie linguistique et efficacité de la communication. Ces forces contradictoires donneraient lieu à la distribution décrite par Zipf (1949) qui représente un écart important entre d'une part un groupe limité de types fréquents et d'autre part un nombre important de types ayant un taux de fréquence très faible. Ce même principe du « moindre effort » serait également responsable d'après Zipf (1949) de la fréquence d'apparition des mots puisque les mots les plus fréquents seraient également les plus courts. Autrement dit, l'indice de fréquence des mots serait directement corrélé au nombre de syllabes des mots. Toutefois l'explication apportée par Zipf (1949) ne parvient pas à justifier la présence de la même distribution dans d'autres disciplines telles que l'histoire, la démographie et l'urbanisme. D'autres approches basées sur des modèles mathématiques probabilistes de la croissance du vocabulaire ont été avancées afin de rendre compte de la distribution de la fréquence du lexique. Selon ces hypothèses, la probabilité d'apparition de nouveaux mots est liée à une probabilité constante. La probabilité qu'un mot apparaisse *x* fois est proportionnelle au nombre total d'occurrences des mots apparus *x* fois et appartenant donc au même rang.

En dépit de toute explication biologique ou probabiliste, la répartition des fréquences observée par Zipf (1949) semble effectivement attestée dans de grands corpus issus de langues typologiquement différentes. À titre d'illustration, la figure ci-dessous (cf. Figure 17) montre une comparaison des courbes rang-fréquence représentées dans un repère bilogarithmique (Baroni 2008). Celles-ci sont issues du corpus *Brown* d'anglais américain, du corpus d'anglais *British National Corpus* (BNC), de l'ouvrage *The war of the words*, du corpus d'extraits du journal italien *La Repubblica* et d'un corpus de pages web en japonais.

[156] La loi de Zipf a été également décrite dans d'autres disciplines telles que l'histoire (entre autres Guerreau 1989 par Baroni et Evert 2005), la démographie et l'urbanisme (entre autres Schaffar et al. 2012 cité par par Baroni et Evert 2005).

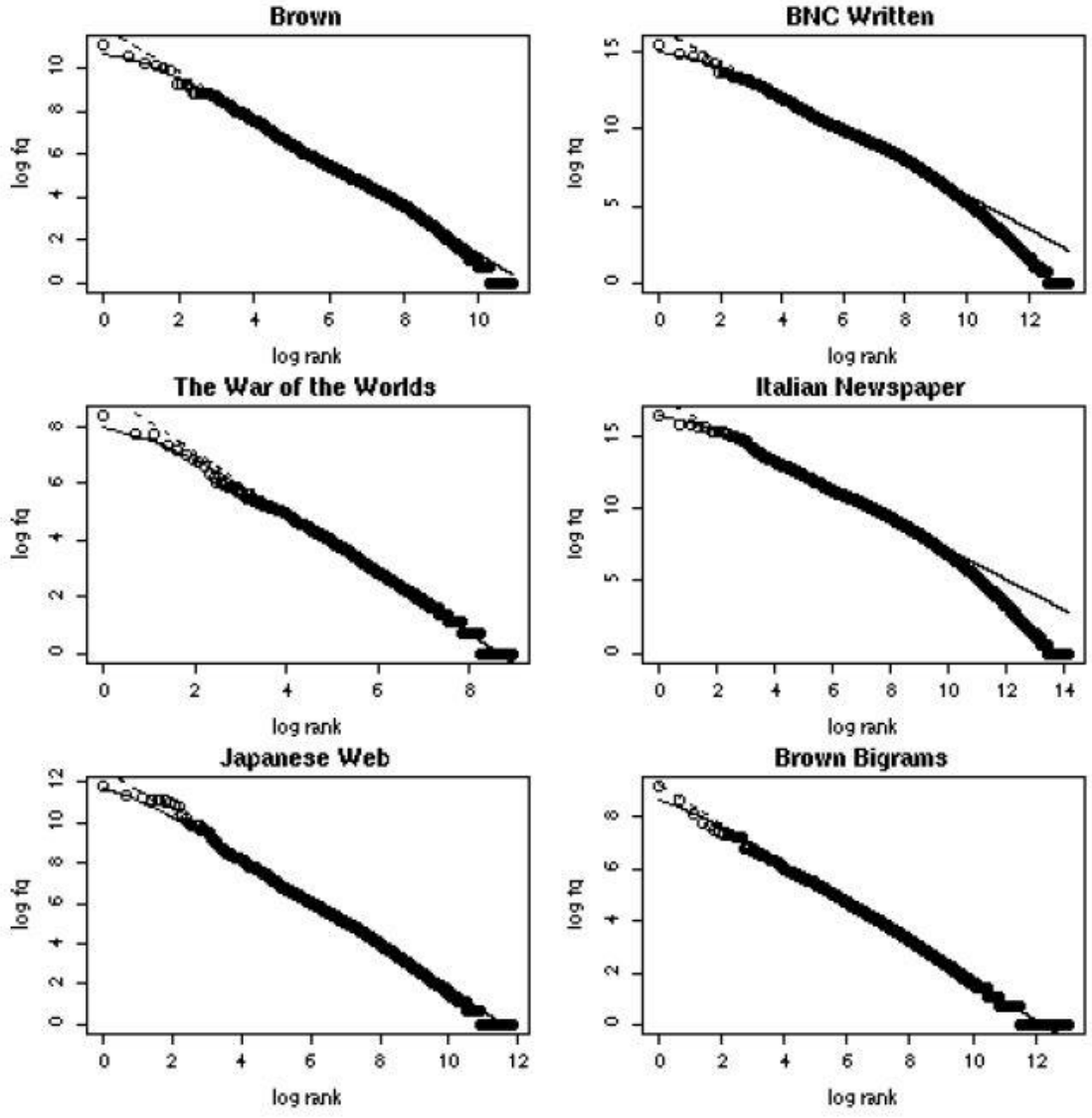


Figure 17 : Comparaison des courbes rang-fréquence représentées dans un repère bilogarithmique (tirées de Baroni 2008, p. 12‑13)

Cette analyse comparative des courbes rang-fréquence montre clairement une distribution en forme de cloche (*bell-bottom*) représentée d'une part, à l'extrémité droite, par les rangs comprenant un nombre important de types de basse fréquence et d'autre part, à l'extrémité gauche, par les rares mots (repartis parmi différents rangs) ayant une fréquence élevée.

1.2. Distribution de Zipf de la fréquence des réalisations totales de la liaison

L'observation de la présence d'une distribution de Zipf parmi les données de réalisation de la liaison indique qu'un groupe restreint de contextes morphosyntaxiques de liaison donnent lieu à la plupart des réalisations totales (catégoriques et variables) de la liaison.

En effet si nous comparons les données des deux études (Boula de Mareüil, Adda-Decker et Gendner 2003 ; Durand et al. 2011 ; Laks, Calderone et Celata 2014) que nous avons cités dans les paragraphes précédents (cf. Paragraphe 1), nous pouvons observer que les 50 types les plus fréquents du corpus BREF dans l'étude de Boula de Mareüil et al. (2003) rendent compte de 98,6% des réalisations totales. Il en va de même pour les résultats de l'étude menée par Laks, Calderone et Celata (2014) où les 50 types les plus fréquents sur 3105 totaux induisent 50% du total des réalisations de liaison (catégorique et variable)[157].

> « The results of the analysis, indicating the existence of a power-laws distribution in the production of liaison, suggested that there is a complex function regulating the number of liaison environments and their token frequency in a corpus ». (Laks, Calderone et Celata 2014, p. 49)

Cette étude a également analysé l'influence des facteurs phonologiques (à savoir la nature de la CL) et sociolinguistiques (ex. âge, sexe, niveau socio-culturel des locuteurs) sur la distribution de fréquence d'apparition de la liaison. Les variables phonologiques ne semblent apporter aucune information supplémentaire à l'analyse de la distribution des réalisations. Il en va de même pour les facteurs externes de nature sociolinguistique bien qu'il semblerait que ceux-ci influencent de façon marginale une faible partie des données correspondant aux types les moins fréquents. Contrairement à ces derniers, les types les plus fréquents, capables de rendre compte de la presque totalité des réalisations, ne seraient pas sujets à l'influence des facteurs externes.

> « External factors, such as the age and the educational level of the speakers, neither provide substantially different pictures of the general distribution function […]. We therefore take this point as one of the fundamental lesson of our corpus study, allegedly bearing important consequences for current sociophonological practice: there are phenomena in the phonology of languages, for which a "core" of frequent lexical uses may be substantially untouched by sociolinguistic variation, while a "periphery" of infrequent uses appears to show significant aspects of style- or speaker-dependent variation ». (Laks, Calderone et Celata 2014, p. 49)

[157] Le nombre est différent pour des raisons d'étiquetage en POS.

2. Présente étude

À la suite des travaux précédents (Boula de Mareüil, Adda-Decker et Gendner 2003 ; Laks, Calderone et Celata 2014) l'objectif de la présente étude est de mener une analyse de la distribution des réalisations de la liaison variable à partir des données de parole semi-spontanée (lectures exclues) extraites du corpus PFC (Durand, Laks et Lyche 2009).

Ce propos se justifie par la constatation du besoin en L1 ainsi qu'en acquisition L2 de disposer d'une description plus précise de la variabilité de la liaison au sein de chaque type de liaison variable.

En effet, comme nous l'avons déjà souligné, de nombreuses recherches portant sur le phénomène de la liaison en français ont pu mettre en évidence l'influence de divers paramètres linguistiques sur la réalisation des liaisons (cf. Partie 1.Chapitre 2). De plus, non seulement la liaison relève de plusieurs niveaux d'analyse mais elle ne constitue pas non plus un phénomène unique et homogène analysable comme un processus unique. Elle est au contraire profondément hétérogène et variable. Plusieurs études (Mallet 2008 ; Durand et al. 2011) menés sur le corpus de français oral ont souligné la forte diversité des comportements qui caractérise les réalisations même à l'intérieur de la même catégorie morphosyntaxique et du même contexte segmental. Multi-dimensionnalité et hétérogénéité rendent donc difficile un traitement classificatoire de liaison sous formes de règles formelles ainsi que sa prédiction (cf. Partie 1.Chapitre 4.4).

Ces observations suggèrent que pour traiter de la liaison, il est nécessaire d'analyser très précisément chaque contexte et chaque occurrence (Morin et Kaye 1982 ; Morin 2005). Morin et Kaye (1982) avaient déjà souligné que le contexte syntaxique de réalisation de la liaison doit être défini catégorie par catégorie mais aussi mot à mot, et la dimension lexicale doit donc être intégrée dans l'étude du contexte syntaxique de réalisation de la liaison (Bybee 2005). Nous essaierons donc de développer une analyse capable de rendre compte de l'hétérogénéité interne aux différents contextes décrits par les classifications traditionnelles et plus précisément de la plus grande productivité de certaines séquences issues du même contexte morphosyntaxique de liaison (Mallet 2008 ; Côté 2013 ; Barreca 2014).

2.1. Corpus et méthodologies

La ressource orale sur laquelle nous nous appuyons est le corpus PFC (Durand, Laks et Lyche 2009). Notre échantillon correspond en particulier à 4296 occurrences de liaisons variables réalisées qui ont été extraites des transcriptions des seuls entretiens guidés et des conversations libres de 394 locuteurs natifs issus de 38 points d'enquête. Dans le cadre de cette étude nous avons décidé d'écarter les données issues des lectures oralisées pour lesquelles la fréquence de réalisation pourrait avoir été influencée par la nature même de la tâche.

Comme nous l'avons évoqué dans le chapitre précédent (cf. Partie 2.Chapitre 1.2.2), l'ensemble de ces données (ainsi que la totalité du corpus PFC précédemment annoté à l'aide de *TreeTragger*) a été soumis à un important travail de ré-annotation en POS à l'aide de l'étiqueteur morphosyntaxique *DisMo* (Christodoulides, Avanzi et Goldman 2014 ; Christodoulides, Barreca et Avanzi 2015). Nous rappelons que ce travail de ré-annotation automatique (suivi d'un post-traitement de correction manuelle de 27000 tokens) nous a permis d'améliorer la précision d'étiquetage des types de liaison. À la suite de ce traitement, nous avons mené une analyse fréquentielle de la liaison à l'aide du package *Zipf-R* développé par Baroni et Evert (2007 ; 2014) et exploitable dans le cadre du logiciel statistique R.

2.2. Résultats

2.2.1. Distribution de la fréquence de la liaison variable

Les résultats de notre analyse montrent que la distribution des réalisations de la seule liaison variable en parole semi-spontanée suit également une distribution Zipf (1949), déjà observée par les études précédentes sur l'ensemble des types de liaison (catégoriques et variables).

La figure ci-dessous (cf. Figure 18) représente la distribution des fréquences des liaisons variables réalisées. Sont représentés en abscisse les types (*rank* représentant 114 types différents de liaison) et en ordonnée les tokens (à savoir les occurrences de liaison réalisée représentant 4296 tokens). Les 9 premiers types, c'est-à-dire les 9 contextes de liaison variable les plus fréquents sur 114 types totaux, rendent compte de 76% des réalisations totales (à savoir 3280 / 4296 tokens).

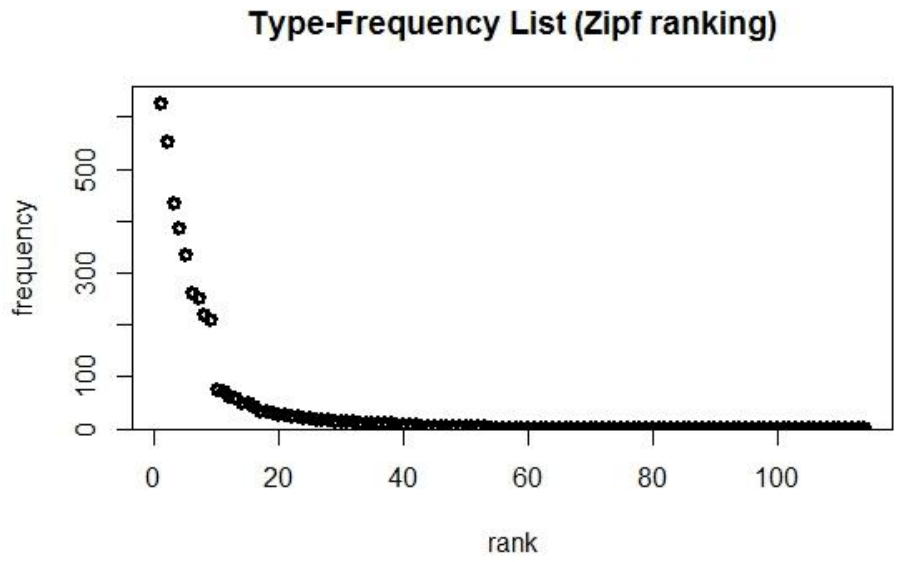


Figure 18 : Distribution de la fréquence des réalisations de la liaison variable dans la parole semi-spontanée du corpus PFC

Autrement dit, un nombre restreint de types permettent de rendre compte de la presque totalité des réalisations de la liaison variable. Celles-ci correspondent aux constructions-types constituées d'autant de mots issus des classes fermées (prépositions, conjonctions et adverbes) que des classes ouvertes (verbes et adjectifs)[158]. Le tableau ci-dessous (cf. Tableau 8) résume ces types et le nombre de réalisations de la liaison correspondante (tokens)[159].

Type	Token	Type	Token
CON:sub-PRO:per:sjt	627	PRP-NOM:pro	263
PRP-DET:ind	553	ADV:deg-ADJ	251
VER:pres-DET:ind	435	VER:pres-PRP	220
PRP-NOM:com	386	ADJ-NOM:com	209
VER:pres:aux-VER:ppas	336		

Tableau 8 : Types de liaison variables réalisées les plus fréquents (n. tokens > 100)

L'observation d'une telle répartition des fréquences indique qu'il existe de nombreux types possédant très peu de réalisations (n. tokens < 100) de liaison ; par contre, le nombre de types qui possèdent un taux élevé de réalisation de liaison est limité.

[158] Le jeu d'étiquette utilisée pour l'annotation des types de liaison est celui proposé par l'annotateur morphosynatxique DisMo (cf. Partie 2.Chapitre 1.2.2.1).

[159] Les exemples issus de l'ensemble des types étudiés sont présentés dans la première annexe (cf. Annexe 1) où, pour raison d'espace, les différents types sont regroupés selon la classe morphosyntaxique principale d'appartenance à savoir selon les types adoptés par les classifications traditionnelles de la liaison (ex. « CON:sub - PRO:per:sjt » sous la classe « CON - PRO »).

Afin de vérifier la validité de ces résultats, nous avons mené à l'aide du package *Zipf-R* (Baroni et Evert 2007 ; Baroni et Evert 2014) une analyse d'ajustement (*goodness of fit*) d'un modèle statistique paramétrique *Large Number of Rare Events model* (dorénavant modèle LNRE) (Baayen 2001) pour la distribution des fréquences d'une population (en l'occurrence, la fréquence de réalisation de la liaison)[160] aux données observées. Cette analyse statistique paramétrique basée sur un test multi-varié du *khi* carré (Baayen 2001) nous permet donc d'estimer la distance existant entre le spectre de fréquence observé et celui attendu sur la base du modèle LNRE. L'ajustement du modèle statistique aux données observées est d'autant plus valide que le *khi* carré (x^2) est petit et que la valeur *p* d'ajustement des données de liaison au modèle LNRE est importante. Autrement dit, la valeur *p* doit être plus grande que 0,05 ($p > 0,05$) pour que l'hypothèse soit vérifiée. L'estimation de l'ajustement (*goodness of fit*) du modèle testé aux données observées est détaillée dans le tableau ci-dessous (cf. Tableau 9).

Goodness of fit (multivariate chi-squared test)		
x^2	*df*	*p*
1,7	3	0,6

Tableau 9 : Résultats de l'analyse d'ajustement (*goodness-of-fit*) du modèle statistique paramétrique LNRE aux données observées

Le coefficient du *khi* carré est très petit : 1,7. En revanche, la valeur *p* est plus grande que 0,05 ($p = 0,6$). De ces résultats, nous pouvons conclure que la réalisation de la liaison variable observée suit une loi de puissance de type Mandelbrot-Zipf.

L'ajustement est également illustré dans la figure ci-dessous (cf. Figure 19) qui représente une comparaison des deux spectres de fréquences (cf. Paragraphe 1.1) issus du modèle LNRE (*fZM model* cf. note 160) et des données observées. D'une part on observe les colonnes de l'histogramme en noir correspondant à la fréquence de réalisations de la liaison variable observée (*observed*). D'autre part l'histogramme en rouge représente le spectre de fréquence des réalisations de la liaison variable estimé à partir du modèle LNRE (*fZM model*). Pour des raisons d'espace, ce graphique ne représente que les types ayant de 1 à 15 occurrences.

[160] Il s'agit plus précisément du modèle finite Zipf-Mandelbrot (LNRE.fZM) (cf. Evert 2004).

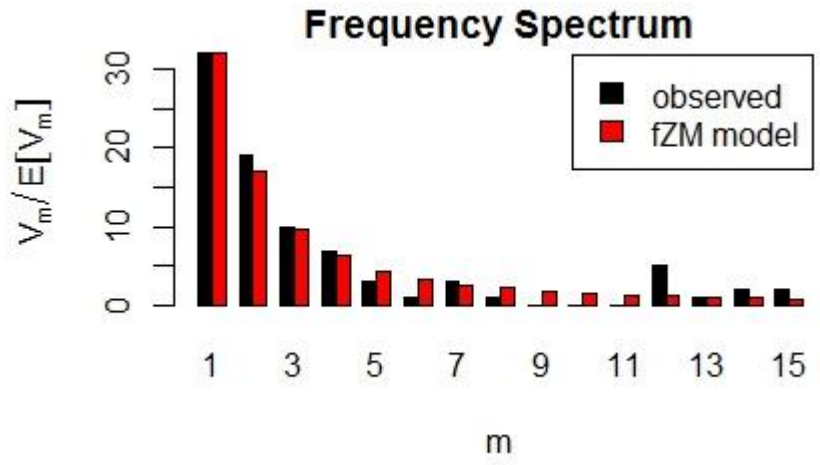


Figure 19 : Comparaison des deux spectres de fréquences issus du modèle LNRE (*fZM model*) et des données observées (*observed)*

2.2.2. Idiosyncrasies de la liaison variable

L'analyse fréquentielle de la liaison variable nous a permis d'identifier les 9 différents types de liaison les plus fréquents sur un total de 114 qui, comme nous l'avons vu, permettent de rendre compte de 76% des réalisations totales de la liaison variable. Les détails des résultats extraits de chaque type sont présentés dans la première annexe (cf. Annexe 1). Si nous examinons de plus près les résultats de ces types de liaison variable, nous observons une forte idiosyncrasie inter-catégorielle (cf. Annexe 2).

Ces différences sont évidentes si nous comparons les résultats des 9 premiers types de liaison issus de l'analyse fréquentielle (en gras dans le tableau ci-dessous) (cf. Tableau 10) avec ceux des types similaires appartenant au même contexte morphosyntaxique de liaison tel qu'il est décrit par les classifications traditionnelles de la liaison (ex. « CON-PRO ») (cf. Partie 1.Chapitre 4.3). Le tableau ci-dessous résume (cf. Tableau 10) les résultats pour chaque type le nombre de liaisons réalisées (LR), non-réalisées (LNR) ainsi que le taux de réalisation de la liaison.

	Type	LR	LNR	%LR
CON-PRO	**CON:sub-PRO:per:sjt**	**627**	**205**	**75%**
	CON:coo-PRO:per:sjt	1	857	0%
PRP-DET	**PRP-DET:ind**	**553**	**159**	**78%**
VER-DET	**VER:pres-DET:ind**	**435**	**1476**	**23%**
	VER:cond-DET:ind	0	37	0%
	VER:fut-DET:ind	0	13	0%
	VER:impe-DET:ind	0	7	0%
	VER:impf-DET:ind	34	976	3%
	VER:inf-DET:ind	2	283	1%
	VER:ppas-DET:ind	12	285	4%

	VER:ppre-DET:ind	1	7	13%
	VER:simp-DET:ind	0	4	0%
	VER:subp-DET:ind	1	26	4%
PRP-NOM	**PRP-NOM:com**	**386**	**10**	**97%**
	PRP-NOM:pro	**263**	**23**	**92%**
VER:aux-VER	**VER:pres:aux-VER:ppas**	**336**	**735**	**31%**
	VER:cond- VER:ppas	0	45	0%
	VER:fut- VER:ppas	0	2	0%
	VER:impe- VER:ppas	1	2	0%
	VER:impf- VER:ppas	17	331	5%
	VER:subp- VER:ppas	0	11	0%
ADV-ADJ	**ADV:deg-ADJ**	**251**	**46**	**85%**
	ADV:neg-ADJ	1	106	0%
VER-PRP	**VER:pres-PRP**	**1152**	**220**	**16%**
	VER:cond-PRP	0	15	0%
	VER:fut-PRP	0	7	0%
	VER:impe-PRP	2	23	8%
	VER:impf-PRP	21	922	2%
	VER:inf-PRP	1	552	0%
	VER:ppas-PRP	13	341	4%
	VER:ppre-PRP	1	9	10%
	VER:simp-PRP	0	2	0%
	VER:subp-PRP	0	6	0%
ADJ-NOM	**ADJ-NOM**	**46**	**209**	**82%**

Tableau 10 : Comparaison entre les types de liaison variable les plus fréquents (n. tokens > 100) (en gras) et les types issus des mêmes classes morphosyntaxiques

Les données issues de cette analyse ne font que confirmer les tendances déjà observées par les études précédentes menées sur des corpus linguistiques (cf. Partie 1.Chapitre 4.3.2). En effet, la liaison est plus fréquente au présent (« VER:pres ») (en particulier après les formes du verbe « être ») qu'après les autres temps verbaux. En revanche la liaison après les adverbes de négation (« ADV:neg ») est caractérisée par un taux de réalisation faible.

Cependant ces résultats ne nous disent rien quant aux fréquences de réalisation au sein de chaque type de liaison. En effet, comme nous l'avons vu dans la première partie de cette étude (cf. Partie 1.Chapitre 4.4) une description basée uniquement sur les contextes morphosyntaxiques de liaison ne permet pas de prédire la réalisation de la liaison.

Il nous est donc apparu nécessaire de mener une étude plus approfondie des occurrences de liaison afin de mettre en évidence les éventuelles idiosyncrasies intra-catégorielles pour les types de liaison analysés jusqu'ici. Les résultats de l'étude des occurrences de liaison pour les types de liaison les plus fréquents (cf. Tableau 8)[161] seront présentés dans les pages qui suivent.

[161] Sauf le type « VER:aux-VER:ppas » pour lequel l'analyse des M1 (à savoir les auxiliaires) ne s'avérait pas nécessaire.

Une telle analyse quantitative servira de base dans la dernière partie de ce travail (cf. Partie 5) pour le développement d'une classification plus fine de la liaison variable. Ces données seront également utilisées pour évaluer les productions des apprenants de français langue étrangère (cf. Partie 4.Chapitre 2.1.2.1.3).

2.2.2.1 Conjonctions suivies d'un pronom (CON-PRO)

Une forte idiosyncrasie est observée au sein des conjonctions monosyllabiques parmi lesquelles trois conjonctions sont les plus attestées dans notre échantillon : « quand », « mais » et « puis » (cf. Tableau 11). Le tableau présente, pour chaque conjonction des types « CON:sub-PRO:per:sjt » et « CON:coo-PRO:per:sjt » : (a) le nombre total d'occurrences de laisons réalisées (dorénavant LR) ; (b) le nombre total d'occurrences de liaisons non-réalisées (dorénavant LNR) ; (c) le taux de réalisation de la liaison (dorénavant %LR).

	Type	LR	LNR	%LR
CON-PRO	**CON:sub-PRO:per:sjt**	**627**	**205**	**75%**
	Quand	626	187	77%
	Soit	1	12	8%
	Sinon	0	6	0%
	CON:coo-PRO:per:sjt	**1**	**857**	**0%**
	Mais	1	427	0%
	Puis	0	430	0%

Tableau 11 : Taux de réalisation de la liaison par conjonction des types « CON:sub-PRO:per:sjt » et « CON:coo-PRO:per:sjt »

Si la liaison est presque toujours réalisée après la conjonction de subordination « quand », presque aucune liaison n'a été observée après les conjonctions de coordination « mais » et « puis ». Nous reviendrons sur les raisons de cet écart important entre productions dans les paragraphes qui suivent (cf. Paragraphe 2.3).

2.2.2.2 Prépositions suivies d'un déterminant (PRP-DET)

Si la liaison n'est jamais produite après les prépositions polysyllabiques elle semble néanmoins avoir un comportement relativement homogène après les prépositions monosyllabiques (« chez », « dans », « en », « sans » et « sous ») sauf après « vers » et « près » (cf. Tableau 12). Bien que la liaison soit plus fréquente après la préposition « dans » à cause de la fréquence plus importante de cette préposition (« dans - z - DET:ind »), elle n'est pas beaucoup moins productive après les autres.

	Type	LR	LNR	%LR
PRP (monosyllabique)-DET	**PRP-DET:ind**	**553**	**157**	**78%**
	Chez	22	8	73%
	Dans	506	44	92%
	En	15	7	68%
	Près	0	5	0%
	Sans	8	0	100%
	Sous	2	0	100%
	Vers	0	5	0%
PRP (polysyllabique)-DET	Après	0	23	0%
	Depuis	0	11	0%
	Devant	0	10	0%
	Pendant	0	44	0%

Tableau 12 : Taux de réalisation de la liaison par préposition du type « PRP-DET:ind »

2.2.2.3 Verbes suivis d'un déterminant (VER-DET)

Une plus grande hétérogénéité est observée dans le cas de la liaison entre verbes et déterminants où les seules formes « est » et « était » sur un total d'environ 500 formes verbales (tous temps et modes verbaux confondus) rendent compte d'environ 90% des réalisations. Cependant, il convient de souligner que le taux de non-réalisation reste très élevé (67% dans le cas de « est ») même après ces formes (cf. Tableau 13). Cela confirme l'idée que les facteurs syntaxiques et lexicaux ne suffisent pas à décrire la variabilité de la liaison. Nous reviendrons sur cet aspect dans les prochains paragraphes (cf. Paragraphe 2.3).

	Type	LR	LNR	%LR
VER-DET	**VER:pres-DET:ind**	**435**	**1476**	**23%**
	Est	420	861	33%
	Suis	4	63	6%
	Fait	3	81	4%
	Sont	3	16	16%
	Avons	2	8	20%
	As	1	58	2%
	Es	1	13	7%
	Sortent	1	0	100%
	Autres 142 formes verbales	0	376	0%
	VER:impf-DET:ind	**34**	**976**	**3%**
	Était	25	359	7%
	Avait	5	312	2%
	Avaient	2	18	10%
	Avais	1	77	1%
	Faisait	1	30	3%
	Autres 109 formes verbales	0	180	0%

	VER:ppas-DET:ind	**12**	**285**	**4%**
	Fait	11	197	6%
	Ouvert	1	3	33%
	Autres 27 formes verbales	0	85	0%

Tableau 13 : Taux de réalisation de la liaison par forme verbale des types VER:pres-DET:ind, VER:impf-DET:ind et VER:ppas-DET:ind

2.2.2.4 Prépositions suivies d'un nom commun et propre (PRP-NOM)

De manière générale, nous pouvons observer un comportement assez homogène au sein des prépositions monosyllabiques après lesquelles la liaison est réalisée à hauteur de 96% devant les noms communs, et 92% devant les noms propres. Des constructions-types telles que « en - n - NOM:com » émergent de l'ensemble des occurrences mais uniquement en raison d'une plus grande probabilité d'apparition (Côté 2013 ; cf. Paragraphe 2.3.1.2) de ces prépositions dans ce contexte de liaison. Au contraire, il n'a été enregistré aucune réalisation de liaison après les prépositions polysyllabiques (cf. Tableau 14).

	Type	**LR**	**LNR**	**%LR**
PRP-NOM	**PRP-NOM:com**	**386**	**10**	**97%**
	En	364	5	99%
	Sans	20	2	91%
	Sous	2	0	100%
	Après	0	3	0%
	PRP-NOM:pro	**263**	**23**	**92%**
	En	262	6	98%
	Sans	1	0	100%
	Chez	0	4	0%
	Dans	0	8	0%
	Après	0	2	0%
	Devant	0	1	0%
	Depuis	0	2	0%

Tableau 14 : Taux de réalisation de la liaison par préposition des types « PRP-NOM:com » et « PRP-NOM:pro »

2.2.2.5 Adverbes suivis d'un adjectif (ADV-ADJ)

Ce type de liaison est sans doute l'une des plus représentatives de la variabilité de la liaison associée à une même classe morphosyntaxique (Mallet 2008 ; Côté 2013 ; Barreca 2014) (cf. Partie 1.Chapitre 4.3.2.3.1).

	Type	**LR**	**LNR**	**%LR**
ADV (monosyllabique)-ADJ	**ADV:deg-ADJ**	**250**	**38**	**87%**
	Très	157	9	95%
	Plus	83	16	84%
	Moins	3	8	27%

	Trop	7	5	58%
	ADV:neg-ADJ	**1**	**105**	**1%**
	Pas	1	105	1%
ADV (polysyllabique)-ADJ	**ADV:deg-ADJ**	**1**	**8**	**11%**
	Tellement	0	4	0%
	Beucoup	1	4	20%
	ADV:neg-ADJ	**0**	**1**	**0%**
	Jamais	0	1	0%

Tableau 15 : Taux de réalisation de la liaison par préposition des types « ADV:deg-ADJ » et « ADV:neg-ADJ »

Les résultats de notre analyse confirment les tendances déjà évoquées par les études précédentes. La liaison est beaucoup plus productive après les adverbes monosyllabiques « très », « bien » et « plus »[162] (« très - z- NOM », « bien - n - NOM » et « plus - z- NOM ») qu'après les adverbes monosyllabiques « moins », « trop » et surtout après l'adverbe de négation « pas » (cf. Tableau 15). Comme nous le verrons dans les prochains paragraphes (cf. Paragraphe 3), des analyses des probabilités transitionnelles de ces adverbes ont été proposées (Côté 2013) afin de rendre compte de cet écart de réalisation.

2.2.2.6 Verbes suivis d'une préposition (VER-PRP)

Le même profil que celui observé pour le type VER-DET (cf. Tableau 13) émerge de l'analyse des liaisons entre verbes et prépositions où encore une fois la liaison tout en étant plus productive dans les constructions « est - t - PRP », « sont - t - PRP » et « était - t - PRP » montre un pourcentage élevé de non-réalisations.

	Type	**LR**	**LNR**	**%LR**
VER-PRP	**VER:pres-PRP**	**220**	**1152**	**16%**
	Est	182	395	32%
	Sont	23	74	24%
	Suis	4	86	4%
	Sommes	2	13	13%
	Autres 193 formes	9	584	2%
	VER:impf-PRP	**21**	**922**	**2%**
	Était	10	259	4%
	Étaient	3	26	10%
	Étais	3	139	2%
	Autres 192 formes verbales	5	498	1%

Tableau 16 : Taux de réalisation de la liaison par préposition des types « VER:pres-PRP » et « VER :impf-PRP »

[162] En tant qu'adverbe d'intesité (cf. « ADV:deg »).

2.2.2.7 Adjectifs suivis d'un nom (ADJ-NOM)

Ce type de liaison représente, comme nous l'avons vu dans la première partie de cette étude (cf. Partie 1.Chapitre 4.3.2.3.2), l'un des contextes (en même temps que le type ADV-ADJ) qui a fait l'objet de plusieurs révisions dans les descriptions de la liaison. Les études menées sur des corpus linguistiques ont en effet montré que la liaison dans ce contexte, bien que très fréquente et limitée à un groupe restreint d'adjectifs, ne pourrait pas être considérée comme catégorique. Dans notre étude nous n'avons observé que 30 lemmes au total issus de l'ensemble des adjectifs singuliers et pluriels.

	Type	LR	LNR	%LR
ADJ-NOM	**ADJ-NOM**	**209**	**46**	**82%**
	Bon	24	1	96%
	Autres	23	17	58%
	Petit	18	3	86%
	Petites	14	0	100%
	Petits	9	3	75%
	Grand	8	4	67%
	Autres 47 ADJ	18	113	14%

Tableau 17 : Taux de réalisation de la liaison par préposition du type « ADJ-NOM »

Parmi ceux-ci les seules formes « bon - n - NOM », « autres - z - NOM », « petit - t - NOM », « petites - z - NOM », « petits - z - NOM » et « grand - t - NOM » représentent 77% des réalisations totales (cf. Tableau 17). Ces premières données ont mis en évidence d'importantes différences entre les adjectifs singuliers et pluriels qui nous ont donc amené à développer une étude supplémentaire qui sera présentée dans les prochains paragraphes (cf. Paragraphe 2.2.3).

2.2.2.8 Bilan

Les données issues de l'analyse conduite à l'intérieur des types variables montrent qu'un nombre restreint d'items lexicaux permet de rendre compte de plus de la moitié des productions à l'intérieur du même type de liaison. Ces unités ne relèvent pas seulement des classes fermées mais également des catégories morphosyntaxiques telles que les adjectifs (cf. Tableau 17) et les verbes (cf. Tableau 13 ; Tableau 16).

Le tableau ci-dessous (cf. Tableau 18) résume les constructions observées pour les types les plus fréquents de liaison variable.

Type	Constructions	% sur LR tot.	Type	Constructions	% sur LR tot.
CON-PRO (tot. CON = 5)	**Quand** - t - PRO	100%	**ADV-ADJ (tot. ADV monosyll = 5)**	**Très** - z -ADJ **Plus** - z - ADJ	96%
PRP-DET (tot. PRP monosyll = 7)	**Dans** - z - DET	92%	**VER:pres-PRP (tot. VER = environ 400)**	**Est** - t - PRP **Sont** - t - PRP	93%
VER-DET (tot. VER = environ 300)	**Est** - t - DET **Était** - t -DET **Fait** - t -DET	95%	**ADJ-NOM (tot. ADJ = 53)**	**Bon** - n - NOM **Autres** - z -NOM **Petit** - t - NOM **Petit(e)s** - z - NOM **Grand** - t - NOM	77%
PRP-NOM (tot. PRP monosyll= 8)	**En-** n -NOM	96%			

Tableau 18 : Constructions internes aux types de liaison variable les plus fréquents (n. tokens > 100)

Les seules constructions décrites dans le tableau ci-dessus (cf. Tableau 18) permettent de rendre compte de 64% des réalisations totales de la liaison variable (à savoir 2755 occurrences de liaisons réalisées sur 4296 liaisons totales réalisées).

2.2.3. Étude supplémentaire de la liaison après adjectifs épithètes

Les données issues de l'analyse fréquentielle de la liaison après adjectif épithète antéposé nous ont amené à approfondir l'étude de ce contexte et à élargir la recherche à un autre corpus de français oral le « Corpus de Français Parlé Parisien » (dorénavant *CFPP2000*) (Branca-Rosoff et al. 2012). Grâce à l'analyse des deux corpus PFC et CFPP2000 (Barreca et al. 2015), nous avons tenté d'apporter des éléments de réponse aux hypothèses suivantes formulées par les études qui ont précédé : (a) la liaison est variable dans toute la classe d'adjectifs mais il existe des contextes catégoriques (ex « petit », « grand ») (Laks 1977 ; Durand et Lyche 2008 ; Mellet 2008) (cf. Partie 1.Chapitre 4.3.2.4.2) ; (b) la réalisation de la liaison est plus fréquente après les adjectifs pluriels qu'après les adjectifs singuliers (Morin et Kaye 1982 ; Bonami et al. 2005 ; Mallet 2008) ; (c) les adjectifs singuliers susceptibles d'apparaître en contexte de liaison possèdent une distribution limitée (Bybee 2005). Des propositions de réponses à ces hypothèses seront présentées dans les pages qui suivent. Avant de passer aux spécificités de la présente étude il nous paraît nécessaire de faire un bilan des approches théoriques proposées par la littérature pour l'analyse de ce contexte de liaison.

2.2.3.1 Approches théoriques et traitement de la liaison après adjectif épithète

La liaison après adjectifs épithètes a été l'objet de nombreux débats. Parmi les approches proposées pour l'étude de ce contexte, nous pouvons citer l'analyse par troncation proposée par Schane (1965) (cf. Partie 1.Chapitre 5.3.1.1) ou encore l'analyse morpho-phonologique de Steriade (1999) qui, dans le cadre de la théorie de l'optimalité suggère d'interpréter la forme de l'adjectif masculin en contexte de liaison et la forme du féminin comme dérivées de la même représentation sous-jacente qui comprend la CL. Celles-ci seraient utilisées en combinaison avec un nom masculin, si ce choix permet d'éviter le *hiatus*. En revanche, contrairement aux approches précédentes, dans les analyses autosegmentales (Clements et Keyser 1983 ; Tranel 1990 ; Booij 1984 ; Wetzels 1987), le recours à la notion de flottance permettra de distinguer le statut des CL de celui des consonnes fixes (tout en gardant une représentation sous-jacente unique). Cependant ces analyses rendent difficilement compte des formes des adjectifs masculins en contexte de liaison (dorénavant FMSL) non prédictibles à partir des formes du féminin (dorénavant FF). C'est le cas par exemple des adjectifs présentant une alternance /s/ - /z/ (« bas », « doux », « faux » et « gros ») ou encore /d/ - /t/ (« grand », « profond » et «second »). De plus, le test expérimental mené par Morin (1992) montre que les sujets ne sont pas capables de produire des FMSL qu'ils ne connaissent pas (ex. « franc entretien ») et donc de dériver automatiquement, par une règle intériorisée, la FMSL de la FF. La FMSL « doit être apprise » (Morin 1992, p. 242). Si les analyses supplétives proposées par Bonami et Boyé sous forme de paradigme à 5 cases (Bonami et Boyé 2003) ou d'espace thématique (Bonami et Boyé 2005) permettent d'éviter de postuler une dérivation directe des FMSL des FF elles semblent toutefois présenter quelques inconvénients. En effet Bonami et Boyé considèrent les cas des adjectifs analysés par Morin (1992) (« blanc » et « franc ») et d'autres adjectifs (« chaud », « froid », « brun » et « sot ») comme des exemples d'adjectifs ayant un paradigme défectif pour la FMLS. Nous avons montré à travers des requêtes réalisées sur les deux corpus oraux PFC et CFPP2000 et un corpus écrit FFT (*Corpus Free French Treebank*)[163] (Hernandez et Boudin 2013) que ces adjectifs ne sont pas défectifs pour la FLMS, mais qu'ils sont tout simplement rarement antéposés. En effet, on ne trouve jamais (ou très rarement) de formes de ces adjectifs[164] en position antéposée, que ce soit en contexte de liaison ou de non-liaison.

[163] Ce corpus écrit est constitué de 28.000 articles de presse *frwikinews* pour un total de 2.535.396 tokens.

[164] Nous avons enregistré des occurrences d'une seule séquence « chaud partisan » où l'adjectif « chaud » était antéposé.

De plus l'approche supplétive, à l'instar de l'analyse par troncation, ne permet pas de rendre compte des cas de liaison non attendue ou de substitution attestés chez les natifs (Desrochers 1994)[165]. Des analyses des CL comme consonnes épenthétiques, comme consonnes morphémiques (préfixales ou suffixales) (cf. Partie 1.Chapitre 5.2.4), ou comme consonnes intégrées à des constructions sembleraient en revanche offrir un cadre théorique plus adéquat pour l'étude de ces formes. Nous reviendrons sur cet aspect dans les paragraphes qui suivent.

Quant aux sources de ces travaux, la plupart des approches que nous venons d'évoquer s'appuient sur des données extraites : (a) des exemples obtenus par introspection (Fouché 1959 ; Delattre 1966); (b) des lectures oralisées (Morin et Kaye 1982 ; Durand et Lyche 2008 ; Post 2000 ; Sampson 2001 ; Morin 1992). Celle de Durand et Lyche (2008) est à notre connaissance la seule étude à avoir analysé ce type de liaison à partir de la parole semi-spontanée (entretiens guidés et conversations libres) de 10 points d'enquête PFC.

2.2.3.2 Objectifs et corpus

À la suite de ces travaux, l'objectif de cette sous-étude est donc d'analyser le comportement de la liaison de l'adjectif épithète antéposé à partir des données de parole semi-spontanée extraites du corpus CFPP2000 (Branca-Rosoff et al. 2012) et du corpus PFC (Durand, Laks et Lyche 2009) en essayant de fournir des éléments de réponse aux hypothèses que nous avons évoquées plus haut. L'échantillon des données que nous avons recueillies correspond à 403 occurrences de liaison potentielle correspondant à 34 lemmes d'adjectifs.

2.2.3.3 Résultats

Les résultats totaux de notre échantillon montrent des taux de réalisations de la liaison très élevés d'environ 87%. Cela rejoint les conclusions des études précédentes qui décrivent la liaison dans ce contexte comme très fréquente mais pas catégorique d'un point de vue statistique. Ces premiers résultats nous permettent de vérifier la variabilité de ce type de liaison et plus précisément de savoir s'il existe des contextes catégoriques de liaison (hypothèse a) (cf. Paragraphe 2.2.3.2). Nos données suggèrent que la réalisation de la liaison est variable dans toute la classe d'adjectifs même pour les adjectifs les plus fréquents comme « petit » et « grand ».

[165] Parmi celles-ci nous pouvons citer des exemples extraits de l'étude de Desrochers (1994) : « j'ai fait un long [z] article » ; « un long [t] article » ; « gros [t] homme », « gros [t] avion » ; « grosse [t] enveloppe » et « ils ont un sacré [t] accent ».

Ensuite, afin d'apporter des éléments de réponse à l'hypothèse formulée par Morin et Kaye (1982) (hypothèse b) (cf. Paragraphe 2.2.3.2) (avant d'être reprise par des études suivantes) selon laquelle la réalisation de la liaison serait plus fréquente au pluriel qu'au singulier, nous avons mené une étude séparée des deux contextes. En ce qui concerne les adjectifs singuliers, le nombre d'adjectifs susceptibles de se trouver en contexte de liaison serait très limité (parmi d'autres Bybee 2005) et correspondant, en ordre de réalisations (LR), aux 12 lemmes présentés dans le tableau ci-dessous (cf. Tableau 19).

Lemme	**LR**
Bon	29
Petit	25
Grand	19
Saint	16
Certain	7
Mauvais	6
Plein	3
Ancien	2
Bas	1
Moyen	1
Prochain	1
Gros	0
	TOT LR 110

Tableau 19 : Réalisations de la liaison après adjectifs singuliers (lemme) du type ADJ-NOM

En revanche, le nombre d'adjectifs pluriels susceptibles de se trouver en contexte de liaison est bien évidemment plus important qu'en contexte singulier : 29 lemmes comprenant les 12 lemmes déjà rencontrés au singulier (cf. Tableau 20).

Lemme	**LR**	**Lemme**	**LR**
Autre	48	Long	3
Petit	38	Propre	3
Grand	29	Ancien	3
Bon	20	Différent	3
Même	12	Futur	3
Dernier	12	Joli	2
Vieux	11	Énorme	1
Beau	9	Pauvre	1
Nouveau	9	Certain	1
Gros	7	Fameux	1

Jeune	6	Prochain	1
Vrai	5	Nombreux	0
Meilleur	4	Immense	0
Faux	4	Véritable	0
Mauvais	3		
			TOT LR 239

Tableau 20 : Réalisations de la liaison après adjectifs pluriels (lemme) du type ADJ-NOM

À la lumière de ces résultats, plutôt que d'essayer de comparer les deux contextes de liaison il nous a paru plus pertinent de parler de deux types différents de liaison. D'un côté la liaison après un groupe restreint d'adjectifs masculins singuliers. Cette liaison est caractérisée par une plus grande variabilité de CL contrairement à la liaison après les adjectifs pluriels qui constitue une liaison morphologique (comme cela a déjà été souligné par des études précédentes Morin 1992 ; Côté 2005) où la CL /z/ est grammaticalisée et constitue une marque du pluriel. C'est dans ce sens que la liaison après les adjectifs pluriels peut être considérée comme plus productive puisqu'elle s'applique à un plus grand nombre d'adjectifs pluriels masculins et féminins.

Ces observations nous ont amené à nous interroger sur les caractéristiques distributionnelles de ces adjectifs. Plus précisément nous essayerons de vérifier l'hypothèse, formulée par Bybee (2005) (hypothèse c cf. Paragraphe 2.2.3.2), selon laquelle les adjectifs singuliers susceptibles d'apparaître en contexte de liaison possèdent une distribution limitée[166]. Nous avons donc mené une analyse distributionnelle de ces adjectifs, tous contextes confondus dans les deux corpus oraux PFC et CFPP2000.

[166] Cette idée n'est pas nouvelle, bien que des combinaisons adoptées par certains linguistes modernes (ex. « sot ami ») fassent sourire (Durand et Lyche 2008). En effet, Clédat (1890) rappelait déjà à la fin du XIXème siècle que : « lorsqu'un adjectif masculin terminé [*phonétiquement*] par une voyelle se trouve placé (**ce qui est rare**) devant le nom auquel il se rapporte, et que le nom commence par une voyelle, on lie l'adjectif au nom par la consonne qui s'ajoute pour former le féminin » (Clédat 1890, p. 40).

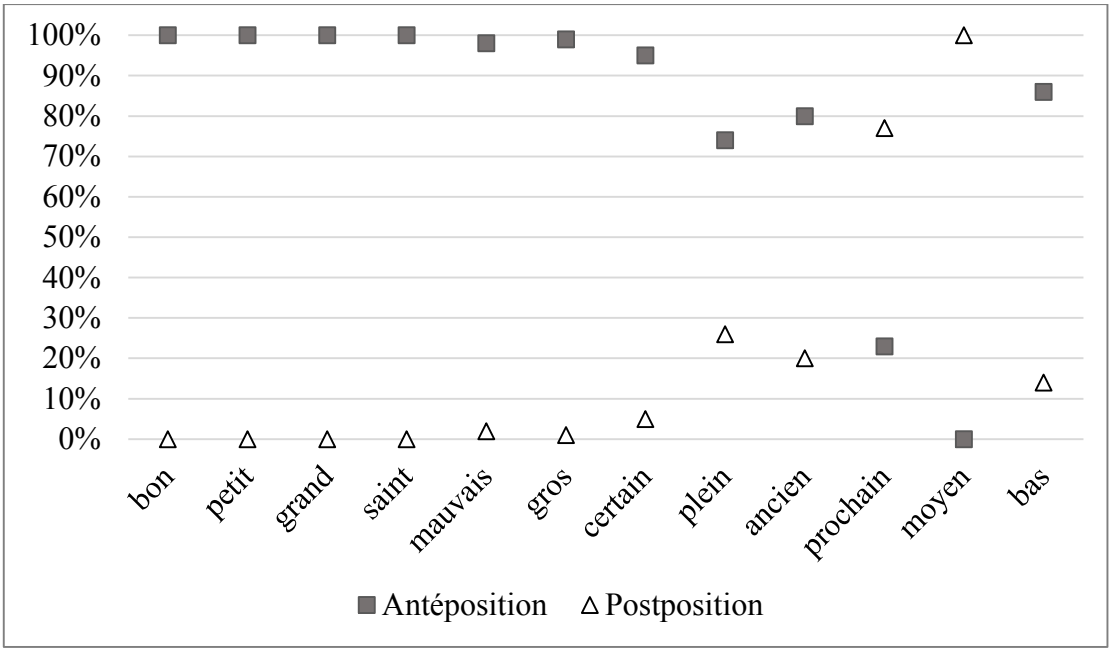


Figure 20 : Caractéristiques distributionnelles (antéposition / postposition par rapport au nom) tous contextes confondus des adjectifs singuliers observés en contexte de liaison

Comme nous pouvons l'observer dans la figure ci-dessus (cf. Figure 20), la plupart des adjectifs que nous avons observés en contexte de liaison se trouvent en position presque uniquement antéposée (marquée par des carrés) (« bon », « petit », « grand » et « saint » et « gros ») dans tous contextes confondus. En revanche, d'autres adjectifs (« certain », « plein », « ancien », « prochain » et « moyen ») possèdent une distribution moins régulière qui reflète la distribution complémentaire de ce groupe d'adjectifs. Quant à l'adjectif « bas », pour lequel nous n'avons enregistré qu'une seule occurrence, il semble suivre une distribution libre.

Toutefois il existe des cas d'adjectifs majoritairement antéposés qui n'apparaissent que rarement ou jamais en contexte de liaison. C'est le cas de l'adjectif « gros » déjà évoqué par (Durand et Lyche 2008) et « long » (cf. Tableau 21).

	Antéposition			**Postposition**
Lemme	Contexte liaison ADjsg	Contexte liaison ADjpl	Tout autre contexte	Tout contexte confondu
Gros	1	5	323	3
Long	0	4	47	12

Tableau 21 : Caractéristiques distributionnelles (antéposition / postposition par rapport au nom) tous contextes confondus des adjectifs « gros » et « long »

Nous pouvons supposer que les locuteurs évitent d'antéposer les adjectifs « gros » et « long » puisqu'ils ne savent pas quelle CL utiliser (Morin 2003 ; Durand et Lyche 2008). Cette hypothèse serait étayée par les cas de substitutions de la CL attestés par Desrochers (cf. Partie 1.Chapitre 4.2.4.3) (ex. « gros [t] enfant »).

Nous pouvons en conclure que si la liaison est limitée à un groupe restreint d'adjectifs singuliers (Bybee 2001a ; Bybee 2005) (à savoir des adjectifs majoritairement antéposés et des adjectifs ayant une distribution complémentaire), la distribution ne représente cependant pas le seul facteur à rendre compte de la réalisation de la liaison après adjectifs prénominaux. En effet l'absence d'antéposition des adjectifs « long » et « gros » en contexte de liaison suggèrent que des stratégies d'évitement peuvent être mises en place par les locuteurs. Notamment des stratégies d'évitement : (a) d'un schéma peu accessible en faveur d'un schéma plus productif (« un gros - z - écran » moins accessible par rapport à « un grand - t - écran ») ; (b) d'une CL rare en faveur d'une postposition (« long - g - article » / « un article long »).

On pourrait à ce moment-là se poser la question de savoir si la fréquence de réalisation de la liaison est liée au type de distribution de l'adjectif singulier et donc notamment si la fréquence relative des liaisons est plus importante après des adjectifs majoritairement antéposés (« bon », « grand » et « petit ») qu'après des adjectifs en distribution complémentaire (« ancien », « prochain » etc.).

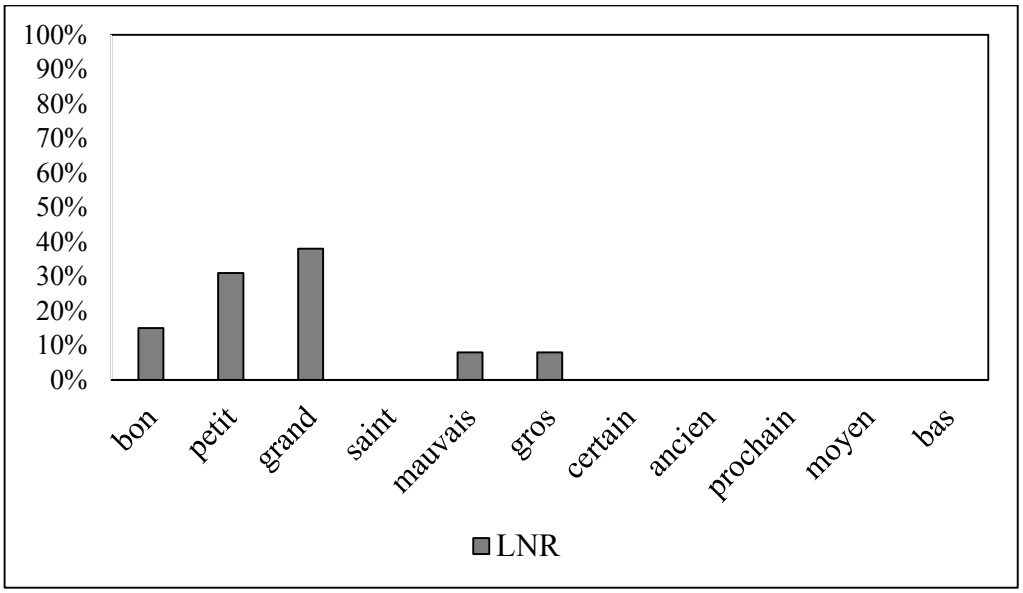


Figure 21 : Fréquence relative des liaisons non-réalisées par adjectif

Cette dernière hypothèse ne semble toutefois pas vérifiée puisque les adjectifs majoritairement antéposés (« bon », « grand » et «petit » etc.) rendent compte de 80% des liaisons non-réalisées (LNR) (cf. Figure 21).

2.2.3.4 Bilan

En résumé, les résultats de cette étude supplémentaire semblent nous indiquer que :

- la réalisation de la liaison est variable dans toute la classe d'adjectifs (à savoir l'hypothèse a n'est pas vérifiée cf. Paragraphe 2.2.3.2) ;
- les différences de comportements entre les deux contextes ne permettent pas d'affirmer si la réalisation de la liaison est plus fréquente après les adjectifs pluriels qu'après les adjectifs singuliers (à savoir l'hypothèse b n'est pas vérifiable cf. Paragraphe 2.2.3.2). C'est la raison pour laquelle nous avons préféré parler de deux types de liaisons différents ;
- la liaison est limitée à un groupe restreint d'adjectifs singuliers (Bybee 2001a; 2005) à distribution limitée (l'hypothèse c est vérifiée cf. Paragraphe 2.2.3.2) mais il existe des cas d'adjectifs majoritairement antéposés (« long » et « gros ») qui n'apparaissent rarement ou jamais en contexte de liaison ;
- la fréquence relative des liaisons non réalisées n'est pas plus importante après des adjectifs singuliers en distribution complémentaire (« ancien », « prochain » etc.) qu'après des adjectifs singuliers majoritairement antéposés (« bon », « grand » et « petit »). La distribution n'est donc pas le seul facteur à rendre compte de la liaison après adjectifs épithète et en particulier de sa variabilité.

Pour conclure, ces résultats semblent étayer l'hypothèse formulée par l'approche exemplariste (Bybee 2001a ; Bybee 2005) qui considère que la CL des adjectifs est encodée dans des constructions (ex. ADJ-NOM). Cette analyse permettrait de rendre compte de la productivité de ce type de liaison qui pourrait être expliquée comme le résultat de la distribution limitée des adjectifs épithètes à l'oral et donc du nombre restreint de schémas accessibles dans ce type de liaison. Une telle approche permettrait également d'expliquer les cas de substitution de la CL et d'évitements que nous avons évoqués dans les paragraphes précédents. Nous proposons donc deux représentations différentes pour les deux contextes de liaison. D'une part des constructions de liaison limitées à un groupe restreint d'adjectifs majoritairement antéposés (ex. « grand - t - NOM ») et à des adjectifs en distribution complémentaire (ex. « ancien - n - NOM »). D'autre part une représentation basée sur un schéma plus ouvert et productif correspondant à la liaison grammaticalisée après les adjectifs pluriels (ex. « ADJ - z - NOM »). Ces représentations permettent de rendre compte des : (a) substitutions de la CL par analogie avec des constructions plus productives (« gros [t] enfant » par analogie avec « grand - t - enfant ») en particulier dans les cas d'évitement de liaison après des adjectifs rarement

antéposés et / ou possédant une CL rare (« franc [t] entretien » ou « long [t]article ») ; (b) des liaisons non attendues résultant de la sur-généralisation de la construction du pluriel (« ils ont un sacré [z] accent »). Ces conclusions partielles engendrent un certain nombre de questions et en particulier une question principale qui reste encore sans réponse : quels sont les facteurs qui influencent la variabilité (même si faible) de la liaison dans ces types de liaison ? Nous reviendrons sur l'étude de ce type de liaison variable dans les pages qui suivent (cf. Paragraphe 3).

2.3. Discussion

L'analyse fréquentielle menée sur la parole semi-spontanée dans le corpus PFC nous a permis de montrer que la liaison variable est aussi limitée à un petit nombre de types qui rendent compte de 80% des réalisations totales.

De plus au sein de ces types de liaison variable, plus de la moitié des réalisations correspondent à un nombre restreint de constructions très productives. Il reste cependant à identifier les causes de cette forte productivité. Ceci nous oblige à revenir à la définition de la notion de cohésion de la séquence de liaison que nous avons abordée au début de ce travail (cf. Partie 1.Chapitre 2.2.3.1).

Deux hypothèses différentes seront discutées. La première relève des approches exmplaristes d'étude de la liaison (cf. Paragraphe 2.3.1.1) et attribue un rôle central aux variables lexicales comme facteurs explicatifs de la productivité de la liaison. La deuxième est issue de l'étude de Côté (2013) (cf. Paragraphe 2.3.1.2) et voit dans les probabilités transitionnelles qui caractérisent les séquences de liaison une possible explication de la frequence de réalisation de la liaison.

2.3.1. Cohésion et productivité des constructions spécifiques : entre syntaxe, fréquence et probabilité transitionnelle

Comme nous l'avons déjà observé (cf. Partie 1.Chapitre 2.1), la notion de cohésion syntaxique a reçu plusieurs définitions de nature différente : syntaxique (entre autres Damourette et Pichon 1968 ; Delattre 1955), prosodique (cf. Côté 2012), lexicale (Bybee 2005 ; Chevrot, Chabanal et Dugua 2007) et statistique (Côté 2013)[167]. Le degré de cohésion d'une séquence de liaison a souvent été considéré comme un facteur jouant un rôle crucial dans sa productivité et donc plus en général dans l'étude de la variabilité de la liaison.

[167] Nous reviendrons sur les autres définitions de cohésion dans les pages qui suivent (cf. Partie 2.Chapitre 2.2.3.1).

2.3.1.1 Hypothèse exemplariste

D'après l'approche exemplariste (Bybee 2001a ; Bybee 2005), les séquences de liaison M1-M2 ayant des fréquences de cooccurrence élevées (et donc une cohésion syntaxique plus importante) seraient aussi le plus souvent réalisées puisque plus facilement accessibles dans la mémoire du locuteur (dans laquelle elles seraient stockées sous formes de constructions). Selon cette hypothèse, la liaison serait plus productive après les M1 ayant des indices de fréquence plus élevés.

En revanche, Chevrot et al. (2007) tenteront de considérer cette productivité comme le résultat du rapport existant entre le nombre d'occurrences du M1 dans un contexte donné et le nombre de différents M2 susceptibles d'apparaître après M1 à l'intérieur d'une séquence de liaison (cf. Tableau 22).

$$\text{Fréquence de la séquence de liaison} = \frac{\text{Nombre d'occurrences du M1}}{\text{Nombre de M2 différents}}$$

Tableau 22 : Calcul de l'indice de la fréquence d'une séquence de liaison d'après Chevrot et al. 2007 (tiré de Côté 2013)

Malgré ces différences, les approches exemplaristes s'accordent sur le fait que les taux les plus importants de réalisation de la liaison associés aux M1 les plus fréquents (ex. « est »), au sein du même type de liaison variable (ex. « VER-DET »), pourraient être interprétés comme le résultat d'une catégorisation de ces items qui apparaissent de façon récurrente dans le contexte gauche du type en question (Bybee 2005, p. 26). Ceux-ci donnent lieu à des constructions facilement accessibles aux locuteurs qui produisent presque uniquement des liaisons au sein de ces séquences.

Il convient cependant de souligner que le taux de non-réalisation même après les items lexicaux les plus fréquents (ex. « est » - DET) reste parfois très élevé (ex. 67% dans le corpus PFC) ce qui confirme que les facteurs syntaxiques et lexicaux ne suffisent pas à décrire la variabilité de la liaison. En effet, la définition de ces repères ne permet pas de prédire l'occurrence d'une liaison variable.

2.3.1.2 Hypothèse basée sur la probabilité transitionnelle

D'après Côté (2013), les différentes hypothèses de fréquence formulées par les approches constructionnistes ne semblent pas expliquer l'hétérogénéité observée au sein de certains types de liaison (ex. unités monosyllabiques). C'est pour cette raison que Côté (2013) proposera une analyse basée sur la probabilité transitionnelle des séquences de liaison à partir du corpus des

variétés de français laurentien. Elle conduira en particulier une étude sur la base d'un ensemble de mots invariables monosyllabiques constitué par : des adverbes (« mieux », « moins », « pas », « plus », « très » et « trop ») des prépositions (« chez » et « dans ») et des connecteurs (« mais », « dont » et « quand »).

Le tableau ci-dessous (cf. Tableau 23) présente les résultats de l'étude menée par Côté (2013) et qui illustre dans l'ordre : (a) le M1; (b) le nombre de tokens à savoir le nombre d'occurrence du M1 en contexte de liaison ; (c) le taux de réalisation de la liaison ; (d) le nombre des différents M2 observés après un M1 donné en contexte de liaison ; (e) le ratio entre le nombre d'occurrence (tokens) du M1 et le nombre des différents M2 observés après le M1 en contexte de liaison (Chevrot et al. 2007).

M1	Nombre de tokens	% Réalisation liaison	Nombre de M2 différents	Nombre de tokens / Nombre de M2 différents
Très	25	84	10	2,5
Quand	80	78	4	23,5
Dans	78	72	3	26
Dont	3	67	2	1,5
Chez	10	60	4	2,5
Plus	22	59	17	1,3
Trop	7	14	5	1,4
Pas	163	0,6	40	4,1
Puis	162	0	13	12,5
Mais	122	0	20	6,1
Moins	10	0	7	1,4
Mieux	2	0	2	1

Tableau 23 : Résultats des réalisations de la liaison après des mots invariables monosyllabiques (Côté 2013)

Sur la base de l'ensemble de ces données Côté (2013) rejette l'hypothèse exemplariste selon laquelle la liaison serait plus souvent réalisée après les M1 les plus fréquents puisqu'elle observe que la liaison n'est jamais produite après des adverbes et des connecteurs monosyllabiques très fréquents telles que « puis », «pas » et « mais » (cf. Tableau 23).

> « The prediction is that the words that appear most frequently in a liaison context, as indicated in the second column, will be associated with a higher rate of liaison realization. This is not what we observe here: the three most frequent words (*puis, pas, mais*) never trigger liaison, and liaison may or may not be productive with the less frequent ones ». (Côté 2013, p. 161)

Il en va de même pour l'hypothèse formulée par Chevrot et al. (2007) d'après laquelle la productivité de la liaison serait conditionnée par la fréquence de la séquence M1-M2 (cf. Tableau 22). En effet malgré le ratio important obtenu par « puis », le taux de réalisation de la liaison observé après ce connecteur serait très faible.

> « In other words, the productivity of liaison (third column) is expected to be positively correlated to the average frequency of specific W1–W2 sequences (fifth column). Again, this measure does not appear to account for the grouping in Table 3. *Quand* and *dans* pattern as expected: they have the highest average frequency of W1–W2 sequences and they trigger liaison quite productively. However, liaison is excluded with *puis,* despite its relatively high ratio, and words with very low average frequencies are found equally in both groups ». (Côté 2013, p. 163)

À la suite de cette analyse, Côté (2013) propose de considérer la productivité d'une séquence de liaison en termes de fréquence statistique en s'appuyant sur la probabilité transitionnelle des séquences de liaison. L'idée de base est que la productivité d'une liaison entre un M1 et un M2 (ex. « quand il ») est conditionnée par la probabilité transitionnelle (Seidenberg 1997; Saffran 2003)[168] qu'un M2 (ex. « il ») ou sa catégorie morphosyntaxique (ex. pronom personnel sujet) suive le M1 (ex. « quand »).

> « […] liaison remains productive with words that surface almost exclusively in the same construction and whose transition probability to a given category of words approaches 100% […]. The productivity of liaison between W1 and (the category of) W2 is determined by the transition probability from W1 to (the category of) W2 ». (Côté 2013, p. 163)

Autrement dit, parmi les connecteurs, « quand » et « dont » déclencheraient un nombre plus important de liaisons puisqu'elles sont suivies dans la plupart des cas par les seuls pronoms personnels « elle (s) », « il(s) » et « on ». En revanche, bien que les connecteurs « mais » et « puis » apparaissent plus souvent que « quand » et « dont » devant ces pronoms, elles possèdent des taux de réalisation de la liaison très faibles puisqu'elles sont susceptibles d'être suivies par un plus grand nombre de M2.

Côté (2013) étend également cette analyse basée sur des probabilités transitionnelles aux séquences de liaison ayant comme M1 des adverbes monosyllabiques. Le tableau ci-dessous (cf. Tableau 24) montre les probabilités transitionnelles de chaque adverbe d'être suivi par un adverbe ou un adjectif ainsi que le taux de réalisation de la liaison.

[168] « The simultaneous satisfaction of multiple probabilistic constraints provides a way of rapidly and efficiently combining different sources of information. Importantly, the interactions among constraints are nonlinear: bits of information that are not very constraining in isolation become quite informative when taken in conjunction with other bits of information ». (Seidenberg 1997, p. 1602)

M1	Nombre total de tokens	Nombre de tokens exclus	Nombre de tokens suivis d'un adverbe / adjectif	Probabilité transitionnelle	% Réalisation liaison (corpus codé)
Très	521	108	412	100% (412/413)	84
Plus	787	140	627	97% (627/647)	59
Trop	255	9	131	53% (131/246)	14
Moins	303	92	89	42% (89/211)	0
Pas	5189	623	1297	28% (1297/4566)	0,6
Mieux	132	5	7	6% (7/127)	0

Tableau 24 : Probabilités transitionnelles des adverbes monosyllabiques (Côté 2013)

La productivité de la réalisation de la liaison après les adverbes « très » et « plus »[169] pourrait donc être expliquée comme la conséquence de la forte probabilité transitionnelle de ces adverbes d'être suivis par des adjectifs et des adverbes. En revanche, dans le cas des autres adverbes « trop », « moins », « pas » et « mieux » après lesquels la liaison est peu productive, la probabilité transitionnelle d'apparaître devant des adjectifs et des adverbes varie de 6% à 53% puisqu'ils sont également susceptibles de se trouver devant d'autres catégories (ex. prépositions, des verbes, etc.).

C'est dans une même perspective que Côté (2013) compare la diversité de production qui caractérise la liaison de l'adjectif antéposé et postposé au nom.

> « Determiner - adjective sequences, however, are always followed by a noun (unless the adjective itself is nominalized). The transition probability from a determiner - adjective combination to a noun is therefore maximal, and liaison is expected to be productive, despite its low frequency. The opposite combination noun - adjective, where liaison is marginal and restricted to plural contexts, offers an interesting contrast. A determiner - noun sequence may be utterance-final (e.g. *j'ai vu l'étudiant* 'I saw the student') or followed by a variety of elements: an adjective (e.g. *l'étudiant français* 'the French student'), a verb (e.g. *l'étudiant arrive* 'the student arrives'), a prepositional phrase (e.g. *l'étudiant dans la classe* 'the student in the classroom'), etc. The transition probability from a noun to any given category is therefore necessarily limited. I suggest that the different transition probabilities within adjective - noun and noun - adjective sequences explain the observed difference in the productivity of liaison ». (Côté 2013, p. 165)

[169] L'adverbe négatif « plus » n'est pas analysé par cette étude puisqu'il correspond dans la variété de français laurentien à une prononciation distincte [py] qui exclut la réalisation d'une liaison.

« *Plus* 'more': This word poses difficulties since the written form plus corresponds to three different prononunciations: [ply], [plʏs] and [py]. The distribution of the three variants is complex, but may be summarized as follows. The reduced form [py] is found only in negative contexts (e.g. *je parle plus* [py] 'I don't speak anymore'), while the positive contexts are shared between [plʏs] and [ply] (e.g. *je parle plus* [plʏs] 'I speak more'; *plus* [ply] *vite* 'faster'). The form [ply] is also found in a number of locutions (e.g. *plus ou moins* 'more or less', *non plus* 'neither') and it may replace [py] and, in some cases, [plʏs] in very careful registers. Crucially, only [ply] may be followed by a liaison [z] and is relevant here. All instances of *plus* pronounced [py] or [plʏs] were therefore excluded from consideration ». (Côté 2013, p. 160) (voir aussi Côté 2012).

Bref, d'après l'étude de Côté (2013), les comportements de la liaison à l'intérieur d'une même classe morphosyntaxique (dans ce cas celles des adverbes, des connecteurs monosyllabiques) seraient influencés par la diversité des probabilités transitionnelles qui caractérisent les différentes unités et non pas par des variables fréquentielles[170]. Dans les conclusions, Côté (2013) avance que les causes de la productivité de la liaison pourraient être recherchées dans des contraintes psycholinguistiques qui conduiraient le locuteur à réaliser plus de liaisons lorsque le M2 est le plus prévisible et donc le plus accessible.

> « […] a possible psycholinguistic motivation for the productivity of liaison. Liaison consonants resyllabify with the initial vowel of the following word, which potentially masks the onset of W2 and interferes with lexical access. It may not be a coincidence, then, that liaison is maintained when the category of W2 is fully predictible, that is precisely when the lexical space of W2 is reduced and lexical access facilitated ». (Côté 2013, p. 165)

Si l'étude proposée par Côté (2013) semble affirmer que la syntaxe[171] n'est pas en mesure de rendre compte des différences de productivité observées au sein de certains types de liaison (ex. adverbes monosyllabiques), il nous semble que les tendances dégagées par l'analyse des probabilités transitionnelles constituent l'effet de la diversité qui caractérise les rapports de dépendances syntaxiques des mots de liaison.

Dans l'analyse des adverbes monosyllabiques, l'adverbe « pas » est syntaxiquement gouverné par le verbe précédent alors que « très » dépend de l'adjectif suivant avec lequel il réalise la liaison de façon presque systématique. De même, la probabilité transitionnelle des adverbes « moins » et « trop » peut varier en fonction des valeurs de dépendance différentes de ces adverbes qui peuvent être gouvernés par le verbe comme par l'adjectif.

Il en va de même pour l'adverbe « plus », qui dans des varietés autres que celle du français laurentien est également susceptible de déclencher une liaison lorsqu'il possède une valeur négative (ex. « ils ont mal, ils arrivent plus à marcher »). En effet, si on adopte une approche purement probabiliste, la présence de cette position comme site de liaison potentielle réduirait les probabilités transitionnelles de « plus » d'être suivis d'un adjectif ou d'un adverbe.

Les figures ci-dessous illustrent les résultats de l'analyse syntaxique de deux énoncés tirés du corpus PFC contenant respectivement d'une part l'occurrence de l'adverbe « plus » à valeur négative (cf. Figure 22) et de l'autre (cf. Figure 23) l'occurrence de l'adverbe d'intensité « plus ». L'analyse syntaxique réalisée à l'aide de l'analyseur syntaxique *FRench*

[170] La seule exception à cette généralisation serait représentée par le taux élevé de réalisations attesté après la forme verbale « est » pour lequel l'influence de fréquence a été plusieurs fois évoquée dans la littérature (cf. Partie 1.Chapitre 2.2.4.2).

[171] « […] in the same adverb - adjective combination (e.g. *très / plus / bien / pas utile* 'very / more / quite / not useful'), syntactic or prosodic structure does not predict the lexical differences observed » (Côté 2013, p. 157).

MetaGrammar (FRMG Wiki)[172] (De la Clergerie et al. 2009) montre clairement que dans le premier énoncé (« ils ont mal, ils arrivent plus à marcher ») (cf. Figure 22) l'adverbe négatif « plus » est gouverné par le verbe « arrivent » alors que l'adverbe d'intensité issu du deuxième énoncé (« C'est plus intéressant d'enseigner au lycée ») (cf. Figure 23) dépend de l'adjectif suivant. Dans cette perspective l'adverbe négatif « plus »[173] comme « pas » aurait moins de probabilité d'être suivi par un adjectif ou par un adverbe puisqu'il dépend du verbe qui le précède. En revanche les adverbes d'intensité « plus » et « très » ne pourront qu'être suivis d'un adjectif ou d'un adverbe par lesquels ils sont gouvernés.

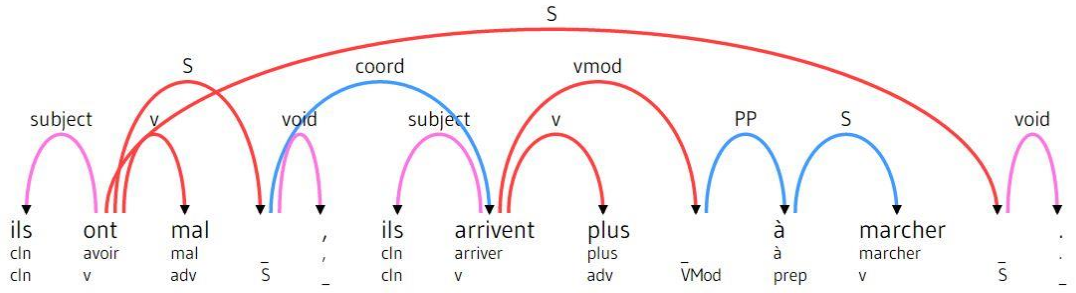


Figure 22 : Analyse syntaxique de l'énoncé « ils ont mal, ils arrivent plus à marcher » (Corpus PFC, Grenoble, 38aca1gg) réalisée par le biais de l'analysateur syntaxique *FRMG Wiki* (De la Clergerie et al. 2009)

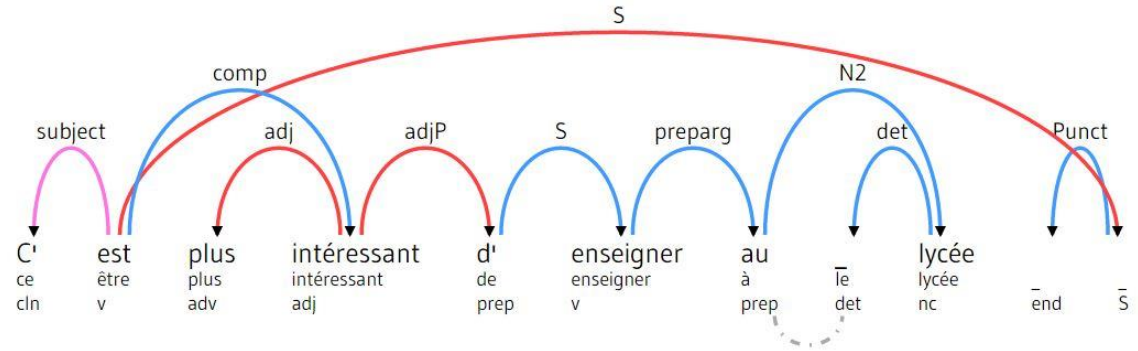


Figure 23 : Analyse syntaxique de l'énoncé « C'est plus intéressant d'enseigner au lycée » (Corpus PFC, Toulouse, 31afb1lg) réalisée par le biais de l'analysateur syntaxique *FRMG Wiki* (De la Clergerie et al. 2009)

[172] L'analyseur FRMG s'appuie sur la grammaire FRMG, méta-grammaire du français développée par l'équipe ALPAGE et correspondant à une description syntaxique sous forme de contraintes qui permet de produire une grammaire d'arbres adjoints (TAG).

[173] Dans l'analyse des adverbes monosyllabiques nous distinguons le comportement de l'adverbe négatif « pas » des adverbes d'intensité (« très » etc) d'autant plus que cette différence se reflète également dans la prononciation des différentes valeurs que l'adverbe « plus » peut recouvrir dans la variété de français laurentien (cf. Note 169) (Côté 2013).

Quant aux connecteurs, il nous paraît que la diversité des probabilités distributionnelles observée entre d'une part « quand » et « dont » et d'autre part « mais » et « puis » est la conséquence des rapports de subordination et coordination qui opposent ces connecteurs. La plus grande probabilité que les connecteurs « quand » et « dont » soient suivis par un pronom clitique (cf. Côté 2013) pourrait donc être liée au fait qu'ils introduisent une proposition subordonnée (contrairement aux connecteurs « mais » et « puis »).

Les figures ci-dessous présentent ces deux types de relations qui permettent de distinguer « quand » de « mais ». D'une part la subordination (marquée par l'étiquette « SubS ») (cf. Figure 24) et de l'autre la coordination (marquée par l'étiquette « coord3 ») (cf. Figure 25).

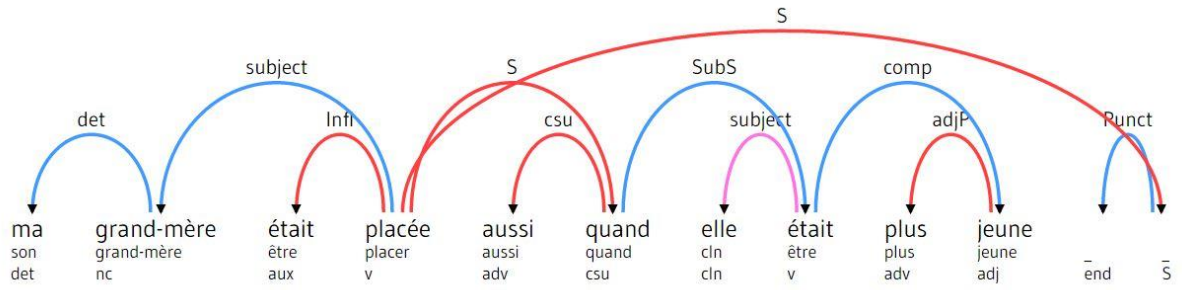


Figure 24 : Analyse syntaxique de l'énoncé « ma grand-mère était placée aussi quand elle était plus jeune » (Corpus PFC, Grenoble 38agp1lg) réalisée par le biais de l'analysateur syntaxique *FRMG Wiki* (De la Clergerie et al. 2009)

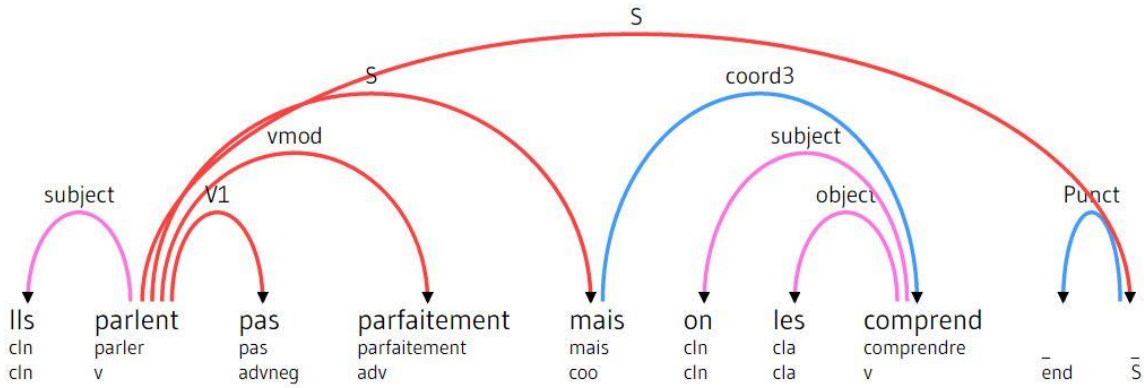


Figure 25 : Analyse syntaxique de l'énoncé « ils parlent pas parfaitement mais on les comprend » (Corpus PFC, Toulouse 31atc1gg) réalisée par le biais de l'analysateur syntaxique *FRMG Wiki* (De la Clergerie et al. 2009)

En ce qui concerne la liaison après les adjectifs épithètes, si la probabilité transitionnelle représentait un facteur explicatif capable de distinguer le comportement de la liaison de l'adjectif antéposé de celui de la liaison après adjectif postposé au nom (cf. Côté 2013), nous devrions alors pouvoir observer une différence entre les adjectifs majoritairement antéposés

(ex. « grand », « petit » etc.) de ceux ayant une distribution complémentaire (ex. « ancien » , « prochain » etc)[174]. Cependant, comme nous l'avons vu dans les paragraphes précédents (cf. Paragraphe 2.2.3.3) cette hypothèse n'est par vérifiée.

Pour conclure, ces observations nous suggèrent que si une analyse syntaxique superficielle des catégories morphosyntaxique d'appartenance des mots de liaison est insuffisante, une étude plus large des rapports de dépendances syntaxiques qui gouvernent les mots de liaison permettrait au moins en partie de rendre compte de la diversité de comportements qui caractérisent les types de liaison décrits par les classifications traditionnelles. Cela ne revient pas à refuser une analyse basée sur les probabilités transitionnelles mais simplement à la traiter comme la conséquence de différences syntaxiques plus profondes.

3. Conclusion

À l'issue de ce travail, nous espérons que l'analyse fréquentielle menée sur les corpus PFC ré-annotés a contribué d'une part à rendre compte des idiosyncrasies internes aux contextes morphosyntaxiques de liaison, en dépassant les limites d'une classification basée sur les catégories morphosyntaxiques d'appartenance des mots de liaison. D'autre part, la distribution fréquentielle (distribution de Zipf) de la liaison variable mise en évidence par notre analyse remet davantage en discussion le traitement unitaire de la liaison proposé par les approches prescriptives et par les modèles phonologiques génératifs.

En effet, afin de rendre compte de l'hétérogénéité et de la productivité de la liaison, il nous semble que, comme l'a suggéré Côté (2005), il serait opportun d'éviter un traitement unitaire du statut de la liaison et surtout de la liaison variable.

Comme nous l'avons montré, l'analyse de la liaison variable reçoit une analyse différente selon le contexte. Si la liaison après les conjonctions et les prépositions monosyllabiques semble être dans l'ensemble assez homogène, d'autres, comme la liaison après des adverbes monosyllabiques, demandent une analyse syntaxique des dépendances qui dépassent les limites d'une classification basée sur les catégories morphosyntaxiques d'appartenance. En revanche d'autres types de liaison semblent être conditionnés principalement par des facteurs d'ordre lexical. C'est le cas par exemple de la liaison après des formes verbales où les seules formes « est » et « était » rendent compte de la presque totalité des réalisations, indépendamment du

[174] Voir Benzitoun (2014)

contexte syntaxique dans lequel elles se trouvent (VER-DET, VER-PRP cf. Tableau 10). Enfin, la liaison après les adjectifs épithètes, à laquelle nous avons consacré une étude supplémentaire, doit, comme Morin (1992) l'avait déjà suggéré, recevoir un traitement qui distingue le contexte singulier du contexte pluriel. Comme nous l'avons montré, le contexte singulier est sujet à des contraintes d'ordre distributionnel qui limitent le nombre d'adjectifs susceptibles d'être antéposés à l'oral. Cependant, ni la distribution ni la probabilité transitionnelle n'influencent la productivité de la liaison puisque, comme nous l'avons vu, la liaison n'est pas plus fréquente après des adjectifs majoritairement antéposés qu'après des adjectifs en distribution complémentaire. De plus, des adjectifs tels que « gros » et « long » majoritairement antéposés se trouvent rarement en contexte de liaison (ce qui suggère une mise en place de stratégies d'évitement de la part du locuteur) (Morin 1992 ; Durand et Lyche 2008 ; Barreca et al. 2015). Si le facteur lexical ne semble pas pouvoir expliquer la variabilité de la liaison après les adjectifs épithètes singuliers (les adjectifs les plus fréquents « grand » et « petit » présentent le taux le plus important de non-réalisations), le nombre limité d'adjectifs ainsi que les stratégies d'évitement observées confirment cependant une lexicalisation sous forme de constructions de ces contextes de liaison. En revanche, la liaison après les adjectifs pluriels où la CL est grammaticalisée, laisse la place à plusieurs traitements qui considèrent la CL comme autonome par rapport au M1 et M2 (épenthétique, morphémique et exemplariste) (cf. Partie 1.Chapitre 5.2).

L'ensemble de ces considérations nous amène donc à nous demander si un traitement unitaire de la liaison variable basé sur des catégories morphosyntaxiques ne pouvait pas être complété par une étude plus fine des constructions (« dans - z - DET », « est - t - DET » , « était - t - PRP », « grand - t - NOM » etc.) qui émergent de cette analyse fréquentielle (cf. Tableau 18). En effet, indépendamment de leur type d'appartenance ce petit nombre de constructions rend compte de 64% des réalisations totales de liaison variable (cf. Paragraphe 2.2.2.8). Les occurrences restantes de liaison, comme nous l'avons vu dans les pages précédentes, relèvent d'un nombre important de types différents où, selon l'hypothèse formulée par Côté (2015), les mots de liaison sont caractérisés par des probabilités transitionnelles moins importantes. À la suite de Laks (2005b), nous pouvons supposer que la production de ces liaisons variables est davantage tributaire de l'influence des représentations orthographiques de mots stockées dans le lexique mental. Dans le cadre du modèle interactif bimodal (cf. Partie 1.Chapitre 6.3.4), nous proposons de considérer cette influence comme un facteur concurrent par rapport à l'émergence des constructions de liaison variable.

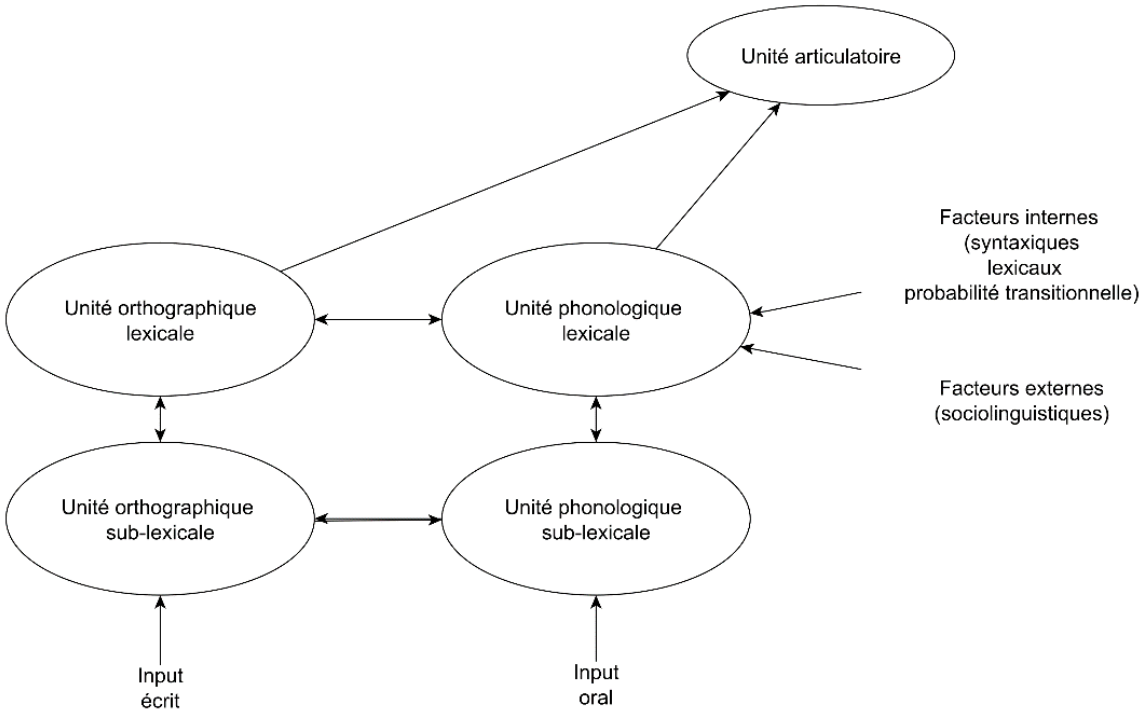


Figure 26 : Interactions entre représentations phonologiques et orthographiques dans la production de la liaison (Grainger et Ferrand 1996)

Ainsi la fréquence la plus importante de certaines occurrences de liaisons attestées par la présente étude pourrait s'expliquer par une plus grande accessibilité de ces constructions favorisée par l'ensemble des facteurs lexicaux, syntaxiques, distributionnels et des probabilités transitionnelles que nous avons évoqués dans les pages précédentes. En revanche, dans le cas des occurrences de liaison moins fréquentes, se situant au milieu et à la fin de la courbe distributionnelle mise en évidence dans cette analyse (cf. Figure 18), la production s'appuierait davantage sur des représentations orthographiques des mots stockées dans le lexique mental du locuteur.

Ces conclusions, bien évidement partielles, représentent pourtant un enjeu important pour l'acquisition et la didactique du français langue étrangère auxquelles seront consacrées les prochaines sections de ce travail. Les résultats issus de l'analyse fréquentielle du corpus PFC pourront ainsi servir de modèle de référence pour l'étude des productions des apprenants de français langue étrangère.

Partie 3. ENSEIGNEMENT ET ACQUISITION DE LA LIAISON EN FRANÇAIS LANGUE SECONDE ET ETRANGERE

Chapitre 1. Repères théoriques sur l'acquisition d'une langue seconde et étrangère

La première partie de ce travail nous a permis de retracer les origines historiques de la liaison, de la situer par rapport aux autres phénomènes de *sandhi* attestés en français. Nous avons par la suite présenté l'ensemble des variables d'analyse, les conditionnements auxquels la liaison est soumise ainsi que les descriptions et les modèles principaux (à la fois linguistiques et psycholinguistiques) qui ont été proposés afin de rendre compte de ce phénomène. Au terme de cette brève introduction de la liaison trois aspects problématiques semblent émerger. Premièrement, la modélisation de ce phénomène ainsi que le statut de la CL sont encore aujourd'hui, et ce depuis presque un siècle, au centre des débats. De plus, la complexité de formalisation de la liaison semble également se traduire dans une difficulté à représenter, à l'aide des classements traditionnels (principalement basés sur des critères syntaxiques), la pluralité des facteurs qui influencent la réalisation de la liaison. Enfin, cette variabilité, (en particulier diaphasique et diastratique), qui fait de la liaison un fort indicateur social, conduit également à s'interroger sur la notion de norme de prononciation dans la description de la liaison.

Comme nous le verrons dans cette troisième partie le débat autour de ces trois sujets sera enrichi par les apports des travaux menés sur la liaison dans le domaine de l'acquisition et de l'enseignement du français langue étrangère. Une telle analyse exige pourtant une introduction, même brève, des repères théoriques sur l'acquisition d'une langue seconde ou étrangère qui nous permettront de mieux situer la description de l'apprentissage de la liaison dans le cadre du processus d'acquisition du français langue étrangère.

1. Introduction

La linguistique acquisitionnelle se propose de décrire l'apprentissage d'une langue seconde (dorénavant L2)[175], en essayant de rendre compte de la variation et du fonctionnement du processus acquisitionnel à travers l'élaboration de modèles descriptifs et l'étude des facteurs qui le conditionnent. Parmi ceux-ci, à la suite d'Ellis (1994), nous pouvons opérer une distinction entre les facteurs linguistiques (ou internes) (influence de la L1, facteurs cognitifs,

[175] Dans cette troisième partie, nous utiliserons le terme « langue seconde » (L2) pour désigner de manière générale l'acquisition d'une langue autre que celle maternelle sans préciser pour l'instant le type de situation d'apprentissage (cf. Partie 3.Chapitre 1.3.2.2).

universaux linguistiques et saillance) et extralinguistiques (ou externes). Ces derniers peuvent être individuels (motivation, attitude, personnalité, style cognitif et stratégies d'apprentissage), sociaux (âge, sexe, niveau social et contexte d'apprentissage) ou liés au type et à la fréquence de l'input. Malgré ces distinctions, Ellis (1994) rappelle que ces variables interagissent entre elles.

> « The categorization of explanations in terms of external factors, internal factors, and individual learner differences is justified because it reflects identifiable orientations in SLA *[Second Language Acquisition]* research and because it provides a means of organizing the information provided by research. However, it should be emphasized that there is no single and no simple explanation of L2 acquisition. A complete explanation will need to consider all the three components and how they interact ». (Ellis 1994, p. 194)

Sans vouloir parcourir l'évolution des différentes théories d'apprentissage des L2 (Bouton 1974 ; Klein 1989 ; Gaonac'h 1991) ni analyser l'ensemble des variables influençant l'apprentissage d'une L2, nous proposons dans ce chapitre d'examiner quelques aspects théoriques de l'acquisition d'une L2 qui permettront de mieux analyser le processus d'appropriation de la liaison.

Nous nous attacherons tout d'abord à décrire le système linguistique de l'apprenant et ses caractéristiques. Cela nous amènera par la suite à rechercher, parmi les facteurs internes et externes, des éléments capables de rendre compte de l'évolution de l'interlangue. L'analyse de l'interlangue fera l'objet de la dernière partie de ce chapitre dans laquelle nous décrirons les approches proposées pour l'étude de ce système intermédiaire ainsi que les parcours acquisitionnels et les stratégies (d'acquisition et communication) mises en place par l'apprenant.

2. Le système linguistique des apprenants : l'interlangue

Quel que soit le contexte et le type d'acquisition de L2, l'apprenant doit toujours s'exprimer avec le répertoire linguistique de L2 dont il dispose à un moment donné, que ce soit pour communiquer réellement comme dans l'acquisition ou à travers des simulations ou des jeux de rôle. Bien que ces systèmes soient incomplets par rapport à la langue cible, ceux-ci représentent la variété d'expression à savoir le lecte d'apprenant ou interlangue. Celle-ci se développe à mesure que l'apprenant est confronté à des éléments de la langue cible. Plusieurs dénominations ont été proposées selon les auteurs : système approximatif, compétence transitoire, système intermédiaire etc.

> « Ce que l'on a dénommé, ici ou là, système approximatif (Nemser, 1971), compétence transitoire (Corder, 1967), dialecte idiosyncrasique (Corder, 1971), système intermédiaire (Porquier, 1974), interlangue (Selinker, 1972), système approximatif de communication, langue de l'apprenant ou système approché (Noyau, 1976) […] recouvre, malgré certaines dispersions théoriques ou méthodologiques, un même objet […]. C'est ce que nous avons précédemment appelé grammaire intériorisée par l'apprenant, et que nous appellerons également ici interlangue ». (Besse et Porquier 1991, p. 216)

Quelle que soit la terminologie adoptée, les différentes études concordent quant à la nature progressive du développement de l'interlangue qui résulte à la fois orientée vers la langue cible et dépendante de la langue de départ. Malgré cette instabilité, d'après Corder (1981) les interlangues, tout comme les langues naturelles, sont régies par des règles et peuvent donc être décrites d'un point de vue linguistique. Cependant, ces règles ne renvoient pas forcement à celles de la L1 ou de la langue cible, mais à un système linguistique à part. Cette autonomie et le caractère systématique (issu de la présence de règles) s'accompagnent d'une forte variabilité, qui reflète l'évolution de l'interlangue et qui se manifeste à plusieurs niveaux. D'un point de vue longitudinal, l'interlangue évolue de façon continue passant par des états successifs. Chacun d'entre eux se construit sur un état antérieur en le modifiant. Pendant ces phases les connaissances de l'apprenant ne sont pas encore suffisamment consolidées, ce qui explique cette forte variation. De plus, comme le souligne Py (1980), les constituants de l'interlangue (qu'il considère comme un ensemble de « microsystèmes »)[176] sont caractérises par un certain degré d' « autonomie » et de « souplesse » qui permettent de rendre compte des « distorsions », des « transgressions » et des « interférences ». Ces dernières ne dépendent pas que du hasard mais au contraire sont à l'origine même de l'évolution de l'interlangue (Py 1980, p. 50).

Toutefois, l'interlangue ne varie pas seulement dans le temps mais aussi en fonction de l'apprenant et des situations de production linguistique (communication formelle / informelle; entre apprenants / avec locuteurs natifs ; production libre / guidée etc.). Quant à la variabilité intra et inter-individuelle, elle dépend de plusieurs facteurs liés à la personnalité de l'apprenant, à ses expériences d'apprentissage antérieures ainsi qu'au degré de présence de sa L1 ou d'autres langues étrangères.

Cette dualité qui oppose systématicité et présence de règles d'une part et variabilité et instabilité d'autre part rend complexe la tâche de description de l'interlangue. Néanmoins, les linguistes ont tenté de décrire des principes généralisables au-delà des variations intra et inter-

[176] En effet, d'après Py (1980), le recours à la notion de « microsystème » qu'il emprunte à Gentilhomme (1979 cité par Py 1980) « permet de concevoir la langue comme équilibre instable. Au lieu de la définir comme système cohérent d'unités dépendantes, on l'envisage comme un ensemble formé d'un nombre indéfini de microsystèmes. Leur coexistence implique solidarité de fait, mais elle laisse assez de jeu pour que chacun jouisse d'une relative autonomie » (Py 1980, p. 45).

individuelles. En particulier, Selinker (1972) définit des processus cognitifs principaux gouvernant l'évolution de l'interlangue : (a) le transfert de la L1 (d'unités ou de règles) à savoir l'influence linguistique relevant de la L1 des apprenants (cf. Paragraphe 3.1.1) ; (b) le *transfert-of-training* (« transfert par entraînement ») c'est-à-dire l'influence positive ou négative de l'application des règles apprises par l'apprenant ; (c) les stratégies d'apprentissage de la L2 (cf. Paragraphe 4.3); (d) les stratégies de communication en L2 visant au transfert du message plus qu'à l'acquisition (cf. Paragraphe 4.3.2) ; (e) la surgénéralisation des règles de L2 apprises par les apprenants.

> « I consider the following to be processes central to second-language learning: first, language transfer; second, transfer-of-training; third, strategies of second-language learning; fourth, strategies of second-language communication; and fifth, over generalization of TL *[L1]* linguistic material. Each of the analyst's predictions as to the shape of IL *[interlangue]* utterances should be associated with one or more of these, or other, processes ». (Selinker 1972, p. 215)

Corder (1981) recherche également des régularités dans l'évolution de l'interlangue et parvient à la conclusion que les processus d'apprentissage en L2 présentent des caractéristiques communes qui doivent renvoyer à des similarités au niveau de la grammaire de l'interlangue.

> « The concept of interlanguage presupposes that interesting and important generalizations can be made about the process of second language learning, specifically that all language learners do show similarities in their acquisition of a second language and consequently in their interlanguage grammars at various points along the continuum change or development ». (Corder 1978, p. 76)

Pour résumer, l'interlangue doit s'analyser comme un ensemble relativement autonome par rapport à la L1 et à la langue cible, régie par ses propres règles qui sont l'objet d'une évolution continue au cours du processus acquisitionnel. Dans ce contexte, les formes de l'interlangue des apprenants ne sont pas traitées comme des déviations par rapport à la norme, mais, en revanche, sont rapportées à la fonction et à la place qu'elles jouent dans le processus d'acquisition. Une telle analyse permet, comme l'observe Vogel (1995), de comprendre et d'expliquer les phénomènes et les stratégies d'évolution et de planification de la langue qui déterminent l'acquisition des L2 en vue d'élaboration d'aides destinées à améliorer la performance de ce processus d'apprentissage (Porquier 1974 ; Pienemann 1984 ; Pienemann 1985).

Bien que l'interlangue, comme nous le verrons dans les prochains paragraphes, soit influencée par plusieurs facteurs internes (cf. Paragraphe 3.1) et externes (cf. Paragraphe 3.2), Py (1980) rappelle que l'étude de la langue de l'apprenant doit tout d'abord rechercher « les propriétés intrinsèques » capables d'expliquer son évolution.

> « Expliquer le changement, ce n'est pas seulement énumerer les facteurs extérieurs qui le conditionnent, c'est aussi et surtout montrer que le système linguistique impliqué comporte des caractéristiques qui rendent possible l'évolution. En ce qui concerne l'interlangue, cette thèse signifie qu'on ne peut pas la caractériser de manière tant soit peu réaliste sans prendre en considération les propriétés instrisèques qui constituent justement l'apprentissage ». (Py 1980, p. 44)

3. Influence des facteurs internes et externes sur le développement de l'interlangue

3.1. Des facteurs internes

3.1.1. Le rôle de la L1

Parmi les facteurs cognitifs qui conditionnent l'acquisition d'une L2, l'influence de la L1 est probablement la plus évoquée et constitue l'une des variables ayant fait l'objet des premières études sur l'apprentissage des L2. Les premières approches comportementalistes et contrastives considéraient l'interférence de la L1 comme le principal obstacle (*proactive inihibition*) à l'acquisition de la L2. Celui-ci émergeait lorsque l'apprenant employait les habitudes (*habits*) linguistiques de la L1 pour apprendre celles de la L2. La production de la plupart des erreurs était attribuée au recours à la L1 à savoir à l'interférence ou transfert négatif qui avait lieu lorsque les formes de la L1 et de la L2 divergeaient. Dans le cas contraire, le transfert positif permettait à l'apprenant de réaliser des formes correctes. Ces positions seront critiquées et jugées trop mécaniques. Le poids de l'influence de la L1 a été revu ainsi que l'extension du terme « transfert ». Certains auteurs (Sharwood Smith et Kellerman 1986) préfèrent parler d'influence interlinguistique (*crosslinguistic influence*) ou de « marques transcodiques » (Lüdi et Py 1986) ce qui permet de faire référence aux conditionnements de tout système linguistique précédemment acquis par l'apprenant.

> « […] the term 'crosslinguistic influence' […] allowing one to resume under one heading such phenomena as 'transfer' , 'interference', 'avoidance', 'borrowing' and L2-related aspects of language loss and thus permitting discussion of the similarities and differences between these phenomena ». (Sharwood Smith et Kellerman 1986, p. 1)

Dans cette perspective, Andersen propose de considérer l'influence de la L1 comme un filtre qui conditionne la perception de certains traits de la L2. Plusieurs recherches (Dulay, Burt et Krashen 1982 ; Odlin 1989 ; Ellis 1994) ont essayé de rendre compte des facteurs en jeu dans l'influence interlinguistique : (a) niveau linguistique (l'influence du conditionnement serait plus

importante en phonologie[177] (cf. Paragraphe 3.1.1.1) que dans le lexique, la grammaire et en dernier lieu le discours) ; (b) la saillance (les éléments moins marqués seraient plus facilement l'objet d'un transfert) ; (c) la prototypicité (les formes ayant des significations et des usages les plus prototypiques sont plus facilement transférées) ; (d) la distance typologique ou psycho-typologique à savoir la perception que l'apprenant possède des différences entre la L1 et la L2 (les formes perçues comme moins distantes sont plus sujettes à transfert) ; (e) la conformité aux principes de la naturalité qui favorise une correspondance entre forme et fonction déterminant la production d'une erreur.

3.1.1.1 Le transfert dans la phonologie L2

Comme nous l'avons déjà évoqué le transfert de la L1 semblerait exercer une influence plus importante dans l'acquisition de la phonologie de la L2. Major (2008) propose un parcours historique des études menées sur le transfert phonologique en L2, concept qui sera évoqué par Weinreich (1953) dans le cadre des études des « langues en contact » et par la suite sera au centre de l'analyse contrastive à partir des années 1960.

Dans ces recherches sur le bilinguisme, Weinreich (1953) distingue plusieurs typologies d' « interférences » (*interference*). Parmi les transferts phonologiques les plus fréquents nous pouvons citer : (1) la substitution de phones (*sound substitution*) (ex. l'apprenant anglophone utilise le [ɹ] anglais au lieu du [r] espagnol dans l'acquisition de l'espagnol L2)[178] ; (2) la réinterpretation de phonèmes (*reinterpretation of distinctions*) (ex. l'apprenant recombine des caractéristiques de L2 sur la base d'oppositions prèsentes dans la L1, ex. les apprenants germanophones d'anglais L2 interprètent l'opposition entre les voyelles tendues et voyelles relâchées comme une distinction entre des voyelles longues et des voyelles brèves) ; (3) la sous-différenciation de phonèmes (*underdifferentiation*) (cette intérférence a lieu lorsque la langue cible possède des phonèmes qui dans la L1 de l'apprenant n'ont pas de valeur distinctive, ex. [d] et [ð] dans le cas d'apprenants hispanophones d'anglais L2) ; (4) la sur-différenciation des phonèmes (*over-differentiation of phonemes*) (cette intérférence a lieu lorsque la L1 possède des phonèmes qui en L2 n'ont pas de valeur distinctive, ex. [d] et [ð] dans le cas d'apprenants anglophones d'espagnol L2)[179].

[177] « They [*les apprenants*] also make use of their L1, gradually restructuring it as they discover how it differs from the TL [*L2*], a process that Corder suggests is more prevalent in the case of phonology learnt in formal environments ». (Ellis 1994, p. 115)

[178] Les exemples sont tirés de Major (2008, p. 67).

[179] L'étude de Flege (1987) menée sur des apprenants d'anglais et de français L2 a permis de montrer que l'interférence entre la L1 et la L2 est bi-directionnelle, dans la mesure où les apprenants L2 sont capables d'une part d'apprendre à prononcer de nouveaux phones L2 (*new L2 phones*) (des phones de L2 qui par définition sont différents et qui ne se rapprochent d'aucun phone appartenant à la L1, ex. /y/ pour des apprenants anglophones de

Ces catégories seront par la suite ré-élaborées par l'analyse contrastive des années 60, qui sous l'influence de l'école behavioriste considère les erreurs comme le résultat de l'interférence des habits linguistique de la L1 dans la production de la langue cible. Cette approche essayait de prédire les erreurs des apprenants en identifiant les différences entre leurs L1 et la langue cible (Lado 1957). Cette méthologie a fait l'objet de nombreuses critiques qui montraient que les apprenants ne réalisaient pas les erreurs prédites par l'analyse contrastive (ex. certains apprenants germanophones d'anglais L2 ne rencontraient pas de difficulté dans l'acquisition du /r/ anglais). À la suite de ces objections, plusieurs courants plus ou moins modérés ont été développés au sein de l'analyse contrastive (Wardlhaugh cité par Major 2008). Si d'une part, la version forte (*strong version*) (Lado 1957) défend encore le pouvoir prédictif de cette approche, d'autres positions comme celle adoptée par Oller et Ziahosseiny (1970) (*moderate version*) considèrent que le transfert a lieu le plus souvent lorsque les systèmes de la L1 et de la L2 se ressemblent. Oller et Ziahosseiny (1970) ont mené une analyse des erreurs de prononciation réalisées par des apprenants d'anglais L2 répartis en deux groupes selon qu'ils possédaient ou pas une L1 basée sur un système d'écriture alphabétique. Les résultats de cette recherche ont montré que les apprenants ayant des langues (romanes, germaniques) réalisent plus d'erreurs de prononciation que des sujets sinophones, japonophones ou hébreux ayant un système d'écriture non alphabétique. Les auteurs en concluent que la connaissance de l'écriture alphabétique rend plus complexe l'acquisition d'une autre langue issue du même système d'écriture (Oller et Ziahosseiny 1970, p. 188).

En dépit de ces positions plus modérées, l'analyse contrastive continue d'être accusée de négliger l'étude des tous les autres facteurs internes et externes qui conditionnent les productions de l'apprenant. Ces réflexions aboutiront par la suite au développement d'une nouvelle interprétation des erreurs (cf. Paragraphe 4.1) des apprenants, et plus en général de la notion même d'interlangue (cf. Paragraphe 2).

3.1.2. La saillance linguistique

Comme nous l'avons vu dans le paragraphe précédent, des aspects typologiques sont évoqués parmi les facteurs influençant l'acquisition d'une L2. Une approche typologique d'étude de l'acquisition d'une L2 exige une comparaison interlinguistique des traits spécifiques afin d'établir lesquels sont rares et lesquels sont communs dans une langue particulière. Un

français L2), et d'autre part de modifier les modèles de prononciation de L1 sur la base des phones similaires de la L2 (*similar L2 phones*) (à savoir des phones de la L2 qui se rapprochent d'autres appartenant à la L1 tout en étant différents). Ceux-ci seraient classifiés sous une même catégorie (*equivalence classification*), ce qui expliquerait les déficits dans la production et perception d'un phone similaire de L2.

certain nombre de recherches (Eckman 1977 ; Eckman 1981 ; Eckman 1984 ; Rutherford 1984 ; Tarone 1987; Eckman 2008) basées sur la typologie phonologique ont montré que les apprenants acquièrent plus facilement les traits phonologiques non marqués que ceux qui sont marqués. Selon cette hypothèse, des apprenants ayant deux L1 différentes ne rencontrent pas le même niveau de difficultés dans l'acquisition de la même langue cible. Celles-ci varient en fonction des différences de saillance linguistique existant entre les langues sources et la L2 (*directionality of difficulty*) (Eckman 2008, p. 98).

Cependant d'après certains auteurs (Gass 1984 cité par Ellis 1994 ; Rutherford 1984) ce facteur joue un rôle marginal dans l'acquisition d'une L2 et ne paraît offrir que des indices pour des observations et des prédictions.

3.2. Des facteurs externes

Afin de parvenir à une définition plus précise des formes d'acquisition d'une L2, il nous paraît important de présenter brièvement quelques facteurs externes conditionnant l'acquisition d'une L2. Parmi l'ensemble des variables externes jouant un rôle dans l'apprentissage d'une L2 (Ellis 1994 ; Hansen Edwards 2008), nous nous intéresserons tout particulièrement à l'âge, au contexte d'apprentissage ainsi qu'à quelques caractéristiques de l'input linguistique auquel les apprenants sont exposés (notamment l'influence de la fréquence de l'input et de l'input écrit).

3.2.1. L'âge

Il a été souvent observé que les apprenants les plus jeunes apprennent plus facilement que les plus âgés. Cette différence est expliquée en faisant référence à l'hypothèse de la période critique (Lenneberg, Chomsky et Marx 1967) selon laquelle il existerait une période de temps (avant 12 ans) pendant laquelle la plasticité neuronale du cerveau permettrait une meilleure acquisition de la L2. Au-delà de cette période le système nerveux serait moins réactif (à cause du phénomène de la latéralisation) aux stimulations de l'environnement ce qui rendrait plus difficile le processus l'apprentissage. Cette hypothèse a été revue par des études successives (Krashen et al. 1976 ; Jacobs 1988 ; Singleton et Ryan 2004 ; De Keyser 2000; Ioup 2008). Si certains ont proposé d'anticiper la fin de la période critique, d'autres (Singleton et Ryan 2004) défendent l'existence de plusieurs périodes sensibles, une pour chaque niveau linguistique. Les périodes de la phonétique et de la phonologie seraient plus courtes que celles de la morphologie et de la syntaxe[180].

[180] En revanche, la sémantique, la pragmatique et le lexique ne poseraient pas d'importantes difficultés d'acquisition même à l'âge adulte.

Comme le souligne Galazzi (2008a, p. 116-117) ces obstacles neurobiologiques qui empêchent l'acquisition de la prononciation native à l'âge adulte ne sont pas sans lien avec un blocage psychologique tel que la « surdité phonologique » (Troubetzkoy 1949) qui empêche de percevoir et de produire tout son qui n'a pas de fonction distinctive dans le système phonologique de sa propre langue » (Galazzi 2008a, p. 117). D'après Galazzi (2008a) la notion de période critique représente une « reformulation actualisée » du concept de « crible phonologique »[181] élaboré par Troubetzkoy (1949) et donc en rapport direct avec le rôle de filtre joué par la L1 des apprenants (cf. Paragraphe 3.1.1).

Ioup (2008) résume bien l'opposition entre ces deux différentes hypothèses dans le débat concernant l'influence exercée par l'âge sur l'acquisition de la phonologie en L2.

> There are several schools of thought on this; two of the major ones will be reviewed here. One argues that as it matures, the brain undergoes biological changes that make it impossible for the learner to perceive and produce novel sounds. The other contends that the categories developed for the L1 interfere with the perception of new categories in the L2 ». (Ioup 2008, p. 48)

3.2.2. Le contexte d'acquisition d'une L2 : langue seconde et langue étrangère

Dans la linguistique acquisitionnelle on désigne de façon générale sous le terme d'acquisition d'une langue seconde une langue non maternelle (dorénavant L1) apprise après la L1 ou après l'âge de trois ans[182] qui n'est ni la langue de première socialisation ni la langue première dans l'ordre des appropriations linguistiques.

Ensuite, à partir de la deuxième moitié du XXème siècle, la nécessité de décrire les différents types de situations d'appropriation d'une L2[183] conduira à établir une opposition entre l'acquisition d'une langue seconde (LS) et d'une langue étrangère (LE).

En France le terme de français langue seconde (FLS) désignera notamment dans un premier temps le français parlé à l'étranger avec un statut particulier à savoir l'usage du français dans les régions francophones hors de la France métropolitaine. Dans ce contexte le français, bien qu'il ne soit pas la L1 de la plupart de la population, ne peut pas être considéré comme

[181] Le concept de « crible phonologique » renvoie à une situation d'acquisition où l'apprenant d'une L2 perçoit les sons de la langue cible à travers le filtre phonologique de sa L1. Autrement dit, il ne parvient pas à distinguer deux sons qui dans sa L1 n'ont pas de fonction distinctive (ex. l'opposition entre /y/ et /u/ en français constitue une difficulté pour les apprenants pour lesquels ces phonèmes n'ont pas de fonction distinctive dans le système phonologique de leur L1 comme dans le cas de l'italien standard).

[182] Cette opposition ne relève pas seulement de l'ordre temporel et intéresse des aspects du développement cognitif et social qui sont propres de l'acquisition de la L1.

[183] Nous continuerons à utiliser ce terme pour désigner de manière générale l'acquisition d'une langue autre que celle maternelle dans les cas où la définition des types de situation d'appropriation ne s'avèrerait pas nécessaire.

langue étrangère. Dans un premier temps, le terme de FLS était donc employé pour définir une langue ayant un statut privilégié par rapport aux autres langues étrangères présentes dans le même territoire.

> « Il (le FLS) se distingue des autres langues étrangères, éventuellement présentes sur ces aires par ses valeurs statutaires, soit juridiquement, soit socialement, soit les deux et par le degré d'approbation que la communauté qui l'utilise s'est octroyé ou revendique. Cette communauté est bi ou plurilingue. La plupart de ses membres le sont aussi et le français joue, dans leur développement psychologique, cognitif et informatif, […] un rôle privilégié ». (Cuq 1991, p. 96)

Ainsi, on voit apparaître deux terminologies visant à qualifier les contextes d'apprentissage: « homoglotte » et « alloglotte » (Dabene 1990). Le premier se référant à une situation d'apprentissage et d'enseignement où la langue cible est une langue de première socialisation ou elle « est présente dans l'environnement communicationnel des locuteurs et/ou elle a une fonction dans la construction des connaissances et des apprentissages (FLS) » (Coste 2007, p. 215). En revanche, le deuxième qualifie les autres contextes où la langue cible représente une langue étrangère (français langue étrangère) (dorénavant FLE), apprise en général en classe de langue, et qui n'est pas utilisée en concurrence avec la L1 pour les communications quotidiennes. Cependant, l'augmentation du flux de migrants (surtout d'enfants) dans les pays où le français représente la L1 (français langue maternelle ou FLM) a entraîné une évolution de cette notion. Dans un premier temps le français des (élèves) migrants (défini français de scolarisation ou FLSco) est exclu du FLS (Cuq 1991, p. 140) étant donné que « la fin prévisible » du processus d'acquisition est l'assimilation et le monolinguisme en français. Cette position sera ensuite nuancée (Vigner 2009) en faisant converger les problématiques d'enseignement et d'acquisition des deux catégories FLS et FLSsco[184]. Ce rapprochement découle de la constatation que l'usage dans l'éducation d'un français L2, était bien plus ancien, en France comme dans les régions francophones, que le terme même de FLS.

Malgré la pluralité de dénominations qui semblerait opposer différentes réalités d'apprentissage / enseignement, les frontières entre ces catégories restent floues. Le modèle par continuum proposé par Verdelhan- Bourgade (2007) s'avère donc approprié pour représenter l'axe des situations d'apprentissage dont le FLM et le FLE constituent les deux extrêmes, entre lesquels se situe le FLS (bien qu'il soit plus près du FLE). Une telle approche favorise une description des situations d'apprentissage où l'appartenance à une catégorie est définie sur la

[184] « Le français langue seconde, c'est d'abord une solution destinée à assurer la scolarisation d'élèves non francophones natifs. Cette problématique rejoindra progressivement celle de la scolarisation, en France ». (Vigner 2009, p. 5)

base d'un certain nombre de critères (tels que le lieu d'apprentissage, le statut psycholinguistique et sociolinguistique de la L2), ce qui permet de nuancer les frontières entre les catégories. Dans cette perspective, Goï et Huver (2012) invitent à considérer la diversité des contextes d'acquisition non plus à partir des classes fixes dans la mesure où « les catégorisations ne sont plus tant une question de définition que d'argumentation, qui doit intégrer les représentations et les positionnements (divers, évolutifs et situés) des acteurs » (Goï et Huver 2012, p. 33).

3.2.3. Modalité d'acquisition : acquisition guidée et non guidée

Une grande importance est accordée à la modalité d'acquisition selon qu'elle est guidée ou non par l'enseignement. On définit l'acquisition non guidée comme une acquisition qui se réalise à travers la communication quotidienne sans efforts systématiques qui guident le processus acquisitionnel.

> « On désigne ainsi [*acquisition non guidée*] l'acquisition d'une langue étrangère par la communication quotidienne, acquisition qui se développe naturellement, et sans intervention systématique pour guider le processus ». (Klein 1989, p. 29)

La communication quotidienne demande à l'apprenant de mettre en place deux tâches : l'une communicative visant à utiliser le répertoire linguistique dont l'apprenant dispose pour communiquer et l'autre acquisitionnelle qui requiert une adaptation progressive de ce répertoire à la langue cible. Comme le souligne Klein(1989) ces deux aspects, bien qu'ils soient étroitement liés, constituent deux facteurs distincts du processus d'acquisition qui, comme nous le verrons (cf. Paragraphe 4.3), impliquent des stratégies différentes (d'acquisition et de communication).

> « La communication est un facteur stabilisateur : le développement d'un système d'apprenant et son utilisation optimale facilitent la communication ; l'apprenant peut en quelque sorte se sentir chez lui dans la langue et s'y mettre à l'aise. La tâche d'apprendre, en d'autres termes la nécessité de dépasser ces acquis, de les améliorer et de les réorganiser, est un facteur dynamique : il fait avancer le processus d'acquisition ». (Klein 1989, p. 30)

Dans le cadre d'une communication quotidienne l'accomplissement de la tâche communicative et son succès (et non pas la précision formelle de la langue) constituent les objectifs principaux.

Un deuxième aspect qui permet de distinguer les deux types d'apprentissage concerne l'absence ou la présence d'un guidage qui influence de façon volontaire et systématique le processus d'acquisition. Cependant il est évident que même dans un type d'apprentissage non guidé l'apprenant peut être exposé à des corrections ou des explications. D'autres terminologies

ont donc été proposées pour distinguer ces deux contextes d'acquisition : « acquisition en milieu naturel » et « acquisition en milieu institutionnel » (Noyau 1980) ; « acquisition formelle » et « acquisition informelle » (Brougère et Bezille 2007).

Cependant, encore une fois, les frontières entre ces deux catégories ne sont pas aussi nettes et elles ne justifient pas non plus le recours à une distinction entre les deux processus[185]. Comme le fait observer Noyau (1980), les situations d'acquisition sont souvent très hétérogènes et possèdent des caractéristiques qui les éloignent des « cas purs » prototypiques.

> « Lorsque nous distinguons entre acquisition guidée et acquisition non-guidée, nous ne constituons pas deux catégories disjointes. Les cas « purs » ne sont pas les plus fréquents. Pour des apprenants en milieu scolaire, il peut y avoir aussi de l'acquisition non-guidée, s'il existe des contacts en dehors de la classe avec des locuteurs de LC [*langue cible*], avec de la LC par des lectures, par les média, etc., ces contacts pouvant avoir un impact variable sur l'acquisition […]. Dans les cas d'acquisition non-guidée, il peut y avoir un contact plus ou moins fugace avec de l'enseignement (les travailleurs immigrés qui ont suivi quelques cours d'alphabétisation ; un autodidacte qui s'aide de matériel pédagogique commercial). Les cas purs seraient : acquisition guidée : des élèves en classe, en dehors de tout contact avec un milieu de LC, et recevant des données de LC provenant exclusivement du professeur et des supports pédagogiques ; acquisition non-guidée : un adulte (donc non soumis à scolarisation) transplanté dans un milieu de LC ». (Noyau 1980, p. 73)

Quant aux modalités qui influencent systématiquement le processus d'acquisition, elles varient en fonction des méthodes d'enseignement et des matériaux didactiques[186] (Puren 1988 ; Germain 1993 ; Rivenc 2003). De manière générale, nous pouvons souligner que, contrairement au contexte d'acquisition non guidée, dans l'acquisition guidée les données linguistiques sont plus ou moins structurées et préparées pour l'apprenant qui n'est pas contraint (même dans les méthodes les plus proches de l'acquisition naturelle) par la nécessité de se faire comprendre mais plutôt de se rapprocher d'un modèle prédéterminé (cf. Paragraphe Partie 3.Chapitre 3.1).

[185] La distinction entre acquisition « guidée » ou « non-guidée » a conduit Krashen (1981) à postuler l'existence de deux processus différents : d'une part l'apprentissage d'une L2 (*Second Language Learning*) et d'autre part l'acquisition d'une L2 (*Second Language Acquisition*). À la suite de Klein (1989), nous ne reprenons pas cette distinction et nous utiliserons ces termes comme des variantes stylistiques.

[186] Nous nous attacherons à décrire dans les prochains chapitres (cf. Partie 3.Chapitre 3.2) quelques aspects des ressources didactiques destinées en particulier à l'enseignement de la liaison.

3.2.4. Quelques caractéristiques de l'input

3.2.4.1 L'input écrit

La forme orthographique des mots semble jouer un rôle crucial dans l'apprentissage d'une L2[187]. Cette influence est démontrée par plusieurs travaux (Young-Scholten 1998 ; Young-Scholten, Akita et Cross 1999 ; Young-Scholten 2002 ; Bassetti 2007) menées auprès des apprenants L2. Young-Scholten (1998) montre que les apprenants du polonais L2 afin de résoudre la complexité des groupes consonantiques auraient tendance à ajouter un segment épenthétique alors que les enfants polonais adopteraient une stratégie d'évitement, effaçant le groupe consonantique. Young-Scholten (1998) explique cette diversité de traitement par le fait que, contrairement à l'enfant, l'apprenant essaierait de produire toutes les consonnes présentes dans l'input écrit. Néanmoins, l'apprenant semblerait privilégier l'effacement du groupe consonantique quand il est exposé au seul input phonologique (Young-Scholten et al. 1999).

De même, on retrouve chez des apprenants anglophones d'allemand L2, étudiés par Young-Scholten (2002), des prononciations orthographiques. Bien que les consonnes obstruantes finales voisées soient dévoisées en position finale de mot en allemand, les apprenants anglophones d'allemand L2 auraient tendance à les réaliser voisées ([d] au lieu de [t] dans le mot *Bund*) à cause des formes orthographiques correspondantes (Young-Scholten 2002).

Cependant, l'input orthographique peut parfois jouer un rôle positif dans l'apprentissage d'une L2 et renforcer l'opposition phonologique entre phonèmes, favorisant leur catégorisation. En effet, le travail mené par Erdener et Burnham (2005) a montré que l'apprenant produit moins d'erreurs phonologiques s'il acquiert une L2 orthographiquement transparente (ex. espagnol) qu'une langue orthographiquement opaque (ex. irlandais). Cependant, l'étude de Simon et al. (2010) semble indiquer que l'exposition à l'information orthographique ne facilite pas l'acquisition en L2 de l'opposition entre les voyelles du français /u/ (ex. « vous » [vu]) et /y/ (ex. « vu » [vy]) chez des apprenants anglophones.

Ces études montrent que l'input orthographique peut donner lieu à des résultats opposés dans l'acquisition de la L2. Ces données apparemment contradictoires pourraient s'expliquer par une exposition réduite à la langue parlée, qui, combinée à une maîtrise imparfaite du système phonologique de la L2, pourrait ralentir le développement des représentations phonologiques ainsi que la production. Bassetti (2006) illustre bien cette dynamique.

[187] C'est n'est par hasard si dans les premières méthodes structuro-globale audio-visuelles (SGAV) (Rivenc 2003) l'apprenant était exposé uniquement à la langue orale dans les phases initiales de l'acquisition.

> « [...] instructed second language learners are exposed to orthographic representations without mastering the L2 phonology, and much of the input they receive is written rather than spoken. It is therefore possible that the orthographic representations of the second language, besides affecting performance in phonological awareness tasks, could affect L2 pronunciation as well. [...] the specific orthographic representations of second language units (phonemes, words, etc.) affect learners' second language production, in both L2 spelling and pronunciation ». (Bassetti 2006, p. 95)

3.2.4.2 La fréquence de l'input

Plusieurs études se sont attachées à étudier l'influence de la fréquence de l'input sur l'acquisition d'une L2 et en particulier la relation entre celle-ci et le degré de précision de la production (Ellis 1994, p. 271). Cependant, ces travaux sont difficilement comparables et présentent souvent des limites d'ordre méthodologique. Pour cette raison Ellis (1994), contrairement à Larsen-Freeman et Long (1991), se montre prudent quant aux effets d'une éventuelle influence de la fréquence sur l'acquisition d'une L2.

> « Overall there is very little evidence to support the claim that input frequency affects L2 acquisition but there is very little evidence to refute it. Perhaps the safest conclusion is that input frequency serves as *one* of the factors influencing development, often in association with other factors such as L1 transfer and communicative need ». (Ellis 1994, p. 273)

Gass et Selinker (1992) partagent les mêmes perplexités. Quant à Gass et Mackey (2002), ils proposent de considérer la fréquence comme un facteur de l'input pouvant expliquer les avantages apportés par l'interaction dans l'acquisition d'une L2.

> « A consideration of frequency of patterns in the input may be useful for current formulations of questions and explanations about how interaction facilitates L2 learning. As Ellis points out, opportunities for relating form and function are important for developing strong target like associations. Interactional modifications during conversation can present ideal opportunities for relating form and function ». (Gass et Mackey 2002, p. 251)

Cependant, ils soulignent également qu'il existe des cas où la fréquence ne semble jouer aucun rôle dans le développement de l'interlangue de l'apprenant. C'est le cas par exemple de l'apprentissage de la désinence de la troisième personne singulier en anglais. L'acquisition de cet aspect suit un ordre cognitif de développement (cf. Paragraphe 4.3) qui ne tient pas compte de la variable fréquentielle. Néanmoins les auteurs n'excluent pas que ce facteur ne puisse pas intervenir dans un moment successif du processus acquisitionnel.

> « However, the existence of developmental sequences does not necessarily rule out a role for frequency. A number of researchers have argued that, when learners are not at the correct developmental level to make immediate use of input, it may be stored and made available at a later time for processing and use ». (Gass et Mackey 2002, p. 254)

4. Approches de l'interlangue

Depuis la définition de l'interlangue comme système linguistique autonome par rapport à la L2, les études acquisitionnelles ont non seulement essayé d'identifier les facteurs influençant le processus d'apprentissage mais ont surtout recherché les mécanismes cognitifs qui guidaient cet apprentissage. Plusieurs approches ont été proposées pour étudier l'évolution de ce système intermédiaire et retrouver les régularités sur lesquelles il repose. Au début de cette recherche l'analyse des erreurs et des processus psycholinguistiques qui en étaient à l'origine semblait représenter une piste valable. Cependant, une telle approche non seulement présente des problèmes d'ordre méthodologique mais ne permet pas non plus de saisir le caractère dynamique de l'évolution de l'interlangue que nous avons évoqué au début de ce chapitre (cf. Paragraphe 2). En revanche, cet aspect sera mis en avant par des études successives ayant pour objectif la définition des parcours et des stratégies d'acquisition.

4.1. Analyse des erreurs

L'analyse des erreurs (*Error Analysis*) de l'apprenant élaborée par Corder (1974 ; 1981) est une des premières méthodes d'investigation de l'interlangue. Elle a été développée à partir des années 70 et a permis de dépasser les limites des analyses contrastives (cf. Paragraphe 3.1.1.1) qui essayait de prédire les erreurs des apprenants identifiant les différences entre leurs L1 et la langue cible. Contrairement à l'approche constrastive, l'analyse des erreurs suggère de considérer les formes interlinguistiques comme des indices d'une part pour le linguiste et le didacticien (en tant qu'éléments révélateurs du stade de développement de l'interlangue) et d'autre part pour l'apprenant (comme indices pour découvrir les règles de la langue cible). Cette approche propose une nouvelle méthodologie d'étude des erreurs s'articulant en plusieurs étapes : (a) récolte d'exemplaires de l'interlangue de l'apprenant ; (b) identification des erreurs ; (c) description des erreurs ; (d) explication des erreurs ; (e) évaluation des erreurs.

Une fois établie la liste des erreurs (a) (d'un ou plusieurs apprenants de la même L1 et du même niveau)[188], des critères pour l'identification des formes de l'interlangue (b) doivent être fixés. Cette étape comporte le choix d'une norme de référence[189]. Un autre aspect exigeant une attention particulière concerne la distinction entre les erreurs (*errors*) et les fautes (*mistakes*) (Corder 1967). Les premières (*errors*) font référence à des formes erronées qui

[188] Voir Ellis (1994, p. 49) pour la méthodologie de récolte du corpus des erreurs.
[189] Nous reviendrons sur cet aspect dans les chapitres qui suivent (cf. Partie 3.Chapitre 3.1).

résultent d'un manque de connaissances alors que les deuxièmes (*mistakes*) relèvent d'un problème de performance. Corder (1974) suggère de ne traiter que les erreurs dans l'analyse de l'interlangue de l'apprenant. Cependant, comme le fait observer Ellis (1994, p. 51), cette démarche comporte une double difficulté liée d'une part aux problèmes méthodologiques d'identification que cette distinction fait émerger et d'autre part à la variabilité de ces formes. Il est en effet possible que l'apprenant possède des compétences partielles d'une forme de la langue cible et qu'il ne soit donc pas en mesure de produire la forme correcte dans tous les contextes. De plus, au cours de cette phase d'identification, une autre problématique concerne la possibilité d'inclure dans l'analyse les formes inappropriées qui relèvent du niveau stylistique et pragmatique. Cela demande non seulement une prise de position quant à la norme de référence mais aussi au registre de référence.

Pour ce qui est de la phase de description des erreurs (c), elle se réalise à travers une comparaison entre les formes interlinguistiques des apprenants et des reconstructions des énoncés correspondants dans la langue cible. Au cours de cette étape l'analyse se contente d'examiner les aspects superficiels des formes interlinguistiques sans rechercher à les expliquer mais simplement à les décrire à l'aide de taxinomies[190] (ex. absence de liaison, substitution de la CL, ordre syntaxique etc.) en indiquant le taux d'erreur. L'étude interprétative est, en fait, réservée au stade suivant d'explication des erreurs (d) dont l'objectif est d'établir l'origine psycholinguistique des formes interlinguistiques. La plupart des auteurs distinguent entre les erreurs de transfert de la L1 (*transfer errors* ou *interference errors*) et les erreurs intralinguistiques (*intralingual errors*) (Richards 1971 ; Lott 1983 cité par Ellis 1994). Il existe également des formes interlinguistiques pour lesquelles cette distinction n'est pas valable et qui ont donc été classées comme des erreurs inter et intralinguistique (Lott 1983 cité par Ellis 1994) (cf. Paragraphe 4.1.3).

4.1.1. Erreurs interlinguistiques : le transfert interlangue

Ces formes relèvent de l'emploi d'éléments et de structures de la L1 dans les productions de la langue cible. Elles peuvent être distinguées en deux types (Lott 1983 cité par Ellis 1994) :

[190] Cependant, cette phase n'est pas facile à réaliser puisque, comme le fait observer Ellis (1994), il est souvent problématique d'identifier la forme correspondante de l'erreur dans la langue cible et par conséquent de pouvoir classer la forme de l'interlangue au sein d'une taxinomie.

« The description of errors, like their identification, is problematic. Even if the error itself can be easily identified, it is often problematic to determine what the error consists of. […] For exemple if learner produces the following sentence: **I am worried in my mind* it is not clear what constitutes the best reconstruction. One possibility is *I am feeling worried*. Another is *I have a problem on my mind* ». (Ellis 1994, p. 57)

- surgénéralisation par analogie (*overextension of analogy*) résultant de l'usage impropre d'un item ayant des traits communs avec un autre élément appartenant à la L1 (ex. l'emploi du mot *process* à la place de *trial* chez des apprenants italophones pour lesquels le mot italien *processo* possède cette signification) ;
- transfert de structures (*transfer of structure*) apparaissant lorsque l'apprenant utilise des traits (phonologiques, lexicaux, grammaticaux ou pragmatiques) de sa L1 au lieu de ceux de la langue cible. Ce type de formes correspond aux erreurs traditionnelles de transfert.

4.1.2. Erreurs intralinguistiques

Les erreurs intralinguistiques étudié par Richards (1971) concernent les caractéristiques générales des processus d'apprentissage telles que :

- fausses généralisations (*over-generalizations*) : elles apparaissent quand l'apprenant crée des formes incorrectes sur le modèle d'autres structures existant dans la langue cible ;
- non connaissance des restrictions des règles (*ignorance of rules restrictions*) : ce type d'erreur donne lieu à une surapplication des règles dans des contextes où elles ne devraient pas être appliquées;
- application partielle des règles (*incomplete application of rules*) : cette erreur comporte la production de formes incomplètes liées donc à des compétences en cours de formation (*error of transitional competence*) (ex. **You like to sing* ? au lieu de la forme complète de l'interrogative *Do you like to sing ?*) ;
- fausses hypothèses (*building of false systems or concepts*) : ces formes émergent lorsque l'apprenant ne comprend pas entièrement une différence existant dans la langue cible (ex. l'emploi de *was* comme marque du passé dans l'énoncé **One day it was happened*)[191].

4.1.3. Erreurs inter et intralinguistiques

Des erreurs inter et intralinguistiques (*interlingual / intralingual errors*) émergent quand une distinction particulière de la langue cible ne possède pas de correspondance dans la L1 de l'apprenant (ex. l'emploi de *make* à la place de *do* chez des apprenants italophones pour lesquels cette différence n'existe pas dans la langue maternelle).

[191] Ces erreurs sont décrites par Richards (1971) comme des erreurs de développement.

4.2. Limites

Malgré ces distinctions, comme le fait observer Flick (1979 cité par Ellis 1994), non seulement il existe des cas ambigus mais la classification d'une erreur peut être le résultat d'une attribution arbitraire basée sur une interprétation subjective du chercheur. Les limites de l'analyse des erreurs mises en évidence par Ellis (1994) concernent donc trois aspects : (a) les faiblesses de l'approche méthodologique que nous venons d'évoquer; (b) le manque de prise en compte des stratégies d'évitement mises en place par les apprenants. « EA, which focuses exclusively on what learners do, has non way of investigating avoidance and is, therefore, seriously limited » (Ellis 1994, p. 68) ; (c) l'analyse statique des erreurs qui ne permet pas d'évaluer les stades de développement de l'interlangue. Dans ce dernier aspect il est possible de rechercher les raisons qui justifient l'émergence d'une nouvelle approche d'étude de l'interlangue basée sur l'identification d'itinéraires acquisitionnels. Telle méthodologie propose de regrouper les analyses des erreurs isolées afin de parvenir à une description de l'évolution de l'interlangue sous forme de parcours de développement.

4.3. Itinéraires, profils et stratégies d'acquisition

4.3.1. Itinéraires acquisitionnels et profils des apprenants

Les recherches en linguistique acquisitionnelle (Dittmar et Klein 1979 ; Meisel, Clahsen et Pienemann 1981 ; Klein et Perdue 1997 ; Pienemann 1998 ; Bartning et Schlyter 2004) montrent depuis les années 80 que l'interlangue de l'apprenant évolue souvent selon un certain ordre quel que soit le type d'acquisition (guidée ou non guidée) et le contexte (homoglotte ou hétéroglotte). Ces parcours d'acquisition, appelés « itinéraires acquisitionnels » ou encore « parcours développementaux », ont été observés dans les processus d'apprentissage de plusieurs langues cibles[192]. L'examen des itinéraires de différents aspects isolés (morphosyntaxiques, lexicaux, syntaxiques etc.) ont été regroupés afin de définir des stades d'acquisition, caractérisés par un ensemble de traits spécifiques, capables de représenter le niveau du lecte de l'apprenant à un moment donné de l'évolution de son apprentissage. Les premiers parcours[193] ont été établis à partir des données (selon une approche *bottom-up*) recueillies auprès d'apprenants en milieu naturel (Noyau 1980) dans le cadre du projet

[192] Parmi celles-ci : l'anglais, l'allemand, le hollandais, le français et le suédois.

[193] Cependant, il convient de souligner que des études précédentes avaient déjà été menées sur l'ordre naturel d'acquisition des morphèmes en anglais L2 (Dulay et Burt 1979) mais elles avaient fait l'objet de nombreuses critiques pour des raisons méthodologiques et pour l'absence d'un cadre théorique (Larsen-Freeman et Long 1991).

Heidelberger Forschungsprojekt Pidgin Deutsch (dorénavant HPD) (Dittmar et Klein 1979) et le programme *Zweitspracherwerb Italienischer und Spanischer Arbeiter* (dorénavant ZISA) (Meisel, Clahsen et Pienemann 1981). Ces travaux pionniers seront suivis par une étude à grande échelle conduite par le programme de recherche *European Science Foundation* (dorénavant ESF) (Klein et Perdue 1992 ; Klein et Perdue 1997 ; Klein 1998). Les ordres de développement observés par ces études en milieu d'apprentissage naturel seront validés par d'autres recherches conduites auprès d'apprenants ayant suivi une acquisition guidée. C'est le cas par exemple des programmes d'études d'acquisition du FLE lancés en Suède par les universités de Lund et Stockolm (Bartning et Schlyter 2004). Les itinéraires acquisitionnels proposés par l'ensemble de ces recherches semblent être valables pour les deux types d'apprenants (guidé et non-guidé). Cette similarité suggère que le développement de l'interlangue de l'apprenant n'est pas « aléatoire » (Bartning et Schlyter 2004, p. 297) et plutôt conditionné par des principes cognitifs et psycholinguistiques que par des facteurs externes (cf. Paragraphe 3.2). Dans cette perspective, s'appuyant sur les données empiriques issues de l'étude de plusieurs langues (voir données ESF), Pienemann (1998) propose un modèle (*Processability Theory* à savoir la « théorie de la processabilité ») selon lequel l'ordre d'acquisition grammaticale dans une L2 serait déterminée par une hiérarchie de contraintes psycholinguistiques qui conditionnent la processabilité des structures grammaticales.

4.3.1.1 Le projet « Heidelberger Forschungsprojekt Pidgin Deutsch » (HPD) (Dittmar et Klein 1979)

Ce projet pionnier a été mené par Dittmar et Klein entre 1974 et 1978 à Heidelberg auprès de 48 locuteurs d'allemand L2 - hispanophones et italophones pour la plupart - en milieu naturel. L'examen des données issues des entretiens a conduit les chercheurs à établir une première échelle implicationnelle d'acquisition des règles syntaxiques en allemand[194] posant les bases pour le développement du projet de recherche ESF.

4.3.1.2 Le programme « Zweitspracherwerb Italienischer und Spanischer Arbeiter » (ZISA) (Meisel, Clahsen et Pienemann 1981)

Le programme *Zweitspracherwerb Italienischer und Spanischer Arbeiter* (ZISA) (Meisel, Clahsen et Pienemann 1981) comprend une étude transversale conduite auprès de 45 apprenants et une étude longitudinale auprès de 12 locuteurs migrants hispanophones et

[194] Ces indices syntaxiques seront en outre mis en relation avec des variables sociales (telles que le contact avec les natifs au travail et pendant le temps libre, l'âge au moment de l'immigration et le niveau d'étude et la durée du séjour).

italophones. Comme le programme HPD, cette recherche a abouti à l'élaboration d'un modèle implicationel d'apprentissage de la morphosyntaxe de l'allemand dans lequel l'acquisition est le résultat de la combinaison entre le niveau d'appropriation des règles linguistiques et des variables extralinguistiques qui conditionneraient la capacité de mettre en œuvre les connaissances acquises. L'itinéraire acquisitionnel proposé par le programme ZISA sera ensuite confirmé par les études suivantes menées sur l'allemand, notamment par les recherches du projet ESF.

4.3.1.3 Le programme « European Science Foundation » (ESF)

Les recherches du projet HPD et ZISA des années 70 poseront les bases pour l'élaboration du programme de recherche *European Science Foundation* (Dittmar et Klein 1979 ; Klein et Perdue 1992 ; Klein et Perdue 1997 ; Klein 1998) qui donnera lieu à une vaste étude longitudinale conduite auprès de 40 locuteurs en milieu naturel issus de cinq pays européens. Les données recueillies sont issues de six langues sources typologiquement différentes et cinq langues cibles (anglais, allemand, hollandais, français et suédois). De plus les protocoles des études précédentes (principalement basés sur des entretiens) ont été enrichis par plusieurs tâches telles que le récit, des descriptions d'images et des jeux de rôle.

Ce projet a abouti, à partir de l'identification d'une variété de base, à la définition d'un itinéraire acquisitionnel composé par des niveaux : pré-basique, basique et post-basique. Ceux-ci permettent de regrouper, sous forme d'une échelle implicationnelle, les traits morphosyntaxiques caractéristiques de différentes variétés de l'interlangue de l'apprenant qui indiquent un passage des modalités d'apprentissage pragmatiques à d'autres d'ordre syntaxiques (Givón 1979). Si les premières variétés semblent en effet être guidées par des facteurs pragmatiques (s'appuyant principalement sur des ressources de la L2), les deuxièmes présentent une morphologie et une syntaxe plus élaborées.

Dans la variété pré-basique l'énoncé est encore très court et, dans la plupart des cas, nominal organisé autour de quelques mots-clés. La faible autonomie communicative oblige l'apprenant à avoir recours à des stratégies communicatives (demande d'aide au natif, gestes, références au contexte etc.). L'organisation de l'énoncé est donc régie par des principes pragmatiques (présence du seul rhème ou ordre thème-rhème) et la morphologie et la syntaxe sont presque absentes. Ceux-ci commencent à émerger dans les variétés basiques également appelées « stades du verbe non fléchi » (Klein et Perdue 1997) où l'énoncé est organisé autour d'un verbe dont les traits morphologiques (mode, temps et personnes) ne sont pas encore complétement maîtrisés. À ce stade l'énoncé est encore gouverné par des principes

pragmatiques (*Focus last* à savoir l'ordre thème-rhème) et sémantique (*Controller first* selon lequel le syntagme indiquant l'action se situe en début d'énoncé) et la subordination est peu attestée. Dans les variétés post-basiques la modalité syntaxique (Givón 1979) définit l'organisation de l'énoncé. La morphologie devient plus riche et les désinences acquièrent une valeur fonctionnelle (au moins dans le cas des langues indoeuropéennes observées). La syntaxe devient plus complexe avec l'apparition de la subordination.

Ces itinéraires acquisitionnels, établis à partir des données empiriques recueillies dans des contextes d'apprentissage en milieu naturel, seront confirmés par des études successives menées auprès des apprenants ayant suivi une acquisition guidée.

4.3.1.4 Programmes d'études sur les itinéraires acquisitionnels et profils grammaticaux du français L2

La première recherche menée sur les itinéraires acquisitionnels du français L2 a été réalisée par Véronique (1995). Cet auteur propose de distinguer trois phases principales dans l'acquisition du français L2 indépendamment de l'âge et du type d'apprentissage de la langue cible. Celles-ci s'appuient principalement sur des traits grammaticaux différents qui seront ensuite repris par les études suivantes.

Plus récemment des itinéraires acquisitionnels du FLE ont été proposés par des études longitudinales et transversales menées auprès des apprenants adultes suédois en milieu institutionnel. Ces recherches ont été conduites par Bartning et Schlyter (2004) à partir de deux différents corpus basés principalement sur des entretiens, des récits et des traductions. D'une part le corpus « Interlangue Française » (*InterFra*) (Universités de Stockholm) comportant des productions d'apprenants (semi-) guidés et d'autre part le corpus de Lund *Determinarfrasens Utveckling i Romanska Språk* (*DURS*) contenant des productions d'apprenants guidés et non-guidés[195].

Ce programme a mis au point des itinéraires acquisitionnels de différents phénomènes isolés du français langue étrangère. Ceux-ci portent particulièrement sur le développement de la morphologie verbale, de l'accord, du système temporel et modal, de la forme et de la place de la négation ainsi que des parcours développementaux des pronoms et de la subordination. L'ensemble de ces itinéraires ont été réunis en stades, chacun correspondant à un profil d'apprenants qui comprend des traits syntaxiques et grammaticaux caractéristiques d'un

[195] Les locuteurs du niveau débutant sont des apprenants universitaires, ceux de niveau intermédiaire sont en revanche des lycéens alors que les niveaux plus avancés sont représentés par des futurs professeurs de français et des doctorants.

moment précis de son développement (Bartning et Schlyter 2004, p. 281‑282). Les auteurs distinguent six stades qui vont des débuts de l'acquisition jusqu'aux productions des apprenants quasi-natifs : (a) stade initial ; (b) stade post-initial ; (c) stade intermédiaire ; (d) stade avancé bas ; (e) stade avancé moyen ; (f) stade avancé supérieur. Nous présentons ci-dessous la caractérisation des stades proposée par Bartning et Schlyter (2004, p. 293‑296).

Stade 1 : Le stade initial

Au stade initial, l'apprenant se sert d'une structuration en grande partie nominale, mais aussi de moyens grammaticaux. Il/elle emploie des 'formes finies courtes et des formes non finies, ce qui montre que cette opposition n'est pas maîtrisée. Il n'y a pas non plus d'opposition entre les personnes des formes verbales. On trouve pourtant à ce stade l'article défini et indéfini ainsi que le pronom sujet *je*, bien que très souvent accentués et non élidés.

D'autres traits caractéristiques de ce stade sont l'emploi de la négation Nég X (*non grand lit*) et les formules telles que *je(ne)sais pas* et *je m'appelle*. L'apprenant utilise quelques rares formes du passé composé, mais très peu de contextes du passé sont marqués. Il y a aussi un début d'emploi de connecteurs *et*, *mais* et *puis*. Ce stade peut être comparé au *niveau pré-basique* ainsi qu'à la transition vers la *variété basique* (Klein et Perdue 1997).

Stade 2 : Le stade postinitial

Ce stade est caractérisé par l'apparition de quelques phénomènes grammaticaux, mais ceux-ci sont encore très variables. La subordination simple introduite par *quand*, *parce que* et *qui/que* apparaît. Il y a apparition de la négation préverbale *ne* sans pas, à côté de la négation postverbale. L'apprenant commence à utiliser des formes verbales modales (suivis d'un infinitif) et le futur périphrastique. Son emploi du passé composé devient plus productif.

On voit ainsi que certains morphèmes grammaticaux libres apparaissent assez tôt. Quelques apprenants guidés utilisent parfois l'imparfait, le plus souvent sous la forme de *était* et de *avait*. Les formes verbales non finies (*ils *prendre*) utilisées dans un contexte fini sont encore fréquentes, mais le nombre de 'formes finies courtes' augmente (*je parle*). Il y a aussi les premières occurrences de *ils *prend*. L'accord sujet-verbe est marqué par l'opposition entre la 1re et 2e personne au singulier pour les verbes non thématiques être et avoir ainsi que par *nous V-ons* en alternance avec **nous V*, sans désinence. Les pronoms objets sont généralement postposés. L'apprenant à ce stade a aussi souvent recours à *c'est* en tant que constructeur d'énoncés et de structures passe-partout.

Stade 3 : Le stade intermédiaire

À ce stade, l'apprenant semble avoir développé une interlangue plus systématique et régulière, bien qu'encore très simple. Cette interlangue contient des surextensions et des régularisations qui font qu'elle n'est pas toujours en accord avec les normes de la langue cible. L'apprenant utilise cependant la négation (*ne Vfin pas*) comme dans la langue cible. De plus, il emploie le passé avec la plupart des références marquées au passé ainsi que le futur (périphrastique; le futur simple, dans des cas isolés) lorsqu'il fait référence à l'avenir. Les pronoms objets sont placés devant le verbe lexical pour les temps composés et simples - souvent de façon incorrecte après les auxiliaires *est/a*. On observe également un enrichissement de la subordination par des subordonnées causales, temporelles, relatives, complétives et interrogatives et c'est également à ce stade qu'on voit apparaître les premières occurrences du subjonctif. Les formes verbales non finies dans des contextes finis existent

encore, mais se raréfient. *Nous V-ons* est marqué la plupart du temps. L'opposition entre la 3e personne du singulier et du pluriel pour *ont*, *sont*, etc. commence à être plus constante (environ la moitié des cas), mais l'apprenant continue à employer en alternance *ils *prendre*, *ils *prend* et plus rarement *ils prennent*.

Stades 4-6 - Les stades avancés

Ces stades sont représentés par un éventail plus large de différents types de structures (pour les énoncés) et par la grammaticalisation de la morphologie flexionnelle qui devient fonctionnelle, mais avec des zones « fragiles » de développement. La raison pour laquelle la variété avancée est divisée en trois niveaux est due au fait qu'au fur et à mesure que l'interlangue se développe, cette évolution crée une richesse d'expressions qui permettent un plus grand choix dans le répertoire.

Stade 4 : Le stade avancé bas

Au niveau 4, les structures spécifiques du français, plus complexes et variées, apparaissent : le pronom clitique avant l'auxiliaire, le conditionnel, le plus-que-parfait et le subjonctif. Des formes verbales plus complexes commencent donc à apparaître - formes qui sont liées à des structures syntaxiques multi-propositionnelles (comme le subjonctif), ou impliquant une relation discursive (relations temporelles et conditionnelles); ces contextes ne sont cependant pas encore systématiquement marqués par les formes pertinentes. Les négations deviennent complexes, contiennent rien, jamais, personne et elles sont, pour la plupart, correctes et bien placées. À ce niveau, l'utilisation de formes verbales non finies (ils*parler) au lieu d'une forme finie n'existe pratiquement plus. Les formes ont et sont deviennent plus productives et dominent sur ils *a/*est. De plus, le pluriel est souvent marqué sur les verbes lexicaux, du type ils prennent, à côté de ils *prend. La clitisation de l'article est acquise ainsi que celle du pronom sujet. Le genre de l'article pose néanmoins toujours des problèmes à l'apprenant. Il y a diversification des connecteurs ainsi qu'un suremploi significatif des marqueurs polyfonctionnels mais et parce que.

Stade 5 : Le stade avancé moyen

L'un des traits les plus caractéristiques de ce stade est également lié au développement de la morphologie flexionnelle : les formes verbales *ils sont*, *ont*, *etc*. sont employées correctement et, dans la plupart des cas, le plus-que-parfait, le conditionnel et le futur simple. C'est à ce niveau que le subjonctif devient plus productif. L'accord du genre du déterminant et surtout de l'adjectif pose toujours des problèmes. Pourtant, même à ce niveau, on retrouve encore des verbes lexicaux irréguliers conjugués au singulier avec un pronom pluriel comme *ils *sort*. La négation est utilisée en tant que sujet (*personne ne sait…*). On trouve aussi des relatives avec *dont* et le gérondif.

Stade 6 : Le stade avancé supérieur

Ce stade est caractérisé par une morphologie flexionnelle utilisée de manière constante, même dans les énoncés multi-propositionnels. On observe un emploi des connecteurs enfin et donc qui se comparent à celui des locuteurs natifs, un haut degré d'empaquetage, d'ellipse et d'intégration des propositions caractérisantes qui démontrent une capacité à gérer plusieurs niveaux informationnels au sein du même énoncé. On observe également un emploi des relatives macro-syntaxiques presque comparable à celui des locuteurs natifs.

Parmi les itinéraires des phénomènes isolés décrits par ces auteurs, deux nous paraissent pertinents pour notre étude. D'une part l'itinéraire proposé pour la morphologie verbale de l'accord (sujet-verbe au pluriel) et d'autre part celui du développement des pronoms. Ces deux itinéraires acquisitionnels serviront de cadre théorique pour la description de l'apprentissage de la liaison catégorique dans la quatrième partie de ce travail (cf. Partie 4.Chapitre 3).

Itinéraire acquisitionnel de l'accord sujet-verbe au pluriel

Bartning et Schlyter (2004, p. 286) proposent d'étudier l'acquisition de l'accord sujet-verbe (personne / nombre, tous verbes confondus) sur la base des étapes suivantes :

- l'opposition au singulier est bien maıtrisée pour les copules et auxiliaires (« je suis » ; « il est »; « j'ai » ; « il a » etc.) ;
- l'accord de la 1ère personne du pluriel « nous V-ons » est maıtrisé (= au moins 75% des contextes sont marqués) ;
- l'accord de la 3ème p. pl. des verbes en -ont, (« ils sont » / « ont » / « vont » / « font ») est maıtrisé ;
- l'accord de la 3ème p. pl. des verbes lexicaux irréguliers (« ils prennent » etc.) est correct ;
- l'accord de la 3ème p. pl. des verbes lexicaux irréguliers est correct.

Les auteurs soulignent en outre que le développement de cet aspect de la morphologie verbale serait particulièrement long en contexte d'apprentissage tant guidé que non guidé et semblerait suivre la progression suivante : (1) « je suis » ; « il est »; (2) « nous V-ons » ; (3) « ils ont » (4) « ils Vlex-nt ».

Itinéraire acquisitionnel des pronoms objet

L'itinéraire acquisitionnel proposé par les auteurs pour le développement des pronoms objet (Bartning et Schlyter 2004, p. 290) semble être validé par d'autres études précédentes menées auprès d'apprenants anglophones. Il s'articule selon les étapes suivantes :

- les apprenants omettent le pronom ou le remplacent par un syntagme nominal ;
- le pronom (toujours accentué mais dans des formes différentes) apparaît après le verbe lexical : « ∗il dit lui /∗je veux mange toi » (Petra 1) ; « ∗on prend le gaz et refroidir le » (Karl 1) ;

- l'apprenant place le pronom dans une position intermédiaire, entre le verbe modal et le verbe lexical, ou de manière incorrecte entre l'auxiliaire et le verbe lexical ; « je peux le faire / j'ai # j'ai le vu » (Karl 2) ; « il a ass-* il a lui assis » (Petra 2) ;
- la position avant un verbe lexical fini semblerait apparaître peu après ou presque simultanément ;
- à un niveau assez avancé l'apprenant parvient à placer l'objet devant l'auxiliaire: « je l'ai pris / ça m'a changé / je t'ai dit » (Knut 1,3) ;
- les pronoms « en » et « y » n'apparaissent qu'à un niveau très avancé.

Il est intéressant d'observer comment les itinéraires à six stades proposés par ces chercheurs semblent être valables pour les deux types d'apprenants (guidés et non-guidés). Cette similarité suggère que le développement de l'interlangue de l'apprenant n'est pas « aléatoire » et davantage conditionné par des principes cognitifs et psycholinguistiques que par des facteurs externes (Bartning et Schlyter 2004, p. 297). Ces principes seront modélisés par Pienemann qui en 1998 élabore la théorie de la processabilité en postulant l'existence d'une série implicationnelle de contraintes psycholinguistiques capables de conditionner l'acquisition d'une L2.

4.3.1.5 Théorie de la Processabilité

La théorie de la Processabilité constitue un modèle cognitif d'apprentissage de la L2 élaboré par Pienemann (1998) qui se base à la fois sur la Grammaire Procédurale Incrémentale (*Incremental Procedural Grammar*) (Kempen et Hoenkamp 1987) et sur le modèle psycholinguistique de Levelt (1989)[196]. Dans la lignée des études empiriques menées par le projet ESF, Pienemann (1998) propose de considérer l'apprentissage d'une L2 comme un processus soumis à une hiérarchie de contraintes psycholinguistiques qui déterminerait l'ordre de développement acquisitionnel. L'apprentissage et donc la processabilité des structures grammaticales (structures morphologiques et syntaxiques) suit une échelle de stades acquisitionnels qui conduisent progressivement l'apprenant à la production de l'énoncé allant d'un premier accès au lemme jusqu'à l'organisation syntaxique de l'énoncé.

Au sein de cette hiérarchie acquisitionnelle (cf. Tableau 25), les échanges d'information entre les éléments se réalisent graduellement en fonction de la distance syntaxique entre les

[196] Selon ce modèle l'élaboration et la production du message linguistique ont recours à trois mécanismes qui interagissent entre eux : (a) le conceptualisateur qui prépare l'information en vue de sa verbalisation ; (b) le formulateur qui codifie verbalement le message au niveau grammatical et phonologique ; (c) l'articulateur qui produit le message acoustique (cf. Partie 1.Chapitre 6.1.3.1).

unités. Au premier niveau (*Lemma acces,* « mot / lemme »), l'apprenant n'acquiert que les lemmes et ses caractéristiques (forme et sens de l'unité lexicale) sans la morphologie spécifique de la langue cible. Ensuite, au deuxième niveau (*Category procedure,* « procédures de catégorie »), l'apprenant est capable d'attribuer aux lexèmes des catégories syntaxiques spécifiques en utilisant les morphèmes grammaticaux adéquats (les désinences du genre, nombre etc.). Il est donc en mesure, au troisième niveau (*Phrasal procedures,* « procédures intra-syntagmatiques »), d'échanger l'information morphologique et syntaxique entre les constituants du syntagme (ex. les accords de genre et nombre). Les trois derniers niveaux correspondent à la mise en place d'une procédure (*S-procedure*) qui implique un échange d'information inter-syntagmatiques et l'attribution de fonctions syntaxiques des constituants d'abord entre les éléments les plus saillants (*Simplified S-procedure*, « procédures inter-syntagmatiques simplifiées») puis les moins saillants (*S-procedure,* « procédures inter-syntagmatiques »). Ce n'est qu'au dernier stade que l'apprenant est capable de définir des rapports de subordination (*Subordinate Clause Procedure,* « procédures de phrase subordonnée»).

Procédure de processabilité (Pienemann 1998)	**Domaine structural**	**Échange d'information / Marquage d'accord**
5. Proc. de phrase subordonnée (*Subordinate clause procedure*)	Phrase principale et phrase subordonné	Accord entre la phrase principale et la phrase subordonné
4. Proc. Inter-syntagmatique (*S-procedure*)	Phrase simple	Accord entre syntagmes
3. Proc. Intra-syntagmatique (*Phrasal procedures*)	Syntagme	Accord entre les constituants d'un syntagme
2. Proc. de catégorie (*Category procedure*)	Morphèmes lexicaux	Attribution du genre et du nombre
1. Mot / lemme (*Lemma acces*)	Lexème	-

Tableau 25 : Hiérarchie implicationnelle de la théorie de la processabilité de Pienemann (1998)

Comme le souligne Pienemann (1998) le caractère hiérarchique des procédures et des règles de ce modèle découle du fait que chaque étape représente une condition préalable pour l'accès au niveau d'acquisition suivant.

> « The hierarchical nature of this arises from the fact that the procedure of each lower level is a prerequisite for the functioning of the higher level: A word needs to be added to the L2 lexicon before its grammatical category can be assigned. The grammatical category of a lemma is needed before a category procedure can be called. Only if the grammatical category of the head of phrase is assigned can the phrasal procedure be called. Only if a phrasal procedure has been completed and its value is returned can Appointment Rules determine the function of the phrase. And only if the function of the phrase has been determined can it be attached to the S-node and sentential information be stored in the S-holder ». (Pienemann 1998, p. 80)[197]

Un tel système permet d'organiser les formes de l'interlangue de l'apprenant selon des séquences d'acquisition dans la mesure où la processabilité des productions en L2 est soumise à une échelle implicationnelle. L'apprenant ne peut acquérir que les structures de la L2 disponibles au niveau d'apprentissage qu'il a atteint.

> « What is relevant for the structure of the interlanguage is the fact that the lack of access to syntactic information about the L2 lexical item blocks the language production process off, since no phrasal and other procedures can be called, no grammatical information can be exchanged, and phrases cannot be assigned functions. This means that all that can happen at this stage is the mapping of conceptual structures onto individual words and fixed phrases ». (Pienemann 1998, p. 83)

Pienemann (1998 ; 2005) défendra la validité de ce modèle théorique en démontrant son applicabilité non seulement à l'acquisition de l'allemand mais à celle d'un ensemble de langues typologiquement différentes telles que l'anglais, le suédois et le japonais. Ce cadre théorique, fondé sur des données empiriques de développement de la morphosyntaxe à l'oral en L2, nous permettra dans la quatrième partie de ce travail (cf. Partie 4.Chapitre 3) de rendre compte de certains aspects de la l'acquisition de la liaison en FLE.

4.3.2. Stratégies de l'apprenant

Les travaux conduits sur les itinéraires d'acquisition ont sans aucun doute contribué à l'émergence de la question des stratégies de l'apprenant. La notion de « stratégies » rend compte, selon une approche psycholinguistique, de la façon dont l'apprenant s'acquitte de « tâches », que celles-ci soient communicationnelles ou acquisitionnelles (Véronique 1992, p. 6). À la suite de Selinker (1972), Frauenfelder et Porquier (1979 cité par Véronique 1992) proposent de distinguer entre stratégies d'apprentissage et stratégies de communication. Les premières relèvent de l'activité d'acquisition et correspondent à l'inférence, la mémorisation, la répétition mentale, l'association, l'analogie et la régularisation par généralisation, la

[197] Selon Kempen et Hoenkamp (1987), et Pieneman (1998), les fonctions syntaxiques des syntagmes (« phrases ») sont définies par des règles qui varient d'une langue à l'autre (*Appointment Rule*) et permettent d'ancrer les éléments dans la structure de l'énoncé « sentence » (S). La procédure finale S (*S-procedure*) organise et structure l'information syntagmatique de l'énoncé.

simplification des formes marquées (à différents niveau phonologique, morphologique et syntaxique) et le recours à la L1. Quant aux stratégies de communication, elles renvoient aux tentatives de l'apprenant et de son interlocuteur de s'accorder sur un sens dans des situations communicatives où les structures linguistiques et sociolinguistiques ne sont pas partagées (Tarone 1979). Parmi celles-ci le recours à la paraphrase, la circonlocution, la simplification, le mime, des stratégies d'évitement des structures marquées et peu connues par l'apprenant et la recherche conséquente de formes d'expression alternatives. D'après Tarone (1979), les apprenants utiliseraient également des stratégies de production ayant pour objectif l'utilisation optimale de leur interlangue. Certains auteurs comme Bialystok et Smith (1985) rejettent la distinction entre stratégies d'apprentissage et stratégies de communication qu'ils jugent non explicative puisque la notion même de stratégie ferait référence à une dimension descriptive plutôt qu'explicative. Ils proposent donc de distinguer entre des stratégies de mise en œuvre des connaissances (*knowledge-based strategies*) basées sur la connaissance de la langue (créations lexicales, paraphrases et circonlocutions) et des stratégies de contrôle (*control-based strategies*) ayant pour objectif de gérer les difficultés résultant d'un manque de connaissances linguistiques ou d'un insuccès de la communication.

Cependant, l'opposition traditionnelle entre stratégies de communication et stratégies d'apprentissage semblent être adoptée par le Cadre européen commun de référence pour les langues. Dans le référentiel de niveau B2 pour le français Beacco et al. (2004) distinguent entre stratégies d'assimilation (saisie et intégration) et stratégies d'exécution des tâches bien qu'ils précisent que les deux sont souvent « imbriquées » (Beacco, Bouquet et Porquier 2004, p. 357).

> « Il s'agit des moyens mobilisés par l'apprenant : pour appréhender des données nouvelles, les intégrer à sa compétence, les conserver en mémoire, s'entrainer à les utiliser ; pour effectuer des tâches d'apprentissage ; pour évaluer et améliorer ses compétences et ses progrès. Ces stratégies peuvent être suggérées, incitées voire inculquées. Cependant, elles relèvent en bonne partie des apprenants, de leur choix et de leur style d'apprentissage ». (Beacco, Bouquet et Porquier 2004, p. 357)

Le bref examen de ces stratégies montre que l'étude de l'interlangue de l'apprenant comporte la prise en compte de la dimension cognitive et psycholinguistique de l'acquisition d'une L2.

4.3.3. Limites

Bien que ces études se fondent sur des données empiriques solides, Ellis (1994, p. 111) souligne qu'il serait plus prudent de parler de régularités d'apprentissage plus que d'itinéraires définis étant donné : (a) l'attention principalement portée à l'analyse des éléments

grammaticaux[198]; (b) la présence de certains problèmes méthodologiques qui cependant ne semble pas compromettre la robustesse des résultats[199] ; (c) l'existence d'une variabilité inter et intra apprenants observée dans ces parcours de développement.

En dépit de ces observations, Ellis (1994) semble reconnaître la validité de cette approche d'analyse qui prouve l'autonomie du système dynamique de l'interlangue par rapport à la langue cible postulée par Selinker (1972).

> « The discovery that learners do follow identifiable routes has generated immense interest. Any theory of L2 acquisition will need to account for developmental patterns. The theory that has been dominant in SLA, interlanguage theory (Selinker 1972), was initially formulated to provide such an account ». (Ellis 1994, p. 114)

5. Conclusion

En résumé, l'ensemble des données recueillies par les études que nous avons brièvement présentées ont abouti à l'identification des itinéraires acquisitionnels qui permettent de décrire les stades d'appropriation de plusieurs aspects morphosyntaxiques dans l'interlangue de l'apprenant. L'application de ces parcours à plusieurs langues typologiquement différentes ainsi que la grande quantité de données empiriques fournissent des arguments solides en faveur de l'existence de stades d'acquisition dans l'apprentissage d'une L2.

Cependant, selon Larsen-Freeman et Long (1991, p. 92), ces parcours de développement ne peuvent pas être établis à partir de l'identification de stades délimités par des frontières nettes. En revanche, ils devraient être inscrits dans la dynamique de l'interlangue de l'apprenant où chaque étape du développement est définie et ordonnée par rapport à la précédente. Une telle approche permet d'atténuer la rigidité des limites entre les stades tout en gardant leur caractère implicationnel.

La liaison étant, comme nous l'avons déjà souligné (cf. Partie 1.Chapitre 4.4), un phénomène multidimensionnel, son acquisition demande à l'apprenant de mettre en place un système d'apprentissage complexe qui s'articule sur plusieurs dimensions d'analyse.

[198] Sauf quelques exemples d'études menées sur l'acquisition du lexique et de la phonologie (voir Ellis 1994, p. 112‑113). Dans le cas de la phonologie, des itinéraires semblent émerger pour certaines formes de l'interlangue mais pas pour d'autres (Ioup et Weinberger 1987). En revanche, en acquisition L1, différents modèles basés sur des principes universels de distribution des sons (lois universelles de solidarité) (Jakobson 1941) et sur la fréquence (Zamuner, Gerken et Hammond 2005) ont été élaborés pour rendre compte de l'ordre d'acquisition de la phonologie. Yamaguchi (2015) propose un modèle mixte qui combine le pricincipes phonologiques universels et les fréquences des traits.

[199] « Despite all these problems, the work on developmental patterns has been substantial and in many cases convincing ». (Ellis 1994, p. 114)

Les modèles basés sur les itinéraires acquisitionels, comme nous le verrons dans quatrième partie de ce travail (cf. Partie 4.Chapitre 3.3), pourraient représenter un cadre théorique intéressant pour la description de son apprentissage.

Chapitre 2. L'acquisition de la liaison en français langue seconde et étrangère

La présentation des notions théoriques sur l'acquisition d'une L2 nous a permis de décrire les principales caractéristiques de l'interlangue de l'apprenant et de le traiter comme un système en construction dont le développement semblerait suivre un ordre d'apprentissage guidé initialement par des critères pragmatiques et puis syntaxiques. En ce qui concerne l'acquisition de la liaison à ce jour aucune hypothèse pour rendre compte de l'apprentissage en L2 n'a reçu un appui empirique convaincant. Cela est probablement lié à la faible représentativité de populations différentes d'apprenants (pour la plupart anglophones) dont font état les études précédentes, ainsi qu'aux difficultés de comparabilité qui caractérisent ces travaux. Ces aspects empêchent de formuler des généralisations sur les processus d'acquisition de la liaison en L2.

Néanmoins ces recherches ont le mérite d'avoir mis en évidence des typologies d'erreurs de liaison produites par les apprenants de français langue seconde ou étrangère (cf. Paragraphe 1.4.).

Une présentation de ces réalisations de liaison sera proposée dans ce chapitre à la suite d'un bilan et d'une analyse comparative des études menées sur l'acquisition de la liaison. Nous nous appuierons sur l'analyse des typologies des erreurs afin de mener une réflexion sur les stratégies d'acquisition de la liaison en L2. Cela nous amènera à nous interroger sur le rôle que la L1 et l'orthographe jouent dans l'acquisition de la liaison en L2.

1. Etudes précédentes

Jusqu'à présent, de nombreux travaux se sont intéressés à l'acquisition de la liaison en L2. Toutefois la plupart de ces études ont été menées auprès d'apprenants ayant comme langues sources des langues germaniques (anglais et suédois) (Dejean de la Bâtie 1993; Mastromonaco 1999; Howard 2005; Thomas 1998; Thomas 2002; Thomas 2004; Stridfeldt 2005; De Moras 2011) à l'exception d'une étude menée auprès d'apprenants coréens (Harnois-Delpiano, Cavalla et Chevrot 2012). À notre connaissance, contrairement au domaine de l'acquisition de la liaison en L1, il n'existe pas de modèles développementaux qui ont essayé de décrire les étapes d'appropriation de la liaison chez les apprenants de français L2. Cette absence pourrait être la conséquence à la fois du faible nombre de données disponibles et de leur hétérogénéité, qui constituent un obstacle à la formulation de généralisations sur l'apprentissage de la liaison en L2. En effet, les études de la liaison en L2 sont difficilement comparables pour des raisons

d'ordre sociolinguistique, didactique ou méthodologique : (1) hétérogénéité des niveaux observés et des tâches soumises aux apprenants ; (2) contextes et conditions d'apprentissage différents (présence / absence d'apprentissage formel et de correction phonétique, type d'approche d'enseignement utilisée, contexte d'acquisition homoglotte / hétéroglotte, séjour en France etc.) ; (3) modalités d'enseignement ; (4) critères de classification de liaison choisis par l'étude[200]. La diversité de ces aspects est présentée dans le tableau (cf. Tableau 26) qui résume les caractéristiques principales de ces travaux : (a) nombre et origine des apprenants; (b) âge des apprenants ; (c) niveau des apprenants ; (d) contexte d'apprentissage (homoglotte / hétéroglotte / bilingue) ; (e) tâches ; (f) classifications employées par les différentes études.

Étude	Apprenants	Âge	Niveau	Contexte	Modalité d'acquisition	Tâches	Classification
De La Batie et Bradley (1995)	60 anglophones	18-50 ans	débutant /intermédiaire	hétéroglotte	guidée	perception des frontières de mots	–
Stridtfeld (2005)	22 suédophones	–	intermédiaire	hétéroglotte	guidée	perception des frontières de mots	–
Shoemaker et Birdsong. (2008)	15 anglophones	24-55 ans	avancé	hétéroglotte	guidée	perception des frontières de mots	–
Shoemaker (2010)	18 anglophones	19-54 ans	avancé	hétéroglotte	guidée	perception des frontières de mots	–
Tremblay (2011)	33 anglophones	18-31 ans	intermédiaire/ avancé	hétéroglotte	guidée	perception des frontières de mots	–
Thomas (2002 ; 2004)	87 anglophones	–	avancé	les deux	guidée	entretien semi-dirigé + 2 lectures	Delattre, 1947, 1966
Mastro monaco (1999)	68 anglophones	19-21 ans	intermédiaire	région bilingue	guidée	parole spontanée ; lecture ; description d'image	Delattre, 1947, 1966
Howard (2005)	12 anglophones	–	avancé	les deux	guidée	entretien sociolinguistique	Delattre, 1947, 1966
De Moras (2011)	37 anglophones	19[201] ans	A2-B1	les deux	guidée	lectures	Delattre, 1947, 1966
Harnois-Delpiano (2012)	16 koréanophones	–	débutant	hétéroglotte	guidée	dénomination d'image ; test de jugement	Encrevé, 1988
Howard (2013)	5 anglophones	–	avancé	hétéroglotte	guidée	entretien semi-dirigé	Delattre, 1947, 1966
Splendido (2014)	3 suédophones	4 ans	débutant	hétéroglotte	guidée	jeux	Delattre, 1947, 1966

[200] Les catégorisations des liaisons ayant fait l'objet de plusieurs révisions (cf. Partie 1.Chapitre 4.3), les mêmes liaisons peuvent être traitées comme catégoriques ou variables d'une étude à l'autre.

[201] Âge moyen.

Wauquier et Shoemaker 2015	5 bangalophones et 5 sinophones	–	débutant	homoglotte	non-guidée	dénomination d'image	Delattre, 1947, 1966

Tableau 26 : Bilan des études sur l'acquisition de la liaison en L2

1.1. Traitement cognitif et perception de la liaison

Les études de De La Batie et Bradley (1995), Stridtfeld (2005), Shoemaker et Birdsong (2008), Shoemaker (2010) et Tremblay (2011) se différencient en particulier par rapport aux autres travaux par les objectifs et les méthodologies utilisées. Ces recherches s'inscrivent en effet dans un cadre psycholinguistique. À la suite des travaux de Matter (1986 cité par Stridtfeld 2005), elles ont pour objectif de tester le traitement cognitif de la liaison, notamment les effets de la liaison sur la reconnaissance des mots chez des apprenants de FLE. Ces études s'intéressent en particulier aux cas de « liaison apparente » à savoir les contextes où la CL appartenant au M1 correspond à la consonne initiale fixe du M2 comme par exemple dans « un grand talent ». Ceux-ci donneraient lieu à une situation d'ambiguïté momentanée puisque le mot « grand » peut déclencher une liaison en /t/ et le mot suivant « talent » commence par /t/. D'après l'étude De La Batie et Bradley (1995), menée auprès de 60 étudiants universitaires anglophones (australiens) de niveau intermédiaire et avancé en milieu hétéroglotte, ces cas de « liaison apparente » retarderaient la reconnaissance des mots chez des locuteurs natifs ainsi que chez les non-natifs. Toutefois, les premiers contrairement aux apprenants, seraient capables de résoudre l'ambiguïté liée à ces séquences (ex. « C'est un excellent tableau ») s'appuyant sur l'information contextuelle de l'énoncé.

Les résultats de l'étude de Stridtfeld (2005), menée auprès de 22 apprenants suédois de niveau intermédiaire en milieu hétéroglotte, vont dans la même direction. Stridtfeld (2005) postule que la liaison et plus généralement l'enchaînement constituent une difficulté importante pour la compréhension auditive du français (Léon et Léon 1997) surtout chez des apprenants suédophones. En effet, contrairement à l'organisation syllabique et prosodique de l'énoncé français associée au phénomène de l'enchaînement (cf. Partie 1.Chapitre 5), le suédois est caractérisé par une syllabation fermée et une accentuation de mot. Ces différences syllabiques et rythmiques entre le français et le suédois pourraient entraîner des difficultés de perception du français oral chez les apprenants suédophones[202]. De plus, dans le cadre de la phonologie auto-segmentale (cf. Partie 1.Chapitre 5.3.2), Stridtfeld (2005) avance l'hypothèse que le

[202] Cependant, cette hypothèse ne semblerait pas vérifiée par Matter chez des apprenants hollandais (1986 cité par Stridfeldt 2005) ni par certaines études menées chez les locuteurs natifs (cf. Partie 1.Chapitre 6.1.5).

manque d'une représentation des segments flottants pourrait également contribuer à retarder la reconnaissance des mots chez ces apprenants. En l'absence d'une représentation phonologique de la CL flottante, les apprenants seraient amenés à traiter la CL de la même manière que les consonnes de début de mot ce qui entrainerait des problèmes de segmentation. Si les résultats des test psycholinguistiques montrent que la réalisation de la « liaison apparente » (et donc ambiguë) retarde la reconnaissance des mots (ex. dans les cas des pseudo-mots « un navas »), « les résultats sont moins uniformes en ce qui concerne l'effet de la liaison » (ex. « un avas ») (Stridfeldt 2005, p. 764).

Les travaux de Shoemaker et Birdsong (2008) menés auprès d'apprenants anglophones de niveau avancé (étudiants et professeur de FLE à l'Université de Texas) suggèrent que les locuteurs natifs comme les non-natifs ne seraient pas en mesure de découper correctement les contextes des liaison potentielle en s'appuyant exclusivement sur les différences de durée acoustique des CL (Spinelli, Cutler et Mcqueen 2003) par rapport aux consonnes initiales. Le recours à des facteurs post-lexicaux tels que la fréquence lexicale des mots-cibles et la plausibilité des phrases ne semblent pas avoir non plus d'effet significatif sur le taux d'exactitude de l'identification. En revanche, Shoemaker (2010) montre que les apprenants anglophones de niveau avancé (ayant effectué de longs séjours dans un pays francophone) développent une sensibilité aux variations acoustiques leur permettant de distinguer les CL dans des contextes de liaison des consonnes fixes initiales des mots.

Les données expérimentales de l'étude de Tremblay (2011), basée sur une technique de suivi des mouvements oculaires (*eye-tracking*), indiquent que les apprenants reconnaissent plus facilement des pseudo-mots suivant la liaison /z/ (ex. « fameux élin ») que des pseudo-mots à initiale consonantique (ex. « fameux zéla »)[203]. En effet les apprenants feraient appel à la fréquence de distribution des phonèmes à l'initiale de mots afin de lever l'ambigüité sur les contextes de liaison potentielle. Ils seraient donc capables de distinguer les CL des consonnes initiales dans les cas des liaisons en /z/ étant donné le faible nombre de mots commençant en /z/ en français.

À la lumière de ces résultats, apparemment lacunaires, il s'avère difficile de tirer des conclusions sur le traitement cognitif de la liaison et sur les difficultés de segmentation rencontrées par les apprenants au cours de l'acquisition de la liaison.

[203] Dans un premier temps l'apprenant avait uniquement accès aux stimuli écrits affichés sur un écran. Dans une deuxième étape de l'expérimentation les mêmes stimuli ré-apparaissaient à la fois sous forme d'input écrit et oral.

1.2. Production de la liaison

Quant aux études menées sur la production de la liaison, certains travaux ont développé des analyses comparatives entre l'acquisition de liaison en L1 et FLE chez les adultes (Harnois-Delpiano, Cavalla et Chevrot 2012) ou chez les enfants en acquisition précoce (Splendido 2014). D'autres travaux, de nature plutôt sociolinguistique, offrent des analyses quantitatives visant à évaluer l'influence exercée par les variables linguistiques et sociales sur le niveau de maîtrise de la liaison.

L'étude longitudinale de Hanois-Delpiano et al. (2012), conduite auprès de 16 étudiants coréens débutants en milieu hétéroglotte propose une comparaison exploratoire de l'acquisition en FLE et L1 (Dugua 2006) (cf. Partie 1.Chapitre 5.3.5). Si les résultats de production des liaisons catégoriques montrent une progression régulière et significative entre les différentes périodes d'observation des apprenants FLE comme des enfants natifs, cela n'est pas le cas pour la réalisation de liaisons variables. Celle-ci évolue bien plus rapidement chez les natifs et, comme nous le verrons dans les pages qui suivent (cf. Paragraphe 3.1), l'acquisition de la liaison en FLE se distingue par rapport à la L1 quant à la nature des erreurs de liaison. De plus les résultats des tests de jugement menés par Hanois-Delpiano et al. (2012) suggèrent que, contrairement à l'acquisition de la liaison en L1, par référence à la graphie, la capacité à juger les liaisons précède, chez les apprenants de FLE, celle qui consiste à les produire.

L'étude de Splendido (2014), portant sur l'acquisition précoce de la liaison chez trois enfants de 4 ans, indique que chez ce type de sujets, la liaison semble se développer de manière similaire aux monolingues. Cependant, Spendido (2014) précise qu'il est possible d'observer des différences entre les deux groupes dans la phase initiale du développement qui n'a pas été l'objet de son enquête. Un développement proche de celui des natifs ne semble toutefois pas exclure la présence de traits déjà observés chez les apprenants L2 adultes.

> « Le développement des enfants L2 présente des ressemblances avec les L1 des études précédentes mais également avec les apprenants L2 de la littérature. Nous rappelons premièrement la grande variation interindividuelle, ce qui constituerait typiquement un trait L2. En effet les taux de CL réalisés varient énormément entre les trois apprenantes. L'une des enfants produit des liaisons non enchaînées, précédemment uniquement observées chez les apprenants L2 adultes. Le remplacement du mot1 [*à savoir le premier mot de la séquence de liaison*] *un* par *une* pourrait être interprété comme un dernier trait L2 ». (Splendido 2014, p. 254)

La présence de caractéristiques propres à la fois à l'acquisition de la L1 et L2 est également attestée par l'étude menée par Wauquier et Shoemaker (2015) auprès de 10 apprenants adultes bangalophones et sinophones en milieu homoglotte. Contrairement aux

études précédentes, cette recherche a le mérite d'avoir essayé d'examiner l'acquisition de la liaison non seulement chez des apprenants ayant une L1 typologiquement différente (voir aussi Harnois-Delpiano et al. 2012; Detey, Kawaguchi et Kondo sous presse) mais surtout en contexte d'apprentissage non-guidé (où l'acquisition est moins influencée par l'input écrit). Une telle approche a permis de mettre en évidence la présence de non-réalisations de la CL (« Voici deux ours » réalisé [vwa.si.dø.uʁs] au lieu de [vwa.si.dø.zuʁs]) et de liaisons non-enchaînées accompagnées d'un coup de glotte ([vwa.si.dø.ʔuʁs] au lieu de [vwa.si.dø.zuʁs]), mais également des types de substitution de la CL (cf. Paragraphe 1.4) le plus souvent produites en l'acquisition L1 (« Voici deux ours » réalisé [vwa.si.dø.nuʁs] au lieu de [vwa.si.dø.zuʁs]) (cf. Partie 1.Chapitre 5.3.5).

Les travaux de Mastromonaco (1999), Thomas (2002 ; 2004) et Howard (2005 ; 2013) menées auprès d'apprenants anglophones avancés en milieu universitaire au Canada et en Irlande sont de nature plus sociolinguistique. L'objectif de ces études est d'analyser le rôle joué par les facteurs linguistiques, stylistiques et sociaux (modalité d'acquisition, niveau de langue, séjours à l'étranger etc.) dans l'apprentissage de la liaison. Il s'agit, dans la plupart des cas, d'évaluer le niveau de maîtrise de la liaison variable chez les apprenants avancés en le comparant aux taux de réalisation des natifs issus des travaux de De Jong (1988), Malecot (1975) et Ashby (2003).

La recherche menée par Mastromonaco (1999) auprès de 68 étudiants anglophones de l'université de Toronto (ayant suivi une formation différente) vise en particulier à étudier l'influence de la modalité d'enseignement sur l'acquisition de la liaison. Aucune différence significative dans la production de la liaison n'a été enregistrée entre les deux groupes d'apprenants. Pour ce qui est de la réalisation de la liaison variable, Mastromonaco (1999) observe un taux global de réalisation d'environ 14%, tous contextes variables confondus qu'elle compare aux données des natifs, notamment aux résultats des études de De Jong (1988) (21,7%).

Quant à l'étude de Thomas (2002 ; 2004), il s'agit d'une enquête menée auprès de 87 sujets, tous étudiants anglophones avancés de FLE de l'université de l'Ontario divisés en deux groupes en fonction de la nature et de l'intensité de leur expérience linguistique menée en France. L'objectif du travail était d'évaluer les progrès linguistiques accomplis en français par deux groupes d'étudiants avancés ayant passé une année universitaire en France (groupe expérimental) et dans le sud de l'Ontario (groupe témoin).

Contrairement à Mastromonaco (1999), Thomas (2002) adopte une approche beaucoup plus normative en choisissant de comparer les productions des liaisons catégoriques des apprenants aux données issues du corpus de Malecot (1977).

> « La répartition des liaisons réalisées suggère que nos étudiants ont un comportement assez proche du parler de la bourgeoisie parisienne (Malécot, 1975) pour les déterminants, les pronoms et les prépositions. Mais, si on considère l'ensemble des liaisons obligatoires non réalisées, ils font beaucoup plus d'erreurs que les Français, notamment pour les adjectifs (surtout 'grand') et la conjonction 'quand' ». (Thomas 2002, p. 4)

Pour ce qui est de la réalisation des liaisons variables, la comparaison basée sur le corpus radiophonique d'Ågren (1973) montre que les étudiants « réalisent encore une fois beaucoup moins de liaisons que les francophones ». (Thomas 2002, p. 5)

L'étude menée par Howard (2005) vise à évaluer les différences de maîtrise de la variation de la liaison entre deux groupes d'étudiants anglophones en Irlande, un groupe d'étudiants ayant suivi deux ans d'enseignement à l'université d'une part et de l'autre un deuxième groupe possédant exactement le même niveau d'enseignement mais qui avait effectué un long séjour en France dans le cadre d'un programme d'échange international. Pendant ce séjour, les apprenants n'avaient pas suivi de cours formels de français, mais assistaient aux mêmes cours que leurs camarades français. Les résultats de cette recherche montrent bien que le contact informel lié au séjour en France favorise l'acquisition de la liaison sans pourtant donner lieu à une amélioration significative. Il observe en moyenne un taux global de réalisation de 12% dans tout contexte variable. De plus, Howard (2005) enregistre une absence complète de liaison dans certains contextes[204] chez les deux groupes d'étudiants ce qui le conduit à qualifier les non-réalisations de la liaison chez les apprenants comme défectives parce que constituant des « zones de fragilité » (cf. Partie 3.Chapitre 1.4.3) importantes dans l'acquisition de la liaison.

> « Ces différences qui font que l'apprenant se distingue nettement du locuteur natif constituent des zones de fragilité (Bartning 1997) importantes dans l'acquisition de la liaison par l'apprenant L2. Voire, bien que nos apprenants ayant fait un long séjour en France témoignent d'un taux de réalisation de la liaison plus élevé que les autres intervenants dans l'étude, ils ne se rapprochent tout d'abord pas du locuteur natif du point de vue de leur taux d'emploi général de la liaison ». (Howard 2005, p. 10)

[204] Il s'agit tout particulièrement des contextes de liaison suivants: après « quand »; après des adverbes et des prépositions polysyllabiques; entre un verbe et son complément; après des auxiliaires polysyllabiques; et entre un substantif et un verbe.

En revanche les résultats de l'étude de De Moras (2011), menée auprès de 37 apprenants[205] anglophones de niveau intermédiaire, témoignent d'un taux moyen de réalisation des liaisons variables supérieur (22,4%) par rapport aux études précédentes.

En résumé, un bilan de ces travaux sociolinguistiques, semble suggérer que le rôle du milieu de l'acquisition (hétéroglotte ou homoglotte consécutive à la réalisation de séjours linguistiques) dans l'évolution de liaison n'est pas crucial. Néanmoins, des différences (en particulier dans le cas des contextes catégoriques) ont été observées entre les différents groupes (Thomas 2002 ; Howard 2005) ou entre les périodes d'observation précédentes et successives au séjour dans un pays francophone (Howard 2013). À la suite d'une partie de l'acquisition en milieu naturel, les apprenants seraient amenés à réaliser la liaison dans un plus grand nombre de contextes morphosyntaxiques. Quant à l'influence de la stratégie d'enseignement, De Moras (2011), contrairement à Mastromonaco (1999), observe des effets de ce facteur sur l'acquisition de la liaison. Les apprenants ayant suivi un cours basé sur des exercices de répétition phonologique ou d'exemples de séquences de liaison ont obtenu de meilleurs scores de production que ceux ayant participé à des cours théoriques (qui explicitaient les règles de liaison) ou pendant lesquels les erreurs de liaison étaient simplement corrigées.

En ce qui concerne les taux globaux de réalisation de la liaison, si les productions des contextes catégoriques (proches de celui de natifs même dans des tâches de parole spontanée) semblent être presque systématiques (à quelques exceptions près), une faible maîtrise de la liaison variable a été enregistrée par ces études malgré le niveau avancé des apprenants enquêtés[206]. Ces données montrent que les apprenants ne maîtrisent pas la complexité associée à la variabilité de la liaison. Autrement dit, ils ne sont pas conscients des usages réels de la liaison variable chez les natifs. Les résultats de Thomas (1998) et Howard (2005) indiquent en effet que les taux de réalisation de la liaison dans certains contextes (fréquents chez les francophones), (ex. la liaison après un adjectif singulier ou après un adverbe monosyllabique), restent très faibles chez les apprenants, y compris chez ceux qui ont réalisé un séjour dans un pays francophone.

Plusieurs inconvénients nous paraissent cependant affecter ces conclusions. Tout d'abord, comme nous l'avons déjà souligné (cf. Tableau 27), ces travaux renvoient à des modèles de

[205] Ceux-ci ont été divisés en deux groupes selon l'approche didactique adoptée dans l'enseignement de la liaison. Les résultats de cette recherche montrent que les sujets soumis à un apprentissage par répétition présentent de meilleurs scores de réalisation que ceux ayant été exposés à des explications théoriques.

[206] Cependant, dans une étude longitudinale plus récente menée auprès d'apprenants avancés, Howard (2013) indique que les taux globaux de réalisation de la liaison variable chez certains apprenants seraient même plus importants que ceux qui ressortent des études conduites auprès des locuteurs natifs (Ashby 2003).

référence et à des descriptions de la liaison basés sur des enquêtes menées sur des variétés sociolinguistiques de prestige de la bourgeoisie parisienne des années 70 (Malécot 1975) ou encore à partir de registres formels de parole (Ågren 1973). Une comparaison avec des données plus récentes et issues d'une étude systématique de corpus plus vaste pourrait mettre en évidence des écarts moins importants entre les taux de réalisation de la liaison attestés chez les apprenants et chez les natifs. L'analyse de l'acquisition de la variabilité de la liaison ne serait plus formulée en termes de jugements de conformité à une variété de prestige du français (cf. Partie 1.Chapitre 3.1.1) et pourrait être enrichie d'une comparaison avec différentes variétés sociolinguistiques.

De plus, malgré les résultats homogènes de ces études sociolinguistiques (sauf dans le cas De Moras 2001), ces travaux, comme nous l'avons déjà observé (cf. Tableau 27), utilisent des protocoles de recherche souvent très différents. Certaines études s'appuient sur des entretiens semi-dirigés (Thomas 2002), sociolinguistiques (Howard 2005) ou sur la lecture (De Moras 2011) ou encore sur plusieurs épreuves (lectures de phrases ou de textes, description d'image, parole spontanée) (Mastromonaco 1999). Le choix du protocole d'enquête, comme le soulignent De Moras (2011) et Howard (2013), a un impact important sur la variabilité des contextes de liaison variable analysés. Des travaux basés sur des lectures de phrases ou de textes permettent de tester une plus grande variété de contextes linguistiques et donc d'induire les apprenants à produire des séquences de liaison plus rares que l'apprenant n'aurait pas produites spontanément. En revanche, les études s'appuyant sur des tâches de conversation ne permettent de répertorier qu'un nombre restreint de contextes de liaison. La spécificité des tâches linguistiques utilisées par les différents travaux est donc subordonnée à une diversité de contextes de liaison analysés.

De plus, les critères choisis pour identifier les liaisons réalisées dans l'analyse des productions des apprenants varient également d'une étude à l'autre. En particulier, la prise en compte de paramètres différents (ex. degré de conformité à la CL cible, présence ou absence d'enchaînement) déterminent un changement important dans les taux de réalisation des liaisons. Contrairement à Mastromonaco (1999), Thomas (2002) précise que la réalisation de la CL, afin d'être considérée comme une occurrence de liaison valable, doit être prononcée conformément à la cible. Des formes phonétiquement imprécises (cf. Paragraphe 1.4) de la CL sont ainsi exclues du total des réalisations. Quant au traitement des CL non-enchaînées, plusieurs études ne traitent pas les liaisons non-enchaînées comme un phénomène indépendant et intègrent ces liaisons au taux total des liaisons réalisées. En revanche, De Moras (2011) ne considère pas ces

mêmes formes comme des réalisations de la liaison mais plutôt comme de simples prononciations de la consonne finale du M1.

Il nous semble donc difficile de comparer ces études et surtout d'en tirer des généralisations sans essayer de garantir une comparabilité partielle des contextes analysés tout au moins sur la base des catégories morphosyntaxiques observées.

1.3. Contextes de liaison étudiés

Comme nous l'avons souligné dans les paragraphes précédents, les spécificités des protocoles d'enquête de chaque étude ont conduit les chercheurs à observer la production de liaison dans un nombre différent de contextes. Nous avons tenté ici (cf. Tableau 27) de procéder à une comparaison directe des pourcentages de réalisation de la liaison par classe morphosyntaxique des M1 et M2, à savoir par type de liaison. Cependant, dans certains cas cette démarche ne suffit pas à garantir une étude comparative des résultats puisque l'analyse des mots de liaison (M1 et M2) en catégories morphosyntaxiques varie également d'une étude à l'autre. Par exemple, dans certains travaux, comme celui mené par Mastromonaco (1999), aucune différence n'est faite entre les adverbes ou les prépositions monosyllabiques et polysyllabiques ou encore les cas où certains éléments ont été enlevés de la classe des adverbes monosyllabiques (ex. « pas » dans l'étude de Thomas 2002). Il en va de même pour la classe des adjectifs qui, dans certains travaux, regroupe non seulement les adjectifs épithètes, mais également tous les autres adjectifs (numéraux, indéfinis, possessifs etc.) qui adoptent des comportements tout à fait différents face à la liaison (cf. Partie 1.Chapitre 4.3.2.3.2). C'est le cas de l'étude de Thomas (2002).

Type de liaison	**Mastromonaco 1999**			**Thomas 2002**			**Howard 2005**	**De Moras 2011**	**Harnois - Delpiano et al.2012**	**Howard 2013**
	Liaison potentielle	LR	**%**	Liaison potentielle	LR	**%**	**Interv. %[207]**	**Interv. %[208]**	**Interv. %[209]**	**Interv. %[210]**
Adjectifs épithètes - M2	31	28	**90%**	530	264	**50%**	**0% - 50%**	**48,8% - 58,1%**	**23,3% - 31 %**	**

[207] Les indices de réalisation de cette étude ont été calculés en prenant en compte la valeur de pourcentage minimale et maximale de réalisations produites par les 12 apprenants dans chaque contexte de liaison.

[208] Cette valeur a été calculée en prenant en compte les pourcentages minimum et maximum de réalisation T1, T2 et T3 des apprenants anglophones (De Moras 2011, p. 410).

[209] Dans le cas de l'étude de Harnois Delpiano et al. (2012), nous avons indiqué le pourcentage minimal et maximal de réalisation de la liaison des trois temps d'observation (T1, T2 et T3) de l'étude longitudinale.

[210] Nous donnons ici le pourcentage minimal et maximal de réalisation de la liaison des trois temps d'observation (T1, T2 et T3) de l'ensemble des locuteurs.

Adverbes monosyll. - M2	27	21	**78%**	636	546	**86%**	**0%-100%**	–	–	**
Adverbes polysyll. - M2	*	*	*	–	–	–	**0%**	–	–	**
Déterminants - M2	187 [211]	176	**94%**	3026	2891	**96%** [212]	**80%-100%**[213]	**50%-91%**	**52,7%-70%**[214]	**56%-100%**
Conjonctions - M2	17	2	**12%**	64	26	**41%**	**0%-37%**	–	–	–
Expressions figées	17	17	**100%**	–	–	–	**40%-100%**	**62,6%-72,1%**	–	–
Noms pluriel - M2	135	9	**7%**	–	–	–	–	**50, 5%-68,5%**	–	**
Prépositions monosyll. - M2	23	24	**96%**	1418	1362	**96%**	**73%-100%**	**69,4%-83,8%**[215]	–	**88%-100%**
Prépositions polysyll. - M2	*	*	*	–	–	–	**0%-100%**	*	–	**
Pronoms personnels - M2	201	187	**93%**	1410	1350	**96%**	**%**	–	–	**79-100%**
Verbes - M2	266	64	**24%**	**	**	**	**0%-5%**	–	–	(« Être ») **0-63%**

- Contexte non étudié
*Aucune distinction entre monosyllabiques et polysyllabiques
** Faible nombre d'occurrences

Tableau 27 : Comparaison des taux de réalisation de la liaison par type de liaison

Les résultats de cette étude comparative des travaux menés sur l'acquisition de la liaison en L2 sont bien évidemment très partiels à cause des raisons que nous venons d'évoquer. Les considérations faites jusqu'ici nous invitent à la prudence quant aux généralisations que nous pouvons tirer de cette comparaison. Néanmoins, il est possible de dégager quelques tendances. Les apprenants adultes en milieu d'acquisition principalement hétéroglotte semblent maîtriser, dès le niveau intermédiaire, deux contextes de liaison (universellement reconnus comme catégoriques), à savoir la liaison après article défini et indéfini et la liaison après pronom personnel sujet. En revanche, les réalisations de liaison à l'intérieur des expressions figées sont plus sujettes à variabilité, sans doute en raison de l'absence d'un contact direct avec la langue cible. En effet, l'étude de Mastromonaco (1999) menée dans un contexte d'apprentissage bilingue, est la seule à présenter 100% de réalisations de la liaison dans ce type de liaison.

[211] Dans l'étude de Mastromonaco (1999, p. 145), cette classe correspond aux contexes de liaison entre « déterminant - adjectif » ; « déterminant - nom » et « déterminant - pronom ».
[212] Adjectifs numéraux exclus (Thomas 2002, p.4).
[213] Les adjectifs numéraux, possessifs, démonstratifs et indéfinis ne sont pas analysés par Howard (2005).
[214] L'étude de Harnois Delpiano et al. (2012) ne prend en compte que les articles et les adjectifs numéraux.
[215] Cette valeur résulte des contextes « prép+art+nom » et « prép+nom » (De Moras 2011, p. 410).

En ce qui concerne les résultats liés aux contextes variables (selon la classification de Durand et al. 2011), les liaisons après les prépositions « dans » et « en » (Howard 2005) et l' adverbe « très » (Howard 2005) ou encore après les formes verbales « est », « suis » et « sont » (Thomas 2002) présentent des pourcentages de réalisation plus élevés que les liaisons après M1 polysyllabiques.

Hormis ces aspects, il nous paraît difficile de formuler des conclusions, surtout si l'on considère que certains types sont peu répertoriés par l'ensemble des études. C'est le cas par exemple de la liaison après noms pluriels observée uniquement dans les études de Mastromonaco (1999) et De Moras (2011). Les écarts importants entre les résultats de ces deux études (7% - 68,5%) pourraient être la conséquence de la diversité des tâches adoptées par ces travaux dans la mesure où De Moras (2011) est la seule à utiliser exclusivement des lectures.

De même, les pourcentages élevés de réalisations après les adverbes et les prépositions monosyllabiques pourraient s'expliquer par un effet de fréquence selon Howard (2013). Cet auteur signale que les apprenants ainsi que les locuteurs natifs réalisent davantage de liaison avec certains mots dans certaines catégories morphosyntaxiques. Autrement dit les apprenants, tout comme les locuteurs natifs, ne produisent pas la liaison avec tous les mots au sein d'une catégorie morphosyntaxique[216].

Le faible nombre de types de liaison variable comparables entre les différents travaux ainsi que l'attention presque exclusive accordée à l'acquisition chez des apprenants de niveau avancé ne permettent pas de formuler des généralisations sur l'acquisition (cf. Paragraphe 3) de la liaison. Les travaux issus du projet « InterPhonologie du Français Contemporain » (IPFC) (cf. Partie 4.Chapitre 1.1) (Detey et Kawaguchi 2008 ; Racine et al. 2010 ; Racine, Detey et Kawaguchi 2012) que nous nous attachons à présenter dans les pages qui suivent (cf. Paragraphe 2) semblent répondre aux besoins de disposer d'un corpus diversifié de français L2 multilingue (plusieurs langues sources) pour l'étude de la liaison. Toutefois, en dépit de leur hétérogénéité, les études évoquées ci-dessus ont le mérite d'avoir établi une première typologie des erreurs de liaison.

[216] « […] si les apprenants réalisent la liaison variable dans certaines catégories morphosyntaxiques ils ne la réalisent qu'avec quelques mots au sein de ces catégories, et pas avec d'autres. On observerait donc un effet lexical sur la réalisation de la liaison variable chez les apprenants dans ces catégories morphosyntaxiques, impliquant que la liaison variable relèverait de facteurs à la fois lexicaux et morphosyntaxiques dans les lectes d'apprenants ». (Howard 2013, p. 225). Cependant, les résultats de l'étude de Howard (2013) indiquent également que la liaison variable n'est pas réalisée dans certaines catégories morphosyntaxiques bien qu'il s'agisse de types plus ou moins fréquents chez les natifs C'est le cas par exemple de la réalisation de la liaison après « quand » ou encore après «jamais » et « rien ».

1.4. Typologies d'erreurs de liaison

Les difficultés associées à l'acquisition de la liaison, évoquées par les études précédentes, se situent d'une part au « niveau de la macro-planification » (Racine 2014) impliquant la maîtrise ainsi que la connaissance de la variation sociolinguistique de la liaison et d'autre part au « niveau de micro-planification » (Racine 2014) qui comprend les aspects phonético-phonologiques et prosodiques de réalisation de la liaison. Nous proposons ci-dessous un bilan des erreurs de liaison relevées en production par les études précédentes :

a. Non-réalisations de liaisons catégoriques : (ex. « trois ami » [tʀwami]) (Harnois-Delpiano et al. 2012). Il s'agit de cas de non-réalisation de la CL dans des types de liaison catégorique. Il est évident que le pourcentage de ces formes varie en fonction de la classification adoptée par les études ;
b. Réalisations phonétiquement non conformes à la CL cible : les cas de prononciations orthographiques de la CL (ex. « un grand ami » [œ̃gʀɑ̃dami]) (Harnois-Delpiano et al. 2012). Cette catégorie d'erreurs ne dépend généralement pas de facteurs morphologiques mais plutôt orthographiques et comporte l'insertion d'une consonne qui ne correspond pas à une CL existante. En effet, ce type d'erreurs est souvent le résultat de l'influence des valeurs phonographiques des consonnes graphiques finales du M1 ;
c. Substitutions avec une autre CL : (ex. « trois amis » [tʀwanami]) (Harnois-Delpiano et al. 2012). Dans ce cas, les apprenants substituent la CL cible par une autre consonne (cf. Partie 1.Chapitre 4.2.4.3) ;
d. Insertions d'une CL non attendue après un M1 se terminant par voyelle : (ex. « quatre enfants » [katʀzɑ̃fɑ̃]) (cf. Partie 1.Chapitre 4.2.4.3). Il s'agit d'une erreur d'ordre morphologique qui consiste à insérer une CL correspondant à la marque du pluriel ;
e. Insertions d'une CL non attendue devant un mot à initiale consonantique : (ex. « les garçons » [lezgaʀsɔ̃]) (Mastromonaco 1999). Ce type de formes est la conséquence d'une généralisation morpho-phonologique du phénomène de la liaison qui consiste à prononcer une consonne finale devant un mot pluriel commençant par consonne ;
f. Liaisons interdites : (ex. « les homards » [lezɔmaʀ]). Contrairement au cas précédent, cette erreur peut être expliquée par une généralisation phonologique du phénomène de la liaison mais également par l'absence d'un apprentissage explicite des règles qui interdisent la réalisation de la liaison dans certains contextes phonologiques et syntaxiques (Mastromonaco 1999) ;

g. Liaisons non enchainées : (ex. « un homme » [œ̃n ɔm]) (Thomas 2002). Dans ce cas, la CL est suivie d'une pause ou d'une hésitation, ce qui empêche sa resyllabation avec la première syllabe du M2. La CL constitue ainsi une coda consonantique de la dernière syllabe du M1 ;
h. Création : (ex. « des boucles [z] oreilles » pour « des boucles d'oreilles ») (Mastromonaco 1999). Dans ce cas, les apprenants créent un mot qui n'existe pas en L2, avec lequel ils réalisent la liaison ;
i. Interférences : (ex. « las [z] autorités » (L1: Espagnol)) (Mastromonaco 1999). Contrairement à la catégorie précédente, les apprenants emploient dans ce cas un lexème qui est le résultat d'un transfert de la L1 ;
j. Hésitations avec pause remplie : (ex. « les [z] um, et uh » (Mastromonaco 1999). Les locuteurs réalisent une liaison avec l'hésitation suivante.

Les résultats de l'étude menée par Thomas (2002) montrent que 20% des productions réalisées par les apprenants sont erronées et consistent essentiellement en : non-réalisations (a), réalisations phonétiquement non conformes à la CL cible (b) (comme forme phonétique correspondant à la représentation graphique de la CL) et surtout non-enchaînements (8,5%) (g). Ce dernier patron est souvent attesté chez les apprenants anglophones comme le démontrent également les résultats de l'étude menée par Mastromonaco (1999) (7%) et est en revanche absent dans l'étude de Harnois-Delpiano (2012) menée auprès d'apprenants coréens.

Les résultats de l'étude de Harnois-Delpiano et al. (2012) montrent aussi la présence massive de non-réalisations en contexte de liaison variable (41,3%) et également catégorique (35,8%) ainsi qu'une influence probable des valeurs phonographiques des consonnes graphiques (b). De même, Thomas (2002) souligne que les étudiants font des erreurs (b), en particulier quand la CL ne correspond pas à la consonne graphique (Thomas 2002, 106). Cela détermine la production de prononciations orthographiques de la CL (b), attestées dans toutes les études.

En revanche, le taux d'erreurs de substitution (c) est assez réduit en acquisition en L2. Ces formes sont attestées uniquement dans les études de Mastromonaco (1999) et de Harnois-Delpiano et al. (2012) (2% dans le cas de liaisons variables et 6,9% dans le cas des liaisons catégoriques).

Les formes d'insertion d'une CL non attendue après un mot à finale vocalique (d) sont totalement absentes. Cela pourrait s'expliquer par le fait que ces erreurs (produites le plus souvent par les natifs cf. Partie 1.Chapitre 4.2.4.3) impliquent une catégorie morphologique

telle que le pluriel, qui n'est pas encore suffisamment enracinée chez les apprenants (Mastromonaco 1999).

Les erreurs d'insertion d'une CL non attendue devant un mot à initiale consonantique (e) et de formes de liaison interdites (f) sont en revanche plus souvent attestées chez les apprenants en tant que résultats de généralisations morpho-phonologique et phonologique du phénomène de la liaison.

Quant aux liaisons interdites (f) (Mastromonaco 1999, p. 189), les apprenants semblent en réaliser un nombre infime, le plus souvent en parole spontanée.

Pour ce qui est des liaisons non-enchainées (g), toutes les études enregistrent des formes de liaison non-enchaînée. D'après Thomas (2002) les productions de non-enchaînement en L2 n'ont pas la valeur stylistique qu'on retrouve chez les hommes politiques français (Encrevé 1988), mais qu'elles relèvent d'une faible maîtrise du processus de resyllabation de la CL[217].

> « […] la liaison sans enchaînement reflète généralement chez nos étudiants une hésitation devant un mot difficile (ses_ennemis) ou plus simplement une influence de la syllabation fermée caractéristique de leur langue maternelle, l'anglais ». (Thomas 2002)

La L1 des apprenants semblerait donc jouer un rôle important dans le processus de resyllabation de la CL. Selon cette hypothèse l'enchaînement de la CL représente une plus grande difficulté pour les apprenants ayant comme L1 une langue à syllabation fermée (comme l'anglais) à savoir une langue où la congruence entre frontière lexicale et syntaxique est relativement forte (Wauquier 2009) (cf. Paragraphe 4).

En revanche, les erreurs de création (h) et d'interférence (i) sont des erreurs spécifiques à l'interlangue des apprenants (cf. Partie 3.Chapitre 1.2) attestées dans l'étude Mastromonaco (1999) et donc sujettes à une grande variabilité selon le niveau linguistique et la L1 des apprenants.

Pour résumer, les études évoquées dans le paragraphe précédent (cf. Paragraphe 1) nous permettent d'observer une production majoritaire d'erreurs de non-réalisations de la CL (a), des prononciations orthographiques (b) ou encore des formes de liaison non-enchaînées. Les erreurs de substitution de CL (c), caractéristiques de l'acquisition en L1 (cf. Partie 1.Chapitre 5.3.5), ainsi que les formes d'insertion des CL inattendus après M1 se termianant par voyelle (d) (observées chez les adultes natifs cf. Desrochers 1994) sembleraient être absentes chez les apprenants de français L2.

[217] Il propose une intervention pédagogique visant à travailler de façon plus générale sur la syllabation ouverte du français, et à traiter ensuite la liaison comme un cas particulier de resyllabation.

Ces tendances seront vérifiées dans les pages qui suivent par les études issues du projet « InterPhonologie du Français Contemporain » (IPFC) (Detey et Kawaguchi 2008; Racine et al. 2010; Racine, Detey et Kawaguchi 2012) qui assurent, à l'aide d'un protocole et d'une classification de référence communs, une plus large comparaison entre les réalisations et les erreurs produites par des populations différentes d'apprenants.

2. Études issues du projet InterPhonologie du Français Contemporain (IPFC)

Le projet « InterPhonologie du Français Contemporain » (IPFC) (Detey et Kawaguchi 2008 ; Racine et al. 2010 ; Racine, Detey et Kawaguchi 2012)[218] (cf.Partie 4.Chapitre 1.1) piloté par Sylvain Detey (Université Waseda), Isabelle Racine (Université de Genève) et Yuji Kawaguchi (Tokyo University of Foreign Studies) a pour objectif principal la constitution d'une importante ressource de français L2 à visée phonologique pour l'étude des principaux phénomènes du français, au sein desquels un intérêt particulier est porté à l'acquisition de la liaison en L2.

Une partie des travaux consacrés à ce sujet a été rassemblée dans le bulletin suisse de linguistique appliquée (Bulletin Vals-Alsa) (Racine et Detey sous presse) et correspond à des recherches menées auprès de sept différentes populations d′apprenants : anglophones (Tennant sous presse), germanophones (Pustka sous presse), hellénophones (Valetopoulos sous presse), hispanophones (Racine sous presse), italophones (Falbo et al. sous presse), japonophones (Detey, Kawaguchi et Kondo sous presse) et norvégophones (Andreassen et Lyche sous presse). Le tableau ci-dessous (cf. Tableau 28) résume les caractéristiques principales de ces travaux.

Étude	Apprenants	Âge	Niveau	Contexte	Modalité d'acquisition	Tâches	Classification
Tennant (sous presse)	12 anglophones	19 - 24 ans	intermédiaire / avancé[219]	hétéroglotte	guidée	lecture ; entretien guidé ; conversation libre	Durand et al. 2011
Pustka (sous presse)	12 germanophones	26-32 ans	très avancé[220]	hétéroglotte	guidée	lecture ; entretien guidé ; conversation libre	Durand et al. 2011

[218] (http://cblle.tufs.ac.jp/ipfc/).

[219] L'auteur distingue trois groupes d'étudiants anglophone canadiens (Ontario) : groupe A (étudiants en première année de licence ayant suivi un programme d'immersion ou ayant fait leur scolarité en français (niveau B1) ; groupe B (étudiants en première année de licence ayant étudié le français dès l'âge de 9 ans) (programme *Core French*) (niveau A2) ; groupe C (étudiants en quatrième année ayant fait une expérience d'étude en France) (niveau C1) ; groupe D (étudiants en quatrième année n'ayant pas fait d'expérience d'étude en France) (niveau B2).

[220] Il s'agit de futurs professeurs de français.

Valetopoulos (sous presse)	24 hellénophones	18-20 ans	intermédiaire / avancé[221]	hétéroglotte	guidée	lecture[222]	Durand et al. 2011
Racine (sous presse)	20 hispanophones	26-27 ans[223]	avancé[224]	hétéroglotte	guidée	lecture ; entretien guidé ; conversation libre	Durand et al. 2011
Falbo et al. (sous presse)	12 italophones (Milan)	19-22 ans	faux-débutants et débutants	hétéroglotte	guidée	lecture ; entretien guidé	Durand et al. 2011
	12 italophones (Trieste)	19-22 ans	avancé	hétéroglotte	guidée	lecture ; entretien guidé	Durand et al. 2011
Detey, Kawaguchi et Kondo (sous presse)	17 japonophones[225]	20 -24 ans[226]	avancé	hétéroglotte	guidée	lecture	Durand et al. 2011
Andreassen et Lyche (sous presse)	12 norvégophones[227]	17-24 ans	intermédiaire / avancé[228]	hétéroglotte	guidée	lecture ; entretien guidé ; conversation libre	Durand et al. 2011

Tableau 28 : Études sur l'acquisition de la liaison en L2 issues du projet IPFC

L'utilisation d'une méthodologie commune (cf. Partie 4.Chapitre 1.2.3) (protocole de recueil de données, conventions de transcription, codage et outil de décodage) a permis de réaliser des comparaisons à grande échelle tant du point de vue de la population (en fonction de la L1, du niveau des apprenants ou du contexte d'apprentissage) que de la tâche réalisée par les apprenants (lecture *versus* parole semi-spontanée). De plus, la proximité avec le protocole du projet PFC (cf. Partie 2.Chapitre 1) a également permis une comparaison avec les natifs en évitant de faire référence à des données datées ou marquées d'un point de vue sociolinguistique.

[221] Les apprenants appartiennent à deux groupes d'étudiants différents ayant tous suivi des cours de français pendant les trois premières ou les trois dernières années du secondaire : 7 étudiants inscrits en première année de licence et 17 en quatrième année (niveau B2-C1).

[222] Le corpus IPFC-grec (cypriote) est encore en cours d'analyse ce qui explique l'absence des résultats des autres tâches prévues par le protocole IPFC.

[223] Âge moyen des deux groupes d'apprenants (cf. Note 224).

[224] Les apprenants de niveau B2-C1 étaient répartis en deux groupes : 10 avec séjour prolongé en milieu francophone (AS) et 10 apprenants sans séjour (SS).

[225] Les apprenants étaient répartis en deux groupes d'étudiants inscrits en 3ème année de Licence ou 1ère année de Master: 7 étudiants avec séjour prolongé en milieu francophone (AS) et 10 apprenants sans séjour (SS).

[226] Âges moyens des deux groupes d'apprenants.

[227] Les apprenants appartiennent à deux groupes : 6 étudiants en première année de licence à l'Université d'Oslo (niveau B1/B2) ayant réalisé un séjour dans un pays francophones et 6 lycéens en dernière année dans un lycée de Tromsø (niveau A2) n'ayant pas effectué de séjour. Dans tous les autres travaux, les sujets étaient des étudiants.

[228] Cf. Note 227.

2.1. Production de la liaison

Tous les travaux évoqués ci-dessus constituent des études transversales qui ont souvent adopté une approche sociolinguistique visant à étudier les différences de production de la liaison entre groupes d'apprenant sur la base de facteurs tels que le nombre d'années d'études secondaires du français (Tennant sous presse), la réalisation d'un séjour dans un pays francophone (Racine sous presse ; Detey, Kawaguchi et Kondo sous presse ; Andreassen et Lyche sous presse) et le niveau des apprenants (Valetopoulos sous presse ; Falbo et al. sous presse) (cf. Tableau 29).

Étude	**% LR liaison catégorique et variable**	
	Lecture	**Parole**
Tennant (sous presse)	39,6%	45,1% - 48,9%
Pustka (sous presse)	51%	-
Valetopoulos (sous presse)	48,74%	-
Racine (sous presse)	SS = 56,67% AS = 57,38%	SS = 41,77% AS = 33,95%
Falbo et al. (sous presse) (Milan)	41,36%	42,88%
Falbo et al. (sous presse) (Trieste)	46,13%	31,96%
Detey, Kawaguchi et Kondo (sous presse)	SS = 44.63% AS = 47.93%	-
Andreassen et Lyche (sous presse)	Tromsø 8% Oslo 50%	Tromsø 20% - 19% Oslo 44% - 51%

Tableau 29 : Comparaison entre les taux de réalisation (catégorique et variable) par tâche (lecture et parole) extraits des études du projet IPFC

L'étude menée par Tennant (sous presse) montre que les apprenants de niveau plus avancé produisent un nombre plus important de liaisons catégoriques[229]. Toutefois ce facteur externe ne semble pas conditionner la production des liaisons dans les contextes catégoriques après

[229] Pour les contextes de liaison étudiés, voir les sites de liaison contenus dans le texte PFC également adopté par le protocole du projet IPFC (cf. Partie 4.Chapitre 1.4.3.1.4).

pronoms personnels sujets (94%) et après déterminant (86,3%). En revanche la liaison au sein des expressions figées est caractérisée par une forte variabilité (58,3% - 93,1%) qui n'est toutefois pas explicable sur la base du niveau linguistique des apprenants. Detey, Kawaguchi et Kondo (sous presse) montrent également que la réalisation d'un séjour n'influence pas de façon significative la production des liaisons catégoriques. En effet si les apprenants japonophones sans séjour (dorénavant SS) produisent 75,28% des liaisons catégoriques, leurs homologues ayant effectué un séjour (dorénavant AS) réalisent 86,03% des liaisons catégoriques, une différence qui n'est pas significative. De même, ni le niveau des apprenants dans l'étude de Falbo et al. (sous presse) ni le type de tâche n'influencent le taux de production des liaisons catégoriques qui sont dans l'ensemble maîtrisées autant par les apprenants débutants et faux débutants de Milan que par ceux de niveau avancé de Trieste.

En revanche, les résultats de l'étude menée par Racine (sous presse) et par Detey, Kawaguchi et Kondo (sous presse) montrent que de manière générale seuls les apprenants AS produisent un nombre inférieur de liaison par rapport aux natifs[230] et uniquement dans la tâche de conversation. Racine (sous presse) souligne en effet que, comme dans le cas d'autres aspects sociolinguistiques (ex. l'adverbe de négation « ne »), « le séjour permet aux apprenants d′acquérir des formes moins formelles qu′ils ont tendance à sur-généraliser en conversation, alors que le contexte classe de FLE favorise les formes formelles ».

Toutefois, les apprenants AS semblent avoir développé des compétences sociolinguistiques puisque, à l′instar des natifs, ils produisent un taux moins important de liaisons en conversation qu′en lecture. Cet aspect est également mis en évidence chez des apprenants avancés SS, comme c'est le cas des sujets de l'étude IPFC-italien (Trieste) (Falbo et al. sous presse).

2.2. Typologies d'erreurs de liaison

Les tendances qui caractérisent les productions d'erreurs observées par les études précédentes semblent être confirmées par les études systématiques conduites au sein du projet IPFC. La plupart des erreurs correspondent à des réalisations phonétiquement non conformes à la CL cible (b) (cf. Paragraphe 1.4) (ex. des formes phonétiques correspondant à la représentation graphique de la CL). En effet toutes les études enregistrent des productions orthographhiques de la CL en particulier chez des apprenants italophones (Falbo et al. sous presse) et hispanophones (Racine sous presse) de FLE dont la L1 (italien et espagnol) est

[230] Les taux de réalisation des liaisons produites par les natifs ont été extraits des lectures et des conversations de dix locuteurs suisses romands enregistrés dans le cadre du projet PFC (cf. Partie 2.Chapitre 1).

caractérisée par une correspondance transparente entre phonie et graphie. En revanche, un nombre moins important de réalisations orthographiques des CL a été attesté chez les apprenants hellénophones (Valetopoulos sous presse). Quant aux apprenants japonophones, Detey, Kawaguchi et Kondo (sous presse) observent un nombre important de cas (67% des cas totaux) de réalisations d'une CL orthographique [n].De plus, toutes les études enregistrent la présence de liaisons non-enchaînées (g) (cf. Paragraphe 1.4). Le tableau ci-dessous (cf. Tableau 30) résume les taux de productions de ces formes attestées dans les différents travaux.

Étude	**% LR liaison non-enchaînée**	
	Lecture	**Conversation**
Tennant (sous presse)	12,9%	4,6%[231] ; 6%[232]
Pustka (sous presse)	1%[233]	
Valetopoulos (sous presse)	11,2% - 11,3%[234]	_
Racine (sous presse)	SS = 12,50 AS = 14,76%[235]	SS = 7,78% AS = 2,6%
Falbo et al. (sous presse) (Milan)	17,05%	6,68%
Falbo et al. (sous presse) (Trieste)	0,52%	0,19%
Detey, Kawaguchi et Kondo (sous presse)	SS = 6.93% ; AS = 1.68%	
Andreassen et Lyche (sous presse)	Tromsø 37% Oslo 7%	Tromsø 17% ; 10% Oslo 2% ; 7% [236];

Tableau 30 : Comparaison entre les taux de réalisation des liaisons non-enchaînées extraits des études du projet IPFC

[231] Dans l'entretien guidé.

[232] Dans la conversation libre.

[233] Seulement 8 occurrences de liaisons non-enchaînées ont été enregistrées dans cette étude, la plupart des cas relevant de phénomènes d'hésitation.

[234] Il convient de souligner que ce taux important de non-enchaînement chez des apprenants hellénophones pourrait s'expliquer par un choix méthodologique que l'auteur précise dans ses conclusions : « […] une autre question que l′on doit se poser en étudiant nos données concerne la réalisation des voyelles nasales. Dans certains contextes, nous avons codé les liaisons comme étant non enchaînées. Mais il faut revoir éventuellement les productions en s′interrogeant sur l′existence d′une vraie liaison ou une réalisation non nasale de ces voyelles » (Valetopoulos sous presse).

[235] Racine (sous presse) souligne également que les données concernant les réalisations non-enchaînées « […] révèlent en effet qu′un cumul des deux phénomènes, à savoir la réalisation d′une liaison non enchaînée avec une consonne orthographique, est parfois possible, par exemple, dans le texte, 'grand honneur' [gʁɑ̃n-ɔ-noeʁ], 'ont eu' [ɔ̃n-y] et, en conversation, 'sont un peu' [sɔ̃n-ɛ̃-pø] et 'quand on' [kɑ̃n-ɔ̃] […] il paraît toutefois difficile, dans ce cas – et ce même s′il s′agit d′apprenants avancés – de déterminer s′il s′agit réellement de liaisons non enchaînées ou plutôt de la prononciation d′une consonne graphique en finale de mot ». (Racine sous presse)

[236] Les premiers taux de chaque groupe (corpus Tromsø ; corpus Oslo) font référence aux productions des liaisons non-enchaînées issues des entretiens guidés et les deuxièmes à celles réalisées dans les conversations libres.

Comme nous l'avons déjà observé (cf. Paragraphe 1.4), ce patron d'erreur a souvent été attesté chez les apprenants anglophones (Mastromonaco 1999 ; Thomas 2002). Les résultats de l'étude de Tennant (sous presse) montrent un taux de liaisons non-enchainées (7,2%) comparable à celui enregistré chez d'autres apprenants anglophones canadiens (7%) (Mastromonaco 1999 ; Thomas 2002). Toutefois, les résultats de l'étude menée par Tennant (sous presse) indiquent que les apprenants de niveau plus avancé produisent un taux de non-enchaînements moins important. L'étude de Pustka (sous presse) montre que ce type d'erreur disparaît chez les apprenants les plus avancés.

De plus, ce type d'erreur de « micro-planification » (Racine 2014) semble être plus fréquent dans des contextes de liaison erratique et variable (Tennant sous presse ; Valetopoulos sous presse).

À la suite de l'hypothèse formulée par Thomas (2002) et ensuite reprise par Wauquier (2009), Tennant considère les non-enchaînements des liaisons comme le résultat de l'influence de la syllabation fermée caractéristique de la L1 des apprenants anglophones qui constituerait un obstacle à la correcte resyllabation de la CL.

> « It should be recalled that, in L1 French, liaison without enchaînement is an infrequently occurring characteristic of certain public speaking styles (Encrevé 1988), and it is found to be almost non-existent in the PFC corpus (Durand & Lyche 2008). In addition, in native speech, its occurrence is limited to contexts of variable (optional) liaison. While, in the absence of other evidence, the higher rate of non-enchaînement in variable liaisons compared to categorical liaisons might be interpreted as reflecting L2 speakers′ knowledge of these L1 constraints on the phenomenon, an explanation in terms of learner errors would be more plausible. As noted in the literature review, lack of enchaînement of a liaison consonant by an L2 speaker of French can be interpreted as an error relating to the open syllable structure of French generally, which contrasts with the closed syllable pattern of Anglophone learners′ L1 ». (Tennant sous presse)

Néanmoins, la présence de non-enchaînement attestée chez des apprenants ayant des L1 caractérisées par un système syllabique ouvert tel que l'italien (Falbo et al. sous presse) et l'espagnol (Racine sous presse) nuance l'hypothèse évoquée ci-dessus. Plusieurs auteurs (Tennant sous presse ; Racine sous presse) n'excluent pas la possibilité que les liaisons non-enchaînées constituent les traces « de la perméabilité des compétences et connaissances phonologique et orthographique » (Durand et al. 2011) (cf. Partie 1.Chapitre 2.2.1.1).

Cette hypothèse serait appuyée par l'observation d'une plus grande présence de liaisons non-enchaînées en lecture (Racine sous presse ; Tennant sous presse ; Andressen et Lyche sous

presse) qu'en parole semi-spontanée[237]. Racine (sous presse) souligne que « le séjour paraît toutefois avoir un effet bénéfique pour la conversation, puisque le taux de liaisons réalisées sans enchaînement diminue sensiblement chez les apprenants avec séjour par rapport à ceux sans séjour ». Cependant cette tendance n'est pas vérifiée en lecture.

L'étude Andreassen et Lyche (sous presse) propose deux explications différents pour justifier la présence de liaisons non-enchaînées chez des apprenants norvégophones. D'une part les auteurs avancent l'hypothèse que les apprenants sous l'influence du système prosodique de leur L1, tendraient à maintenir un accent lexical contrastif[238] dans leur interlangue en s'opposant au processus de resyllabation déclenché par l'enchaînement de la CL. D'autre part cet aspect serait aggravé par le mode d′apprentissage basé sur l′écrit adopté par la plupart des manuels d'enseignement. En outre, Andreassen et Lyche (sous presse) ont relevé un taux important de liaisons non-enchaînées chez les lycéens ayant un niveau A2, notamment en lecture (23,5%). Ce résultat indique encore une fois l'influence possible de la forme écrite dans la production de ce type de liaison (Racine sous presse ; Tennant sous presse). Néanmoins cette différence inter-tâche n'est pas observée chez les apprenants norvégophones de niveau supérieur (B1-B2) qui produisent presque le même taux de non-enchaînements tant en lecture (7%) que dans les entretiens guidés (2%) et les conversations libres (7%). Il nous paraît donc qu'on ne saurait pas exclure l'hypothèse de l'influence du système prosodique (Andreassen et Lyche sous presse) et syllabique (Thomas 2002 ; Wauquier 2009 ; Tennant sous presse) lors de l'analyse des réalisations des liaisons non-enchaînées.

Quant aux erreurs de substitutions (c), les études issues du projet IPFC confirment le taux marginal de production de ces formes. Tennant (sous presse) observe un faible nombre d'erreurs de substitution (c) (ex. « trois amis » [tʀwanami]) et une absence totale des CL non attendues après M1 se terminant par voyelle (d) (ex. « quatre enfants » [katʀzɑ̃fɑ̃]) et devant M2 à initiale consonantique (e) (ex. « les garçons » [lezgaʀsɔ̃]) (cf. Paragraphe 1.4). Toutefois, il est intéressant d'observer que les formes de substitutions sont le plus souvent produites par des apprenants de niveau avancé. Falbo et al. (sous presse) enregistre également des formes de substitution de la CL (c) (ex. « sont [z] étrangers») chez des apprenants italophones avancés de Trieste. De façon plus générale, ces erreurs sont davantage attestées dans des contextes de liaison variable et erratique (comme dans le cas des liaisons non-enchainées) et en lecture plutôt

[237] Comme le fait observer Tennant (sous presse), cette tendance est également attestée chez les natifs (cf. Partie 1.Chapitre 2.2.1.1) qui produisent des liaisons non-enchaînées de façon marginale en dehors des discours publics (Durand et Lyche 2008 ; Durand et al. 2011).

[238] En tant que trait moins marqué (Eckman 1977) (cf. Partie 3.Chapitre 1.3.1.2) par rapport à l'accent de groupe caractéristique de la prosodie du français (cf. 0).

qu'en parole semi-spontanée. Tennant (sous presse) attribue cette dernière tendance à la difficulté du texte présenté aux sujets.

En revanche, l'étude de Falbo et al. (sous presse) menée auprès d'apprenants italophones avancés de Trieste est la seule étude à relever la présence de liaisons non attendues après M1 se terminant par voyelle (d) (ex. « madame [z] et messieurs »). Ce type d'erreur, absente dans les études précédentes et le plus souvent produite par les natifs, semble suggérer d'après Falbo et al. une généralisation morphosyntaxique de la liaison dans les contextes catégoriques mise en place par les apprenants avancés. Des liaisons non attendues devant M2 à initiale consonantique (e) (ex. « nous sommes [z] rentrés») sont également présentes dans les productions des apprenants italophones (Falbo et al. sous presse) alors que chez les hispanophones (Racine sous presse), des formes telles que « son [n] séjour » pourraient également être interprétées comme des erreurs de prononciation (cf. Note 234).

Quant aux liaisons interdites (f) (ex. « les homards » [lezɔmaʀ]), le nombre réalisé par les apprenants est infime et, contrairement aux études précédentes, a plus souvent lieu dans la lecture que dans la parole semi-spontanée. Comme le soulignent Detey, Kawaguchi et Kondo (sous presse), le nombre de liaisons interdites diminue chez les apprenants AS jusqu'à disparaître chez les apprenants les plus avancés (Pustka sous presse).

3. *Modèles développementaux d'acquisition de la liaison en L2*

Contrairement à l'acquisition en L1 (cf. Partie 1.Chapitre 5.3.5), aucun scénario n'a été élaboré à ce jour quant à l'acquisition de la liaison en L2. L'absence d'une modélisation du processus acquisitionnel est probablement liée d'une part à la faible quantité de données disponibles issues de recherches de nature hétérogène (perceptive, sociolinguistique etc.) et d'autre part à la forte inter-variabilité des productions des apprenants. Celle-ci ne semblerait pas relever de facteurs linguistiques (niveau linguistique et années d'études) (Thomas 2004 ; Wauquier et Shoemaker 2013). Ces aspects constituent donc un obstacle pour la formulation de généralisations concernant l'évolution des stades d'acquisition.

Néanmoins, l'analyse des typologies d'erreurs de liaison réalisées par les apprenants a amené plusieurs auteurs (Wauquier 2009 ; Harnois-Delpiano, Cavalla et Chevrot 2012 ; Wauquier et Shoemaker 2013) à formuler l'hypothèse selon laquelle les apprenants de français L2 utiliseraient une stratégie d'acquisition de type lexical. Ils considéreraient la liaison comme un phénomène lexical appris par le biais d'une représentation orthographique des mots de

liaison. Wauquier (2009) soutient que les apprenants ne possèdent aucune représentation phonologique de la consonne de liaison et que donc « le contenu segmental et la position prosodique de cette consonne finale attendue sont ceux qui font surface dans la forme orthographique du mot » (Wauquier 2009, p. 111). Selon cette théorie, les apprenants traiteraient la liaison « au cas par cas » selon la représentation orthographique du M1 mémorisée et ne seraient donc pas capables, contrairement aux enfants francophones, d'élaborer des généralisations morpho-phonologiques.

> « […] la prégnance de liaisons non enchaînées sans valeur stylistique particulière en contexte obligatoire invite à considérer que ces apprenants recherchent dans l'input oral les mots segmentés qu'ils ont appris par l'écrit. Ceci entraîne un traitement plutôt au cas par cas, lexicalement contraint ». (Wauquier 2009, p. 122)

En revanche Harnois-Delpiano et al. (2012) semblent être plus prudents dans la formulation des conclusions d'ordre théorique. Ces auteurs soulignent que l'influence de l'input pourrait s'expliquer autant à partir d'un modèle exemplariste (Chevrot et Fayol 2000 ; Chevrot et Fayol 2001 ; Chevrot, Chabanal et Dugua 2007 ; Chevrot et al. 2013 pour la L1 et Wauquier 2009 ; Wauquier et Shoemaker 2013 pour la L2) que d'un modèle fondé sur des principes universels (Wauquier-Gravelines et Braud 2005 ; Splendido 2014 ; Ruvoletto 2014 ; Ruvoletto et Wauquier 2015 pour la L1). De façon générale, ils avancent également l'hypothèse que l'input orthographique pourrait jouer un rôle dans le processus d'acquisition de la liaison en L2, mais, contrairement à Wauquier (2009), ils n'excluent pas que les apprenants puissent opérer des généralisations indépendantes du mot écrit en se basant sur le système général des correspondances phonographiques. En effet, les résultats de l'étude menée par Pustka (sous presse) montrent que les apprenants de niveau très avancé parviennent à acquérir une maîtrise de la liaison comparable à celle des natifs dont le développement peut difficilement être expliqué par la mise en place des seules stratégies lexicales.

L'étude récente menée par Wauquier et Shoemaker (2015) a attesté le recours à des stratégies d'acquisition diversifiées chez des apprenants adultes en contexte d'apprentissage non-guidé et homoglotte. Autrement dit, l'acquisition de la liaison en milieu non institutionnel semble entraîner un traitement lexical de la séquence de liaison (comme le démontrent les non-réalisations de la CL et les liaisons non-enchaînées), mais également la mise en place de stratégies telles que celles utilisés par les natifs en L1 (comme le démontre la présence de substitutions de la CL de nature morpho-phonologique)[239] (cf. Partie 3.Chapitre 2.1.4).

[239] « Voici deux ours » [vwa.si.dø.nuʁs] au lieu de [vwa.si.dø.zuʁs]) (cf. Partie 1.Chapitre 5.3.5).

3.1. Comparaison entre processus d'acquisition en L1 et L2

Bien que des scénarii développementaux n'aient pas encore été établis pour l'acquisition de la liaison en L2, l'analyse des typologies des erreurs de liaison suggère l'existence de spécificités dans le processus d'acquisition de la liaison en L2 qui l'oppose aux modèles développementaux proposés en L1. En effet, malgré les interprétations différentes (cf. Partie 1.Chapitre 5.3.5), les scénarii développementaux proposés pour l'acquisition de la L1 postulent l'existence d'une dynamique d'acquisition articulée en trois étapes où l'enfant après une phase de segmentation aboutit à une généralisation d'ordre lexical ou phonologique. En revanche, d'après certains auteurs (Wauquier 2009 ; Wauquier et Shoemaker 2013), l'acquisition de la liaison en L2 semblerait liée à un processus inverse où l'apprenant serait obligé pour réaliser la liaison d'« attacher » les unités lexicales déjà segmentées.

> « […] les apprenants de L1 acquièrent la liaison en détachant les mots les uns des autres, alors qu'elle pourrait plutôt poser problème aux apprenants de L2 quand ils vont attacher les mots ensemble ». (Wauquier 2009, p. 125)

D'après Wauquier (2009), cette hypothèse serait étayée par l'observation d'erreurs de non-enchaînement sans valeur stylistique particulière en contexte de liaison catégorique chez des apprenants avancés (Mastromonaco 1999 ; Thomas 2002 ; Howard 2005). La persistance de ce type d'erreur ainsi que des formes orthographiques des CL montreraient que les apprenants ne disposent pas d'une représentation phonologique de la liaison. Ils utiliseraient plutôt une stratégie d'apprentissage lexical « au cas par cas », traitant la CL de façon similaire à une consonne fixe[240] et s'appuyant sur une représentation orthographique de surface.

Néanmoins, comme le fait observer Wauquier même (2009 ; 2013) et Tennant (sous presse), les occurrences de liaison non-enchainées attestées dans des stades avancés d'apprentissage sont issues des études menées auprès d'apprenants anglophones. En effet, ce même type d'erreur n'a pas été attesté chez les apprenants coréens de l'étude de Harnois-Delpiano et al. (Harnois-Delpiano, Cavalla et Chevrot 2012), ce qui suggère une éventuelle influence de la L1 sur l'acquisition de la liaison (cf. Partie 3.Chapitre 1.3.1.1). Ces observations semblent validées par une comparaison des taux des liaisons non-enchainées entre les différentes études menées auprès d'apprenants anglophones canadiens (Mastromonaco 1999 ; Thomas 2002 ; Tennant sous presse). Néanmoins, comme nous l'avons vu dans le paragraphe précédent, la présence de non-enchainements attestée chez des apprenants ayant des L1 caractérisées par un système syllabique ouvert tel que l'italien (Falbo et al. sous presse) et

[240] « Pour les apprenants de L2, les CL doivent avoir au départ un statut flou, assez proche de celui des consonnes finales fixes ». (Wauquier 2009, p. 122)

l'espagnol (Racine sous presse) nuance l'hypothèse évoquée par Wauquier (2009). Il est en effet également possible, comme le font observer Durand et al. (2011), Tennant (sous presse) et Racine (sous presse), que les liaisons non-enchaînées constituent le résultat de l'effet Buben (1935)[241] , plusieurs fois évoqué par des études précédentes (Chevrot et Malderez 1999 ; Laks 2005b) (cf. Partie 1.Chapitre 6.3).

Ces considérations nous semblent fondamentales pour ne pas exclure la capacité chez les apprenants de français L2 de développer des généralisations d'ordre phonologique qui aboutiraient à une maîtrise complète de la liaison catégorique. En effet, comme le souligne Harnois-Delpiano et al. (2012), la forme écrite des mots pourrait provoquer un ralentissement des généralisations, mais les apprenants pourraient parvenir à opérer des généralisations indépendantes du mot particulier s'appuyant sur un réseau acquis de correspondances phonographiques.

Une stratégie purement lexicale est également rejetée par les études de Tennant (sous presse) et de Racine (sous presse). Racine (sous presse) suggère en particulier que les deux approches proposées par Wauquier (2009) et Harnois-Delpiano et al. (2011) pourraient être également valables en fonction du type de contact avec la L2 (présence / absence d'un séjour) et de la tâche (lecture et parole semi-spontanée).

> « Ainsi, les deux approches mentionnées précédemment pourraient en fait s′appliquer de manière différenciée en fonction de la tâche et du séjour, avec, au début de l′apprentissage, pour les deux tâches, un schéma tel que le propose Wauquier (2009), à savoir un traitement basé sur une représentation orthographique mémorisée au cas par cas, sans généralisation d′ordre morpho-phonologique, ce qui favorise à la fois les productions avec une consonne orthographique et les liaisons sans enchaînement. Le séjour favoriserait la mise en place d′un processus de généralisation indépendante du mot sur la base de correspondances phonographiques (Harnois et al. 2012), ainsi que le développement de la compétence sociolinguistique, notamment au niveau diaphasique ». (Racine sous presse)

[241] Linguiste russe des années 30 qui a essayé d'expliquer l'évolution phonétique du français en postulant l'existence d'une influence de l'orthographe sur la prononciation ayant agi au niveau diachronique.

4. Le rôle de la L1 dans l'acquisition de la liaison

L'influence de la L1 des apprenants que nous avons brièvement évoquée dans le chapitre précédent (cf. Partie 3.Chapitre 1.3.1.1) semble conditionner tant la perception que la production de la liaison en L2. En particulier, comme nous l'avons vu dans les paragraphes précédents, les études perceptives ont, dans certains cas, montré une difficulté des apprenants à résoudre l'ambiguïté des frontières lexicales déterminée par la resyllabation de la CL. Cet aspect semblerait être d'autant plus accentué chez des apprenants dont la langue source serait une L1 où « la congruence entre les frontières lexicales et les frontières syllabiques est beaucoup plus forte que pour le français » (Wauquier 2009). Cette hypothèse semble être étayée par les résultats des études perceptives menées auprès des apprenants anglophones (cf. Paragraphe 1.1) ou encore par les travaux psycholinguistiques réalisés par Cutler et al. (1986) sur les stratégies de segmentation de la chaîne parlée. En particulier, en essayant de reproduire l'expérience de Mehler et al. (1981) en anglais, Cutler et al. (1986) démontrent que des locuteurs monolingues anglais soumis à l'écoute de stimuli français (ex. « balcon » ; « balance ») n'adoptent pas de stratégies de syllabation des stimuli. En effet, les sujets français exposés aux stimuli anglais (ex. *balcony* ; *balance*) tentent de resyllaber l'input inconnu en accord avec la modalité de segmentation la mieux adaptée à leur L1. En revanche, dans le cas de la segmentation du stimulus français « balance », Cutler et al. (1986) observent que les sujets anglais ne sont pas capables de définir les frontières syllabiques et hésitent entre rattacher la consonne à la première ou à la deuxième syllabe.

> « […] the segmentation processes characteristically employed by French speakers and English speakers differ: French speakers make use of syllabification in the segmentation; English speakers do not. We assume that this difference reflects the phonological differences between French and English ». (Cutler et al. 1986, p. 397)

Cette différence de stratégies de segmentation ne serait pas seulement limitée au contraste entre le français et l'anglais mais pourrait être élargie à d'autres langues ayant un accent de mot où la correspondance entre les frontières lexicales et syllabiques est plus forte. Ces résultats sont d'autant plus importants si nous considérons que la liaison constitue un phénomène spécifique du français qui oblige d'un côté l'apprenant à résoudre l'ambigüité des frontières lexicales provoquée par ce phénomène et de l'autre à réaliser en production une resyllabation de la CL à l'initiale du mot suivant.

Du point de vue de la production de la liaison en acquisition L2, ces arguments rejoignent les hypothèses de Wauquier (Wauquier 2009 ; Wauquier et Shoemaker 2013) qui observent une

forte présence de réalisations de liaisons non-enchaînées chez des locuteurs anglophones (Mastromonaco 1999 ; Thomas 2002 ; Howard 2005 ; Tennant sous presse). Il serait pourtant excessif d'imaginer que ces formes de non-enchaînement sans valeur stylistique particulière apparaissent uniquement dans les productions des apprenants ayant comme langue source une langue germanique. En effet, la présence de non-enchaînement attestée chez des apprenants ayant des L1 caractérisées par un système syllabique ouvert tel que l'italien (Falbo et al. sous presse) et l'espagnol (Racine sous presse) remet en question l'hypothèse de l'influence de la L1. Une autre explication éventuelle pouvant rendre compte de ces phénomènes nous vient des études menées sur la L1 (Chevrot et Malderez 1999; Laks 2005b). Celles-ci proposent de traiter ces formes comme le résultat de l'influence de l'orthographe dans la production de la liaison (voir aussi Racine sous presse ; Tennant sous presse). Cette hypothèse serait appuyée par l'observation dans l'étude de Racine (sous presse) d'un « effet bénéfique » apporté par le séjour dans l'enchaînement des liaisons chez les apprenants avec séjour, uniquement attesté en conversation.

> « La présence de liaisons non enchaînées chez les apprenants hispanophones montre que leur présence observée chez les anglophones par Mastromonaco (1999), Thomas (2002) et De Moras (2011) ne peut pas être expliquée par la tendance à la syllabation fermée des langues germaniques, l′espagnol ayant, comme le français, une prédilection pour les syllabes ouvertes. Notre hypothèse qu′il s′agirait d′un effet orthographique, la lecture favorisant le traitement mot à mot, avec un effet plus marqué de la frontière entre les unités dans le texte qu′en parole spontanée doit cependant être nuancée par l′effet du séjour en milieu francophone. La différence inter-tâches n′est en effet présente que pour le groupe avec séjour. Cette influence orthographique serait donc prépondérante sans distinction de tâche chez les apprenants n′ayant pas effectué de séjour prolongé en milieu francophone. Elle s′atténuerait en parole spontanée lorsque l′apprenant est en immersion ». (Racine sous presse)

Cependant Racine (sous presse), comme l'avait déjà observé Valetopoulos (cf. Note 234), invite à la prudence quant à l'interprétation de ces résultats[242].

> « Ces résultats doivent néanmoins être considérés avec prudence. Nos données révèlent en effet qu′un cumul des deux phénomènes, à savoir la réalisation d′une liaison non enchaînée avec une consonne orthographique, est parfois possible, par exemple, dans le texte, 'grand

[242] Toutefois les explications alternatives proposées par Racine (sous presse) et Valetopoulos (sous presse) sont différentes. En effet, si Racine considère que les formes telles que « grand honneur » [gʁɑ̃n-ɔ-noeʁ] encodées selon le codage IPFC en tant que liaison non-enchaînée correspondent de toute manière à des formes de liaison (CL non conforme mais présente dans le mot), Valetopoulos (sous presse) et Falbo et al. (sous presse) s'interrogent sur cette interprétation et sur la possibilité de considérer ces réalisations comme le résultat d'une dénasalisation des consonnes nasales (cf. Note 234). Racine (sous presse) mentionne également cette interprétation dans des contextes non liaisonnant tels que « son séjour » réalisé [sɔ̃n-se-ʒuʁ].

«Comme les apprenants produisent également des formes telles que [sɔ̃n-se-ʒuʁ] ″son séjour″, dans un contexte qui n′est pas liaisonnant, rien ne nous permet de déterminer qu′il ne s′agit pas d′une généralisation de la réalisation de la consonne graphique finale (<n>) de graphèmes représentant une voyelle nasale ([ɔ̃]) ». (Racine sous presse)

> honneur' [gʁɑ̃n-ɔ-noeʁ], 'ont eu' [ɔ̃n-y] et, en conversation, 'sont un peu' [sɔ̃n-ɛ̃-pø] et 'quand on' [kɑ̃n-ɔ̃] […] il paraît toutefois difficile, dans ce cas – et ce même s'il s'agit d'apprenants avancés – de déterminer s'il s'agit réellement de liaisons non enchaînées ou plutôt de la prononciation d'une consonne graphique en finale de mot ». (Racine sous presse)

De plus, comme nous l'avons vu dans les paragraphes précédents (cf. Paragraphe 2), l'absence d'une différence inter-tâche dans la production des liaisons non-enchaînées chez les apprenants norvégophones (Andreassen et Lyche sous presse) de niveau supérieur (B1-B2) nous suggère que l'hypothèse de l'influence du système prosodique (Andreassen et Lyche sous presse) et syllabique (Thomas 2002 ; Wauquier 2009 ; Tennant sous presse) doit sans doute continuer à être prise en compte dans l'analyse des réalisations des liaisons non-enchaînées.

5. Le rôle de l'orthographe dans l'acquisition de la liaison en L2

La forme orthographique des mots joue un rôle crucial dans l'apprentissage de la liaison chez les natifs (Laks 2005b)[243]. Si la graphie et l'identité visuelle des mots exercent une forte influence au niveau cognitif et renforcent la représentation phonologique des mots chez les natifs (cf. Partie 1.Chapitre 6.3), les résultats des études menées chez les apprenants L2 sont quant à eux plus contradictoires (Young-Scholten 1998 ; Young-Scholten et al. 1999 ; Young-Scholten 2002 ; Bassetti 2007) (cf. Partie 3.Chapitre 1.3.2.4.1). En effet, cette variable peut jouer un rôle positif dans l'apprentissage d'une L2 et renforcer l'opposition phonologique entre phonèmes, favorisant leur catégorisation (Erdener et Burnham, 2005 ; Simon et al. 2010). Néanmoins, l'input écrit peut également exercer un effet négatif dans l'apprentissage d'une L2 en provoquant, en particulier dans le cas de l'acquisition de la liaison en milieu hétéroglotte, la production de formes calquées sur l'orthographe comme dans le cas des prononciations orthographiques des CL (cf. Paragraphe 1.4). L'apprenant produit la liaison en se basant sur la valeur phonographique de base des graphèmes et donc en faisant appel à la forme écrite des mots avant de les prononcer. La production de ces formes peut être accentuée, dans le cadre d'un apprentissage guidé en milieu hétéroglotte, par une exposition limitée à l'input oral associée à une maîtrise imparfaite du système phonologique de la L2.

De plus, comme le souligne Racine (sous presse) des difficultés plus importantes apparaissent chez des apprenants italophones et hispanophones de français L2 dont la L1

[243] Laks (2005b) parle d'une sorte d'inculcation qui détermine un dressage linguistique à partir de l'identité visuelle du mot, qui modifie au fur et à mesure la représentation mentale du lexique. À travers l'apprentissage de la norme orthographique, l'enfant parviendrait définitivement à maîtriser la liaison variable (cf. Partie 1.Chapitre 5.3.5).

(italien et espagnol) est caractérisée par une correspondance transparente entre phonie et graphie.

Ces facteurs pourraient ralentir le développement des représentations phonologiques des CL et conduire les apprenants à produire des formes erronées quand la CL n'est pas « identique à la dernière consonne écrite du mot considéré » (Thomas 2002, p. 4). Il convient toutefois de souligner qu'en acquisition L2 ainsi qu'en L1 (cf. Partie 1.Chapitre 6.3) (Morais et al. 1987 ; Chevrot et Malderez 1999), les positions divergent quant au niveau (lexical ou sub-lexical) où les interactions entre phonologie et orthographe ont lieu (cf. Partie 4.Chapitre 3.1.3.1). Quel que soit le niveau, ces considérations posent des questions sur les modalités d'enseignement de la liaison, surtout dans un contexte hétéroglotte où l'acquisition se fait essentiellement par exposition à l'input écrit et en particulier chez des apprenants dont la L1 présente une forte correspondance entre graphie et phonie (Racine sous presse).

6. Conclusion

Malgré les difficultés de comparabilité entre les études, la plupart des travaux ont mis en évidence un certain nombre d'erreurs au niveau de la « micro-planification » (Racine 2014).

Parmi ces difficultés, nous avons évoqué la présence de réalisations orthographiques de la CL ainsi que la persistance de formes de liaison non-enchaînées sans valeur stylistique. Bien que ces résultats renvoient à une influence possible de l'orthographe et de la L1 de l'apprenant dans l'acquisition de la liaison, des études supplémentaires sur des populations plus vastes d'apprenants (plusieurs langues sources) doivent être menées. Deux questions méritent en particulier d'être approfondies: (a) à quel niveau (lexical ou sub-lexical) se situent les interactions entre phonologie et orthographe dans l'acquisition de la liaison ? ; (b) dans quelle mesure le système syllabique de la L1 de l'apprenant influence-t-il la réalisation de l'enchaînement de la liaison dans les productions des apprenants ? Nous essayerons de proposer des réponses à ces questions dans la quatrième partie de ce travail consacrée à l'étude empirique menée auprès d'apprenants italophones de FLE (cf. Partie 4.Chapitre 1).

Quant au niveau de « macro-planification », les études précédentes ont également observé une maîtrise parfois faible de la variation sociolinguistique chez les apprenants L2 par rapport aux productions des locuteurs natifs. Néanmoins, comme le souligne Racine et Detey (sous presse), ces études renvoient le plus souvent à des modèles de prononciation normatifs (Delattre 1947 ; Fouché 1959). Il convient toutefois de souligner, comme nous le verrons dans le chapitre suivant, que d'un point de vue didactique la variation de la liaison constitue une source de

complexité pour l’enseignement (en particulier en milieu hétéroglotte) qui s’appuie le plus souvent sur des listes de règles issues des classifications que nous avons évoquées dans la première partie de la présente étude (cf. Partie 1.Chapitre 4.3).

Chapitre 3. Normes pédagogiques et matériaux didactiques dans l'enseignement de la liaison en français langue étrangère

Les études menées sur l'acquisition de la liaison, comme nous l'avons vu dans le chapitre précédent, ont été le plus souvent conduites dans des contextes d'acquisition guidée. Dans ce cadre, l'input orthographique joue un rôle crucial dans l'apprentissage de la liaison. En effet, dans un contexte d'acquisition institutionnelle et hétéroglotte, la plupart de l'input est écrit et l'apprenant est en outre peu exposé aux variétés des usages oraux. Étant donné que la liaison représente un indicateur social explicite, le choix des ressources didactiques ainsi que de la variété de référence est crucial dans le parcours didactique.

1. Quelle variété de référence pour l'enseignement de la liaison en FLE?

Comme nous l'avons vu, cette question est loin d'être anodine et ne concerne pas seulement la didactique du FLE mais aussi celle du français L1. Cet aspect problématique de la description de la liaison oblige les didacticiens à définir les objectifs de l'enseignement de la langue ainsi que la variété de langue que l'on se propose de faire acquérir. Elle invite d'autre part les linguistes acquisitionnistes à prendre position sur la place à accorder à la variation dans le système linguistique dans leurs analyses ainsi que dans l'évaluation des performances des apprenants. En effet, dans l'enseignement de la liaison comme pour d'autres aspects linguistiques, ce choix n'est pas sans importance étant donné que la liaison représente un indicateur social explicite.

Dans la première partie de ce travail (cf. Partie 1.Chapitre 3), nous avons montré comment la norme de prononciation a pu varier au cours du XXème siècle en fonction des valeurs et des groupes sociaux dominants des différentes époques : des variétés de français parlé par une classe sociale restreinte (haute bourgeoisie parisienne) jusqu'aux modèles basés sur des « usages moyens » tels que ceux du « français standard » (Léon 1968 ; Borrell et Billières 1989) et du « français standardisé » (Carton et al. 1983).

Ces modèles de prononciation ont également servi de base pour la description de la liaison dans le cadre de l'enseignement de la liaison en FLE. Les normes de prononciation adoptées par les didacticiens ont aussi suivi les processus de « démocratisation » et de « décentralisation » (cf. Partie 1.Chapitre 3) favorisés, comme nous l'avons vu, par l'évolution des méthodologies

descriptives (cf. Partie 1.Chapitre 3.2). L'adoption de nouvelles approches didactiques basées sur l'étude des corpus oraux (*data driven learning,* apprentissage sur corpus) a contribué au passage des travaux basés sur la norme de prononciation d'un groupe social limité aux ouvrages proposant des modèles de prononciation géographiquement et socialement plus diversifiés.

1.1. Français parisien cultivé : une tentative de « mettre de l'ordre »

Comme nous l'avons déjà montré (Partie 1.Chapitre 4.3.1), les premières descriptions normatives de la liaison étaient en effet destinées aux étrangers. C'est le cas par exemple de l'étude de Fouché (1959) où le traitement didactique de la variation de la liaison se réduit à la description des variétés de l'élite cultivée parisienne. Ce choix est motivé chez Fouché (1959) par la préoccupation de « mettre de l'ordre » parmi les multiples variétés et usages qui caractérisent la liaison.

> « La question [*quant-on fait la liaison ?*] n'est pas simple, car dans le français dit correct il y a plusieurs variétés à distinguer, et dans chacune d'elles l'usage n'est pas nécessairement le même. Pour mettre de l'ordre dans une matière aussi compliquée, nous choisirons comme base le français tel qu'il se prononce dans la lecture à voix basse d'un texte ordinaire […] dont le style est plus au moins littéraire, et de l'autre la langue de la conversation familière (nous ne disons pas « populaire ») […] ». (Fouché 1959, 437)

La référence à une norme unique de prononciation se présente donc comme une solution didactique nécessaire pour des raisons d'économie et de clarté descriptive[244] en même temps.

La même préoccupation caractérise la description de Delattre (1956) qui désigne comme norme de prononciation les usages de « la conversation naturelle de la classe cultivée ».

> « Sauf indication particulière, tout ce que nous dirons s'appliquera au ton de conversation naturelle de la classe cultivée. Même dans un style aussi défini, il reste une marge de variations sensible; le ton se nuance selon l'interlocuteur, le complet qu'il porte, l'heure du jour et même le temps qu'il fait, et la liaison se plie à chacune de ces nuances. Nous tiendrons rarement compte des facteurs décrits dans notre article précédent, mais il sera toujours sous-entendu qu'ils peuvent encore réduire ou augmenter les possibilités de liaison de toute séquence ». (Delattre 1956, p. 49‑50)

Cependant, comme nous l'avons plusieurs fois observé, Delattre (1966) accorde une place plus importante à la dimension de variation et aura le mérite de reconnaître en premier l'existence de la pluralité des facteurs qui conditionnent la réalisation de la liaison mais que l'on peut difficilement représenter par un traitement didactique basé sur des règles.

[244] Il est intéressant d'observer comment cette même clarté descriptive, assurée par la référence à une seule variété et à l'exclusion implicite des autres usages, a guidé les travaux génératifs dans le choix des études normatives comme sources de données.

En revanche, Malmberg (1968 cité par Galazzi 2015b), reconnaissant la difficulté de déterminer « où est la 'bonne société parisienne' », suggère de faire référence à « une prononciation française que nous jugeons assez répandue et faisant autorité pour servir de norme à l'étudiant étranger » (Malmberg 1968, p.13, cité par Galazzi 2015b).

Pour conclure, si d'un côté ces descriptions semblent adopter une norme de prononciation limitée aux usages des classes sociales privilégiées (dont les critères d'appartenance sont difficilement identifiables cf. Partie 1.Chapitre 3), elles proposent d'autre part des solutions didactiques visant à limiter « la constante fluctuation » du comportement de la liaison et aboutir ainsi à la formulation de règles.

1.2. Recherche d'une nouvelle norme de prononciation : entre nivellement des usages et intégration de la variation sociolinguistique

Comme nous l'avons déjà observé (cf. Partie 1.Chapitre 3.2.2), le développement d'études phonético-phonologiques (Martinet 1971 ; Martinet et Walter 1973) de la deuxième moitié du XXème siècle ainsi que les travaux de dialectologie perceptive (Boula de Mareüil et al. 2008) permettront de dépasser les modèles de prononciations limités aux usages d'un groupe social. Les résultats de ces études encourageront la recherche d'une nouvelle norme de prononciation pour l'enseignement du FLE qui se définira autour de deux tendances opposées. Si en effet la primauté du français parlé par « la bonne société parisienne » sera remise en discussion en faveur du « français standard » des professionnels de la parole publique (Léon 1968)[245] ou encore d'un « français standardisé » (Carton et al. 1983) caractérisé par une « prononciation non aperçue », ces positions en faveur des usages peu marqués[246] seront suivies d'un mouvement opposé animé par la volonté d'intégrer la variation dans l'enseignement / apprentissage du FLE. Galazzi (2015b) illustre bien cette évolution.

> « Dans le monde actuel, vivant à l'heure de la globalisation et du décloisonnement, une ouverture est incontournable sur la francophonie dans toutes ses dimensions (horizontales et verticales), sur les traits phonétiques (non parisiens) qui font consensus dans certains usages sociaux (la langue des jeunes) ou géographiques (le Québec, la Suisse etc.) et même sur les accents "étrangers". À l'opposé de la standardisation neutre et unifiante, vers la fin du XXe

[245] Léon (cité par Galazzi 2015b) considérait le français « standard » comme un français « conforme au modèle attend ». « Il s'agissait en outre de présenter aux étudiants variés un système passe-partout, sans trop tenir compte de toutes les variantes possibles […] Comme les enseignants de langue voulaient qu'on leur fournisse des exemples concrets, on leur disait, pour le FLE, que le français parlé à la télévision (en France, sur la première chaîne, je crois) pouvait servir en quelque sorte de 'norme' du 'bon parler'» (Léon cité par Galazzi 2015b).

[246] Autrement dit « le français standardisé » (Carton et al. 1983) est caractérisé par un élargissement aux traits partagés par les différentes varietés de français à savoir « une prononciation qui tend à être adoptée comme celle du fr. courant central, dans toutes les régions » (Carton cité par Galazzi 2015b).

> siècle, à travers les études de la sociolinguistique américaine et européenne, un mouvement inverse se fait jour, tout aussi puissant, qui consiste à prendre en charge de manière de plus en plus consciente la variation linguistique à tous les niveaux. L'enseignement de la prononciation doit rester pluriel et ouvert ». (Galazzi 2015b)

Comme le soulignent Detey et Racine (2012), ces deux tendances apparemment contradictoires sont également présentes dans le Cadre Européen Commun de Référence pour les Langues (CECR) et ses référentiels pour la langue française. En effet, si la section « Matière sonore » du CECR renvoie à une norme de prononciation « standard » ou « standardisée » du français, nous retrouvons dans les référentiels sous la section « Normes et variations » une proposition de norme pédagogique qui intègre davantage la variation sociolinguistique.

> « Le français parlé connaît de nombreuses variantes nationales (France, Belgique, Suisse, Québec…) régionales et sociales. Toutes les personnes ayant une fonction publique (politiciens, annonceurs,…), quelle que soit leur région, ont généralement une prononciation dont les caractéristiques ont suffisamment de points communs pour qu'on en tire un modèle, dont les marques d'appartenance à un groupe minoritaire sont absentes ou faibles, modèle reconnu comme le français standard, dit aussi français des médias. Il présente beaucoup de variantes, suivant les critères suscités, mais avant tout suivant les situations de communication. On utilise ce modèle comme norme pédagogique. La compétence phonétique des locuteurs doit donc correspondre à l'exigence des auditeurs suivant les situations de communication ». (Beacco, Bouquet et Porquier 2004, p. 272)

L'opposition entre le nivellement et la diversité des usages reflète une recherche d'adaptation des données diversifiées issues de la linguistique du corpus à l'enseignement du FLE.

C'est au sein de ce passage délicat entre l'analyse des données attestées dans les corpus et leur didactisation qu'émerge à nouveau la problématique du « français de référence » dont les contours et les caractéristiques phonético-phonologiques (parmi lesquelles la réalisation de la liaison) restent encore floues (Francard et al. 2000) (cf. Partie 1.Chapitre 3). Certains linguistes et didacticiens se demandent donc si ce français aux contours théoriques et abstraits représente le meilleur modèle pédagogique ou s'il ne constitue qu'« un idéal » (Galazzi 2008a, p. 121).

1.3. Acquisition et enseignement de la variation sociolinguistique

La variation sociolinguistique a été longtemps négligée ou peu abordée dans les études acquisitionnnelles de l'interlangue qui d'après Mougeon et al. (2002) adoptent rarement une approche variationniste.

> « […] les travaux sur la variabilité de l'interlangue ne portent pas nécessairement sur les aspects de la langue cible où l'on peut observer de la variation sociolinguistique. De plus, lorsqu'il arrive à ces travaux d'être centrés sur des éléments sociolinguistiques dont le maniement est régi par des règles variables […] ils examinent rarement ce maniement selon la perspective sociolinguistique variationniste ». (Mougeon, Nadasdi et Rehner 2002, p. 3-4)

Néanmoins, un certain nombre de travaux (Mougeon, Nadasdi et Rehner 2002 ; Gadet et Tyne 2007 ; Galazzi 2008a, 2015b ; Galazzi et Molinari 2007 ; Tyne 2009; Detey et Racine 2012 ; Paternostro 2014 ; Paternostro 2015) ont été consacrés à l'acquisition des usages variationnels en particulier chez des apprenants avancés. Mougeon et al. (2002) ont montré que la maîtrise de la variation sociolinguistique chez des apprenants en milieu institutionnel (et de niveau avancé) était assez « modeste ». En effet les apprenants en milieu institutionnel auraient tendance à utiliser des variantes standard plus souvent que les natifs. Mougeon et Rehner (2001 cité par Mougeon et al. 2012) ont attribué ces résultats au traitement didactique adopté par les enseignants qui amènerait les apprenants à préférer les variantes qu'ils perçoivent comme plus correctes et standard (Mougeon, Nadasdi et Rehner 2002, p. 15). Cet aspect est également mis en évidence par Gadet et Tyne (2007) pour lesquels l'enseignement de la variation sociolinguistique est dans le meilleur des cas reporté aux stades d'acquisition les plus avancés (Dewaele et Mougeon 2002).

> « […] si les apprenants sont plus ou moins sensibles à la variation, leur palette est moindre que celle d'un natif, avec des productions typiquement sur-standardisantes, et ce n'est qu'aux niveaux les plus avancés que l'on voit des éléments de variation comparables à ceux observés en L1. Ainsi, l'appropriation du style constituerait le dernier obstacle dans l'acquisition d'une L2 ». (Gadet et Tyne 2007, p. 132)

Cette vision se reflète également dans les manuels d'enseignement.

> « Les conceptions courantes du style mises en œuvre dans les manuels, de L1 comme de L2, ne proposent aucune représentation un tant soit peu globale du phénomène : elles regardent celui-ci comme localisé en des points linguistiques, dont la maîtrise est une étape ultime, un ' en plus' d'une langue déjà maîtrisée ». (Gadet et Tyne 2007, p. 132)

Selon une telle approche, l'apprentissage des variantes sociolinguistiques demanderait, à cause de la complexité du traitement et de la gestion de la variation (surtout en milieu hétéroglotte), l'acquisition préalable d'un système linguistique relativement solide. Cependant Galazzi (2008a, p. 120) souligne que la validité d'un enseignement qui propose une « progression idéale allant du simple au complexe où l'affectif serait plus difficile à apprendre » reste encore à prouver.

Tyne (2012) souligne également que bien que les apprenants ne parviennent à une maîtrise partielle de la variation qu'à des stades avancés de l'acquisition, ils auraient conscience des variations sociolinguistiques dès les niveaux débutants.

> « […] il semble que même les apprenants débutants soient conscients des variations socio linguistiques (voir à propos du niveau A1 dans J.-C. Beacco et M. Di Giura, 20071), mais que la relative maitrise de cette variation, selon la plupart des études, n'apparaisse qu'au fur et à mesure de l'acquisition ». (Tyne 2012, p. 106‑107)

Tyne (2009) montre que les apprenants de niveaux avancés ne sont pas forcément ceux qui possèdent un plus vaste répertoire de variations (Tyne 2009, p. 271). De plus, les apprenants adultes sont déjà conscients de la variation présente dans leur L1. En dépit des difficultés d'enseignement / acquisition, Tyne (2009) propose d'intégrer la variation sociolinguistique dans les programmes d'enseignement dès les premières phases de l'apprentissage d'une L2.

> « […] we would say that it is difficult to consider style seen thus as a 'final' acquirable feature, the icing on the cake (could we ever conclude that style has been definitively acquired in L1, even?). Rather, we would suggest that it be considered as a part of the cake itself, present from the outset ». (Tyne 2009, p. 262)

Toutefois, bien que l'apprenant de FLE puisse s'appuyer sur les compétences sociolinguistiques apprises dans sa L1, d'après Gadet et Tyne (2007) ces liens « ne sont pas facilement applicables aux apprenants de L2, qui, surtout en contexte guidé, ne forment pas une communauté linguistique » (Gadet et Tyne 2007, p. 133). Il est évident que si l'enseignement du FLE ne peut pas renoncer à une réflexion sur la variabilité sociolinguistique des usages linguistiques, il reste à définir les implications didactiques qui en découlent et qui concernent la norme pédagogique (Galazzi et Molinari 2007).

Galazzi (2008a) suggère de ne traiter le français « standard » que comme une « étape intermédiaire » de l'acquisition qui nécessite d'être suivie d'une confrontation à la pluralité des variétés de prononciation du français.

> « Ce 'standard' ne peut constituer à mon avis qu'une étape intermédiaire dans l'itinéraire d'apprentissage qui, aux niveaux les plus avancés, demande un investissement et apport individuels et une adhésion personnelle au choix éventuel d'un 'accent natif' ». (Galazzi 2008a, p. 121)

C'est dans le cadre de l'approbation d'une sensibilité linguistique que l'enseignement de la variation sociolinguiste semble pouvoir être intégrée dans l'acquisition du FLE.

> « […] la sensibilité au diaphasique, au minimum comme ressource passive, apparaît comme une dimension souhaitable de l'apprentissage d'une langue étrangère, mais on ne peut y recourir que de façon prudente, et en tous cas à condition de ne pas trop interférer avec le

diastratique : pas de familiarisation active avec des formes par trop vernaculaires ou avec l'argot ». (Gadet 2004, p. 26)

1.3.1. L'« acceptabilité » sociolinguistique dans les normes pédagogiques

Le besoin de définition d'une norme de prononciation pour l'enseignement du FLE semble donc se heurter à la variabilité de la réalité linguistique. Cela revient à dire que, comme les soulignent Lyche (2010) et Detey et Racine (2012), toute approche didactique sera obligée d'opérer un choix motivé par des raisons purement socio-éducatives.

> « […] il n'existe pas un français mais des français, chaque système possédant sa propre dynamique. […] Le choix de la variété à enseigner relèvera in fine d'un choix socio-éducatif et non pas linguistique ». (Lyche 2010, p. 160‑161)

Certains auteurs (Valdman 1989 ; Lauret 2007 ; Detey 2010 ; Detey et Racine 2012) ont essayé de définir les critères nécessaires pour opérer ce choix. Lauret (2007) propose d'adopter un modèle de prononciations visant à assurer la plus large intelligibilité pour les apprenants même au détriment d'un éloignement des usages réels des locuteurs.

En revanche, Valdman (1989; 2000) définit des critères d'élaboration des normes pédagogiques en soulignant la nécessité que : (a) elles reflètent la parole réelle des locuteurs natifs dans des situations de communication authentiques ; (b) elles se conforment à la vision idéalisée qu'ont les locuteurs natifs de leur propre utilisation de la langue ; (c) elles se conforment aux attentes des locuteurs natifs et à celles des apprenants étrangers concernant le type de comportement linguistique approprié pour des apprenants étrangers ; (d) elles devraient prendre en compte les facteurs de traitement et d'apprentissage.

À partir de ces critères, Detey (2010) et Detey et Racine (2012) proposent de distinguer deux principes nécessaires pour la définition de la norme pédagogique : l'apprenabilité et l'acceptabilité sociolinguistique.

> « […] exposer les apprenants à trois ou quatre variétés phonologiques différentes (et donc aux productions orales qui leur sont associées), par exemple, ne signifie pas l'imposition de trois ou quatre variétés différentes à maîtriser en production. Car si l'authenticité sociolinguistique de ces variétés en impose l'étude, à un moment ou un autre du parcours d'apprentissage, leur acceptabilité (linguistique) ne doit pas être valorisée abusivement au détriment de leur apprenabilité (psycholinguistique) par les apprenants, étant entendu que la multiplicité des indices phonético-phonologiques, voire, le cas échéant, leurs divergences ou leur incompatibilité, pourrait conduire à un échec en termes d'apprentissage». (Detey 2010, p. 157)

Ces auteurs estiment qu'un seul modèle de prononciation doit donc être présenté dans un premier stade de l'apprentissage afin d'accorder une primauté à un input unique et « standard » qui sera par la suite enrichi par l'apport des usages issus des autres variétés. L'apprenant sera

ainsi confronté et sensibilisé à une variété sociolinguistique qui favorise le développement de son « acceptabilité » sociolinguistique.

Galazzi (2008a) souligne également qu'un enseignement s'appuyant sur un modèle de langue unique dans la perception est non seulement inadéquat mais ne reflète surtout pas la réalité linguistique.

> « […] le fait de s'en tenir à un modèle unique serait tout à fait pernicieux pour l'oreille étrangère condamnée dans les limites étroites d'une seule prononciation, et qui plus est, concrètement inexistante ». (Galazzi 2008a, 123)

Quant à la variété « standard » à adopter au cours des premières phases de l'enseignement, les auteurs semblent renvoyer à une notion de français de référence caractérisé par une variation diatopique[247].

> « […] pour une zone donnée, il n'est pas inconcevable de tenter de cerner ce que pourrait être une variété « de référence » en interrogeant les auditeurs-locuteurs natifs de cette zone sur ce qu'ils jugent être la variété « de référence », ou bien « standard », étant entendu que ce standard peut varier selon des facteurs à la fois géo- et sociolinguistiques ». (Detey 2010, p. 158)

Pour conclure, peut-on donc formuler une réponse à la question posée au début de ce paragraphe ? Le débat reste ouvert tant que certains auteurs (Laks 2002) suggèrent que cette notion de français de référence n'est qu'une « construction purement doxique » alors que d'autres (Bertrand et Schaffner 2010 ; Lyche, 2011) recherchent des solutions applicables à l'enseignement du FLE.

Nous tenterons d'apporter d'autres éléments de réponse à cette problématique dans la quatrième partie de ce travail. Nous nous contenterons pour l'instant de constater que la réponse à cette question implique nécessairement une sélection. Malgré l'importance cruciale jouée par les corpus oraux dans l'enseignement et la sensibilisation de l'apprenant à la variété sociolinguistique, l'enseignement de la liaison doit forcément passer par le choix d'un modèle et par une description / classification qui, pour sa nature, aura un caractère normatif.

[247] Ici les auteurs renvoient aux travaux de dialectologie perceptive (Clopper 2004, Boula de Mareüil et al. 2007) qui ont essayé de définir les traits des français de référence à partir des études perceptives menées auprès de natifs.

Germain (cité par Galazzi 2015b) rappelle également que : « 'français standard' ne signifiait pas la même chose pour les Français de France et pour les francophones du Québec : chacun référait à sa ' bonne norme' ».

2. Matériaux didactiques

Comme nous l'avons vu dans la première partie de ce travail (cf. Partie 1.Chapitre 4.4), le caractère multidimensionnel de la liaison représente un obstacle pour la description de la liaison et donc également pour son traitement didactique. Thomas (1998) souligne que la liaison est souvent négligée et peu abordée par les manuels pédagogiques de FLE à cause de sa variabilité qui rend difficile son traitement classificatoire et la formulation de règles formelles. Toutefois, comme le soulignent Racine et Detey (sous presse), la liaison est un phénomène très fréquent en français[248] et représente « un passage incontournable dans l′enseignement de l′oral en FLE ».

Comme le fait observer Racine (2014), la plupart des ouvrages traitent souvent les aspects de l'acquisition de la liaison issus d'un « niveau de la macro-planification » (Racine 2014) à savoir l'acquisition des contextes des liaisons catégoriques, variables et erratiques. En revanche, les difficultés liées à la production même de la liaison issues du niveau de la micro-planification (Racine 2014), sont souvent peu abordées dans les ressources didactiques. Parmi celles-ci, Racine et Detey (sous presse) citent : (a) la nature de la CL à réaliser (en lien avec l'orthographe) ; (b) le placement de la CL dans la structure syllabique (avec ou sans enchaînement) ; (c) ainsi que « les modifications éventuelles de l′environnement immédiat (ouverture / dénasalisation de la voyelle précédente) ».

Comme le font en outre observer Racine et Detey (sous presse), malgré la richesse des travaux descriptifs et variationnistes de la liaison issus de l'étude des corpus (cf. Partie 1.Chapitre 3.2.2.2), les descriptions fournies par la plupart des manuels de prononciation du français adoptent une approche normative. Ces ressources pédagogiques renvoient souvent à la tripartition traditionnelle entre liaison obligatoire, facultative et interdite et la description de la liaison s'appuie sur des listes de généralisations et d'exemples de contextes de liaison (établis sur la base de critères morphosyntaxiques). En outre ces règles ne correspondent pas toujours aux usages linguistiques réels des locuteurs francophones attestés par les études des corpus oraux[249].

[248] Selon Boë et Tubach (1992 cité par Racine et Detey sous presse et Côté 2015), elle apparaît à peu près une fois tous les 16 mots.

[249] Racine et Detey (sous presse) cite l'exemple de la liaison après les prépositions monosyllabiques « en », « dans », « chez », « sans » et « sous » que Abry et Chalaron (1994) présentent comme un contexte de liaison catégorique alors que, comme nous l'avons vu dans la première partie de ce travail (cf. Partie 1.Chapitre 4.3.2.3.1), ce type de liaison est sujet à variation (entre autres Eychenne et al. 2014).

Comme en effet le soulignent Detey, Kawaguchi et Kondo (sous presse) les apprenants en milieu hétéroglotte[250] remarquent parfois « les écarts entre ce qui leur a été enseigné et les productions de leurs enseignants natifs en classe » (Detey, Kawaguchi et Kondo sous presse).

Les autres variables susceptibles de conditionner la production de la liaison (cf. Partie 1.Chapitre 2) sont rarement traitées sauf quelques aspects liés à la situation de communication et au registre stylistique.

Dans les pages qui suivent, nous allons nous attacher à établir un bilan provisoire des ressources actuellement disponibles pour l'enseignement de la liaison en FLE. Nous allons distinguer tout d'abord à la suite de Morin (2000) les traités de phonétique visant un public n'ayant besoin que « de compléments ponctuels » ainsi que les manuels phonétiques destinés à un enseignement plus systématique. Nous allons ensuite présenter les ressources orales issues de la linguistique du corpus pouvant être adaptées à l'enseignement de la liaison.

2.1. Traités de phonétique

D'après Morin (2000), ces ressources se limitent généralement à une description de l'articulation segmentale des mots ou fournissent simplement des indications en guise de complément pour un lecteur spécialiste. C'est le cas par exemple du traité de Grammont (1933) qui ne donne pas de règles pour la réalisation de la liaison, mais le présente comme un phénomène se produisant toujours à l'intérieur des groupes rythmiques au sein duquel elle fonctionne comme une marque de cohésion sémantique (ex. « savant aveugle »).

> « […] une suite de mots fait une impression psychique toute différente et peut avoir une valeur sémantique toute autre selon que les consonnes finales y sont prononcées ou non devant une voyelle ». (Grammont 1933, p. 416)

Cependant certains auteurs comme Passy (1887) et Fouché (1959) fournissent également des règles. Passy (1887) a été le premier à proposer des transcriptions comme ressources pédagogiques pour l'enseignement de la liaison et du français en général[251]. Comme le souligne Galazzi (2002a, p. 37), Passy a été le promoteur d'un important renouvellement des méthodes d'enseignement en France comme à l'étranger, en mettant l'oral[252] et non plus l'écrit au centre de l'apprentissage des langues. Le recours à la transcription phonétique gagnera au fur et à

[250] Dans le cas de cette étude il s'agit d'apprenants japonophones.

[251] Comme le fait remarquer Galazzi (2014) l'ouvrage « Les sons du français » de Passy (1887) propose une vaste typologie de transcriptions afin d'illustrer les divers styles de prononciation (familière, ralentie, soignée, très soignée et familière rapide) et ainsi que les différentes variétés de français (prononciations du Nord, du Midi, de Suisse et également des patois).

[252] L'oral devient grâce à Passy une « priorité éducative » dans l'enseignement du français langue étrangère (Galazzi 2014).

mesure plus de place dans les ouvrages pédagogiques destinés à l'enseignement du FLE. L'objectif de Passy était celui de « montrer la variation dans toute sa latitude, y compris les prononciations 'marginales' ou caricaturales » (Galazzi 2014, p. 69). Quant à l'ouvrage de Fouché (1959), même s'il s'agit en réalité d'un traité, il est destiné à l'enseignement du FLE. Cependant, l'auteur affirme dans la préface s'adresser également aux natifs. En effet Fouché (1959) déclare vouloir présenter, en réaction à la tradition d'études phonétiques, en particulier Martinon (1913), un ouvrage consacré à l'étude de la prononciation qui n'ait pas une visée purement didactique et qui puisse également offrir une analyse riche en détails susceptibles d'intéresser les « compatriotes eux-mêmes » (Fouché 1959, p. I).

> « Jusqu'ici, d'une façon générale, les ouvrages consacrés à l'étude de la prononciation française ont été rédigés essentiellement, sinon uniquement, en vue des Étrangers […] C'est le trait commun de tous le manuels de prononciation, du sien [celui de Martinon 1913] et des tous les autres, de passer sous silence une multitude de problèmes sur lesquels chacun de nous aimerait à l'occasion être renseigné. Dans l'ouvrage que nous présentons au public, nous avons tâché au contraire d'être le plus complet possible ». (Fouché 1959, p. I)

Contrairement aux autres auteurs, il consacre au traitement de la liaison un chapitre entier (cf. Partie 1.Chapitre 4.3.1.1). Il offrait une liste de règles ayant pour but de décrire les contextes où la liaison doit être réalisée et ceux où elle est erratique. Cependant, bien que cette analyse soit extrêmement fine, elle se prête difficilement à un traitement didactique. La volonté de l'auteur de fournir un grand nombre de détails destinés aussi bien aux natifs qu'aux étrangers limite les possibilités d'utiliser cette référence comme ressource pour l'enseignement. De plus, il convient de rappeler que Fouché (1959) se réfère encore à une « prononciation en usage dans la conversation soignée chez les Parisiens cultivés ».

2.2. Manuels de phonétique

Tous les manuels de phonétique ou de prononciation française contiennent une partie consacrée à l'analyse du phénomène de la liaison. La plupart des ceux-ci n'accordent au traitement de la liaison que quelques pages (d'une à cinq pages) (Carton 1974, p. 87‑88 ; Malécot 1977, p. 32‑37; Léon et Léon 1997, p. 39‑42 ; Charliac et Motron 1998, p. 16‑17 ; Lauret 2007, p. 59-61). Léon (1966, p.118-131 ; 2007, p. 223-237) est le seul à lui consacrer un chapitre entier. Quant à des ouvrages moins récents, Clédat (1890) dans sa grammaire phonétique (« Précis d'orthographe et de grammaire phonétique pour l'enseignement du français à l'étranger »), en dépit du traitement original qu'il propose de la CL (comme consonne résultant d'une « adjonction ») (cf. Partie 5.Chapitre 1.1), ne consacre à la liaison que quelques

lignes, dans chaque chapitre portant sur les catégories grammaticales traditionnelles (ex. liaison après les articles, les noms etc.).

Nous nous attacherons dans les pages qui suivent à présenter les caractéristiques principales de ces manuels qui sont le plus souvent repartis en trois sections : (a) définition du phénomène de la liaison ; (b) classifications des types de liaison ; (c) exercices.

En ce qui concerne les définitions (a), elles sont assez traditionnelles et présentent la CL comme une consonne finale, ordinairement non prononcée, qui est réalisée, dans certains cas, si le mot suivant (M2) commence par voyelle.

> « En français oral, de nombreuses consonnes finales de mots sont muettes (voir p.11). Certaines consonnes finales de mot peuvent, dans certains cas, être prononcées avec la voyelle initiale du mot qui suit. C'est la liaison ». (Charliac et Motron 1998, p. 16)

Seuls certains auteurs font référence au domaine principal de réalisation de la liaison notamment au groupe rythmique (Léon 1966, p. 119 ; Léon 2007, p. 225) ou au «degré de cohésion accentuelle » (Charliac et Motron 1998, p. 34).

> « Règle II : Liaison d'un mot inaccentué à un mot accentué (intérieur du groupe rythmique) ». (Léon 1966, 119)

> « Dans le parler ordinaire, la liaison apparait seulement à l'intérieur du groupe rythmique. Il n'y a pas de liaison après un mot accentué ». (Léon 2007, p. 225)

> « La liaison est généralement impossible entre un mot accentué et le mot suivant ». (Charliac et Motron 1998, p. 18).

De plus, à l'exception de quelques ouvrages (Léon 1966 ; Léon 2007)[253], comme le soulignent Racine et Detey (sous presse), la plupart des auteurs proposent des descriptions sommaires des correspondances entre la graphie et les réalisations phonétiques des CL. Racine (sous presse) met en évidence l'importance de faire appel à une analyse plus approfondie des correspondances entre graphie et phonie afin d'éviter la production massive de réalisations orthographiques de la CL (cf. Partie 3.Chapitre 2.1.4).

> « La question du support graphique doit être mieux prise en compte, en intégrant dès le début de l′apprentissage des activités diversifiées, certaines s′appuyant de manière explicite sur l′orthographe – en travaillant les correspondances phonie-graphie –, alors que d′autres doivent en faire totalement abstraction ». (Racine sous presse)

[253] Léon (1966, p. 121) consacre à la description des consonnes graphiques correspondant aux CL une table détaillée où non seulement il donne des exemples mais essaie également d'établir une classification entre la nature de la CL, le type graphique et la nature des mots de liaison correspondants (ex. CL [n] (ex. « un ami » ; type graphique = <n> ; nature des mots de liaison « un, aucun, en, on, bien, rien, mon, ton, son ; adjectifs en [ɛ̃], type : moyen » (Léon 1966, p. 121).

Ces auteurs relèvent également que « la liaison est généralement traitée de manière totalement déconnectée à la fois des aspects prosodiques et de ceux qui touchent le niveau segmental » et que les explications ou les exercices fournis pour décrire les dénasalisations de la voyelle précédente les CL /n / sont en nombre restreint. En revanche, Léon (1966) illustre en détail « l'effet de la liaison sur la voyelle précédente » adoptant une démarche plus didactique par rapport à la description précédemment proposée par Fouché (1959).

Il nous semble également que, sauf Léon (1966; 2007) qui présente un cadre historique des origines de la liaison, une mise en relation avec le phénomène plus général d'enchaînement consonantique est généralement absente dans les manuels phonétiques.

En particulier Léon (1966, p. 119) après avoir proposé des définitions préliminaires des phénomènes d'élision, enchaînement et liaison précise que bien que la liaison ne se réalise pas après un mot accentué (ex. « il est grand aussi »), l'enchaînement vocalique entre les deux mots subsiste.

Après les explications, une série de règles de réalisation de la liaison (b) (s'appuyant sur les différents tripartitions de la liaison, le plus souvent la classification proposée par Delattre 1966) est proposée sous formes de schémas ou de tableaux classés par catégorie morphosyntaxique.

Comme nous l'avons déjà observé (cf. Partie 1.Chapitre 4.3.2.2) bien que Léon (1966) adopte la tripartition proposé par Delattre (1966) il la situe dans le cadre plus large de deux règles qu'il établit sur la base des frontièrs des groupes rythmiques. Léon (2007) préfère également à cette tripartition une description basée sur la diversité du degré de cohérence syntagmatique. Il a le mérite d'avoir proposé une classification plus descriptive qui dépasse la simple tripartition de la liaison et qui prend en compte des variables syntaxiques, prosodiques et lexicales. Une partie de son chapitre est en outre consacrée à la présentation des rôles (ex. distinctif, morphologique et identificateur) de la liaison qui permettent d'apprécier l'importance de l'apprentissage de ce phénomène. Léon (1966) a également mis en évidence le rôle distinctif joué par la liaison au pluriel (ex. « il entre » *versus* « ils entrent ») et au masculin, dans le cas de la liaison après les adjectifs antéposés (ex. « grande enfant » *versus* « grand enfant »).

Carton (1974) préfère ne pas proposer de classifications des contextes qu'il juge inutiles étant donnée la variabilité de la liaison.

> « On fait traditionnellement la liste des cas où la liaison est obligatoire, interdite ou facultative. En fait une liste des liaisons facultatives n'est guère utile, elle dépend du registre de langue ». (Carton 1974, p. 18)

Il se limite donc à donner des indications de « proportions » de réalisation de la liaison selon le registre de langue[254]. De même, bien que Malécot (1977) fasse appel à l'usage des tripartitions traditionnelles et souligne que la répartition des liaisons possibles dans ces catégories varie selon le style.

> « […] les liaisons facultatives dans la conversation sont généralement obligatoires dans l'élocution [*Ex. verbe+participe passé ; être +complément*], et les liaisons facultatives dans l'élocution sont pratiquement inusitées dans la conversation [*après nom pluriel ; après préposition polysyllabiqu*e] ». (Malécot 1977, p. 32)

Quant aux activités proposées (c), les exercices contenus dans l'« Introduction à la phonétique » de Malécot (1977, p. 52-54) demandent à l'apprenant d'appliquer des règles présentées dans la première partie théorique du texte. L'auteur propose à l'apprenant un exercice à la fois de lecture et métalinguistique, en demandant de préciser si la liaison contenue dans les séquences lues est catégorique, variable ou erratique et d'expliciter la règle correspondante.

À l'opposé, nous avons les exercices contenus dans le manuel de Léon (1966) et de Léon et Léon (1997), où on demande de réaliser des transcriptions phonétiques des phrases écrites contenant des contextes de liaison ou encore de justifier l'absence d'une liaison ou sa fonction (ex. fonction démarcative de l'absence de liaison) (Léon 1966, p. 126)[255].

> « 5.Transcrivez : Ma grande amie et son grand ami sont deux êtres charmants. Chez eux, les invités sont à l'aise
>
> 6. Dans la phrase ; ils se sont retrouvés après vingt ans, pourquoi ne fait-on pas habituellement de liaison après retrouvés ?
>
> 7. Transcrivez : Les heures passent vite avec des hippopotames disent les Hollandais.
>
> 8. À quel problème phonétique se heurte-t-on dans la phrase précédente ?
>
> 9. Transcrivez : Un certain aspect du Moyen Âge, est apparu soudain.
>
> 10. Indiquez entre parenthèses, si la liaison est obligatoire (O) facultative (F) ou interdite (I) dans les énoncés suivants ». (Léon et Léon 1997, p. 41‑42)

Exemple 10 : Excercices pour l'acquisition de la liaison (Léon et Léon 1997)

Léon (1966, p. 128) présente aussi des exercices pour étudier les correspondances entre la graphie et les réalisations phonétiques des CL, dans lesquels on demande à l'apprenant de retrouver la graphie des mots à partir de leurs transcriptions phonétiques (ex. [leero] « les héros »). De plus cet auteur établit une description de la liaison variable où il propose de

[254] Il distingue la conversation familière (où la liaison n'est presque jamais réalisée) de la conversation courante soignée (qui possède une petite proportion de réalisation), le discours télévisé / la conférence de type littéraire (où la majorité de liaison est réalisée) de dictions des vers classiques (dans lesquelles toutes les liaisons possibles sont produites).

[255] Ex. opposition entre le singulier et le pluriel (ex. « enfant // adorable » *versus* « enfants adorables »).

distinguer selon le style entre : (a) les liaisons peu fréquentes (ou en voie de disparition) dans la conversation familière, mais normales dans un style soigné (ex. liaison après tous les adverbes sauf « très », « quand » ex. « vraiment adorable ») ; (b) les liaisons anormales dans le style parlé mais possibles dans un style recherché (ex. liaison après les noms pluriels « des enfants adorables »)[256].

Léon (2003, p. 99‑100) établit également trois fiches explicatives sur la liaison (composées de trois sections : « définition », « fautes à éviter » et « conseils pratiques ») suivies par des exercices. Ceux-ci visent à l'écoute systématique des formes de liaison et sont souvent centrés sur un même mot prononcé dans des contextes différents (ex. « un grand homme », « un grand avion » et « un grand escroc » etc.).

En revanche, Abry et Chalaron (2009 ; 2011) proposent des exercices pour l'apprentissage de la liaison et de l'enchaînement[257], de discrimination et de resyllabation de la CL et d'identification des contextes de liaison (catégoriques, variables et erratiques).

Pour conclure ce bref bilan des ressources didactiques pour l'enseignement de la liaison, il nous paraît que les explications, classifications et activités proposés dans ce type d'ouvrage vont au-delà des spécificités de chaque manuel et se basent essentiellement sur des données obtenues par introspection ou sur des exemples construits *ad hoc* qui reflètent difficilement les divers principes d'élaboration d'une norme pédagogique proposés par Valdman (1989).

Toutefois, comme le soulignent Racine et Detey (sous presse), dans certains manuels (Lauret 2007 ; Pustka 2011)[258] la description de la liaison (au moins un niveau de la « macro-planification ») a été établie sur la base des résultats des études menées sur les corpus oraux. Cela a contribué à un « renouvellement du matériel pédagogique » en faveur de l′enseignement des usages linguistiques réels des locuteurs.

[256] En particulier Léon (1966) distingue quatre styles : (a) conversation familière (« tendance moderne à faire le moins de liaisons possibles ») ; (b) conversation soignée (« tendance à lier les mots vraiment peu accentués aux mots les plus accentués » (ex. « je vais aller » ; (c) style oratoire (« tendance à lier même les mots accentués sauf après une pause évidente » (ex. « films anciens ») ; (d) style poétique « tendance à tout lier sauf s'il résulte une impression de cacophonie « (ex. « je vais aller aussi »). (Léon 1966, p. 130)

[257] Il s'agit d'exercices de discrimination phonétique entre les CL et les consonnes fixes enchaînées à l'aide d'un tableau.

[258] En particulier Lauret (2007) évite d'avoir recours à la tripartition traditionnelle des contextes de liaison et propose une simple listes des types de liaison les plus fréquents. En revanche, bien que les manuels d'Abry et Chalaron (2009 ; 2011) soient très récent ils adoptent encore la classification proposée par Delattre (1947).

2.3. Corpus oraux pour l'enseignement de la liaison

Les ressources offertes par les grands corpus oraux permettent d'apporter des solutions pour la définition des normes pédagogiques. Depuis les années 90, le développement de l'ingénierie linguistique a contribué à la mise en place d'une interaction entre linguistique du corpus et apprentissage des langues sous la forme des approches dites « TaLC » (*teaching and language corpora*) (Stewart, Bernardini et Aston 2004). D'après ces travaux les corpus oraux peuvent être utilisés comme des supports pédagogiques exploitables dans l'enseignement. Parmi les avantages du recours à ces ressources, plusieurs fois soulignées (Chambers 2009 ; Tyne 2009 ; Tchobanov et al. 2009 ; Detey 2010 ; Detey et Racine 2012), nous citons les possibilités offertes par les corpus oraux d'éveiller chez les apprenants et les enseignants une sensibilité à la variation sociolinguistique par rapport aux normes de français « standard » ou « standardisé » (Paternostro 2014). Si la pertinence des études linguistiques des corpus oraux pour l'apprentissage des langues semble être reconnue par la communauté des didacticiens et des linguistes, les positions divergent quant aux modalités d'intégration de ces ressources dans l'enseignement et à l'authenticité de ces documents[259] (Chambers 2009 ; Debaisieux 2009). Si certains défendent la nécessité du recours à de grands corpus (Tchobanov et al. 2009), d'autres (Chambers 2009) proposent l'usage de corpus limités à un seul genre ou à un nombre restreint de genres. De plus, ces données sembleraient surtout être employées pour des activités de compréhension et peu exploitées pour l'apprentissage / enseignement de l'expression orale à cause des difficultés linguistiques posées aux apprenants (Debaisieux 2009). Ces problématiques montrent que l'intégration des corpus oraux dans l'enseignement est ici encore au centre du débat pédagogique. La plupart des auteurs suggèrent une introduction graduelle de ces données dans le processus didactique (Debaisieux 2009 ; Detey 2010 ; Detey et Racine 2012) à l'aide des « Systèmes d'Apprentissage Auto-Dirigés avec Soutien » (SAAS) ou à travers des activités de découverte. Comme nous l'avons souligné dans les paragraphes précédents (cf. Paragraphe 1.3.1), Detey et al. (Detey 2010; Detey et Racine 2012), conscients de l'impact que des données authentiques pourraient avoir sur l'apprenant[260], proposent dans

[259] Voir Widdowson (1998).

[260] « […] l'insertion, dans les programmes d'enseignement, d'input langagier autre que l'input standard, attendu et maîtrisé, comporte des enjeux sur plusieurs plans. Sur un plan socio-idéologique large, la (non-) hiérarchisation des variétés et la gestion de la sécurité/l'insécurité linguistique ; sur un plan cognitif, l'impact de la variation phonétique et phonologique sur l'apprentissage (Detey 2009) ; sur un plan « pragmatique » enfin, la mobilité accrue des locuteurs dans les espaces francophones. De l'enrichissement des descriptions sociolinguistiques à leur utilisation en classe, une étape importante reste encore, semble-t-il, à franchir ». (Detey et Racine 2012, p. 11)

les premières phases de l'apprentissage une exposition à un input « standard » suivie de l'introduction progressive des données issues des autres variétés sociolinguistiques.

> « […] en particulier en milieu hétéroglotte, il est évident que le choix du type de locuteurs (et donc de variétés) employés comme source d'input ne relève en général que de considérations trivialement pratiques – la disponibilité des locuteurs en question – avec, en cas de choix possible en milieu hétéroglotte, une préférence possible pour la norme de référence (bien souvent français standard ou standardisé), essentiellement pour des raisons de reproduction socio-académique. Une étude exhaustive et comparée entre différentes zones francophones des matériaux pédagogiques employés serait toutefois nécessaire pour poursuivre cette analyse ». (Detey et Racine 2012)

Selon une telle approche, la modalité d'intégration des corpus dans l'enseignement devrait se baser sur un équilibre entre une possibilité d'apprenabilité de l'input issu du corpus d'une part et l'acceptabilité de la variation sociolinguistique de l'autre.

> « […] l'objectif est bien d'entraîner les apprenants à gérer la diversité en perception, tout en ne leur fournissant, au moins initialement, qu'un modèle de production, à déterminer en fonction du cadre géolinguistique et didactique. C'est ce que suggérait déjà Valdman en 1996 (p. 2) : « La présentation de documents authentiques, qui sert d'ancrage aux approches actuelles de l'enseignement du FLE ne requiert-elle pas que l'on fasse écouter les voix réelles de la francophonie aux apprenants sans toutefois – et j'insiste là-dessus – leur demander de les imiter […] Si l'équilibre entre les pôles acceptabilité et apprenabilité est parfois difficile à établir à un stade donné de l'apprentissage, c'est bien sur l'axe de la progression d'enseignement / apprentissage que se trouve la solution ». (Detey 2010, p. 162‑163)

Dans cette perspective, et notamment dans le cadre du projet « Phonologie du Français Contemporain – Enseignement du Français » (dorénavant PFC-EF), Tchobanov et al. (2009) ont proposé des activités d'apprentissage de la liaison en s'appuyant sur les données linguistiques extraites du corpus PFC (Durand, Laks et Lyche 2009). Plus précisément, ils ont mis au point deux types de fiches pédagogiques. L'une (la fiche mono-extrait), basée sur des extraits particuliers et abordant plusieurs phénomènes linguistiques et l'autre (fiche multi-extraits), visant l'étude d'un phénomène linguistique en particulier à l'aide de plusieurs extraits. Le premier type d'activité a pour objectif le développement des compétences de compréhension orale et propose ensuite une activité de relecture. En revanche, la fiche multi-extraits est conçue pour l'étude d'un point de grammaire ou phonologie (comme la liaison) spécifique, de manière inductive et progressive, à travers la présentation d'une sélection d'exemples de conversations. Ces matériaux didactiques adaptables pour l'enseignement de la liaison semblent pouvoir être associés aux ressources pédagogiques traditionnelles pour sensibiliser l'apprenant à la variabilité des usages de la liaison. En effet, bien que les classifications de liaison favorisent une analyse explicite des types de liaison, le recours exclusif à ces matériaux associé à une

exposition réduite à l'input oral (surtout en contexte hétéroglotte), risquent de rendre stérile cette modalité d'apprentissage. De plus, comme nous l'avons observé au début de ce travail (cf. Partie 1.Chapitre 4.4), les types de liaison identifiés sur base morphosyntaxique par les tripartitions traditionnelles présentent une forte hétérogénéité interne liée à la variabilité de réalisation de la liaison. Dans un cadre didactique, comme le souligne Harnois-Delpiano et al. (2012), une telle approche pourrait amener les apprenants à traiter la liaison variable comme relevant d'une norme catégorique à apprendre.

3. Conclusion

En tant qu'indicateur sociolinguistique, la liaison représente un aspect crucial de l'acquisition du FLE. Néanmoins, l'hétérogénéité et la variabilité qui caractérisent ce phénomène de *sandhi* externe constituent une source de complexité multiple pour l'apprentissage et l'enseignement de la liaison.

Comme le souligne Racine et Detey (sous presse), la plupart des manuels didactiques renvoient le plus souvent à une description prescriptive et normative des usages de la liaison (Fouché 1959 ; Delattre 1947) qui fait recours à des tableaux ou à des listes de règles construites sur la base de critères morphosyntaxiques. Néanmoins, dans certains manuels (Lauret 2007 ; Pustka 2011), la description de la liaison a été établie sur la base des résultats des études menées sur les corpus oraux. Il convient pourtant de souligner que malgré les avantages indéniables (Chambers 2009 ; Tyne 2009 ; Tchobanov et al. 2009 ; Detey 2010 ; Detey et Racine 2012), le recours aux corpus dans la didactique multiplie les variétés d'usages de la liaison ravivant le débat sur la notion de français de référence déjà évoqué dans la première partie de ce travail (cf. Partie 1.Chapitre 3). Dans ce cadre, nous comprenons pourquoi les descriptions prescriptives représentent une tentative de « mettre de l'ordre » (Fouché 1959). Elles assurent en effet un traitement didactique de la variation de la liaison limité aux usages de l'élite cultivée parisienne. Bien que plus économique, une telle approche renonce à tout enseignement de la variation sociolinguistique associée à la liaison. Detey (2010) et Detey et Racine (2012) estiment qu'un seul modèle de prononciation doit être présenté dans un premier stade de l'apprentissage afin d'accorder une primauté à un input unique et « standard » qui sera par la suite enrichi par l'apport des usages issus des autres variétés. L'apprenant sera ainsi confronté et sensibilisé à une variété sociolinguistique qui favorise le développement de son « acceptabilité » sociolinguistique.

Ces propositions nous paraissent apporter des réponses adéquates aux besoins de la didactique et de l'acquisition de la liaison en FLE, en permettant de sensibiliser les apprenants aux variétés d'usage de la liaison d'une part, et d'éviter un décalage entre les modèles de référence pour l'analyse des productions des apprenants et les usages réels des locuteurs de l'autre.

C'est dans cette démarche que se situent les travaux menés au sein du projet international « Interphonologie du Français Contemporain » (IPFC), parmi lesquels la présente étude qui a été conduite auprès d'apprenants italophones de FLE de niveau intermédiaire. De plus, comme nous le verrons dans la section suivante de ce travail, le recours à un protocole commun permet d'assurer une comparaison entre les études menées auprès de populations différentes d'apprenants (plusieurs langues sources). Nous espérons qu'une telle approche méthodologique apportera des réponses aux questions soulevées dans le chapitre précédent (cf. Partie 3.Chapitre 2.6) à propos des processus d'acquisition et des typologies d'erreur de la liaison en L2.

.

Partie 4. UNE ETUDE LONGITUDINALE DE L'ACQUISITION DE LA LIAISON CHEZ DES APPRENANTS ITALOPHONES DE FRANÇAIS LANGUE ETRANGERE (FLE) : L'ENQUETE IPFC-ITALIEN (ROME)

Chapitre 1. L'enquête « Interphonologie du Français Contemporain - italien » (IPFC-italien) (Rome) : méthodologies, corpus et traitement des données

L'acquisition de la liaison en L2 a fait l'objet d'un nombre relativement limité de travaux (cf. Partie 3.Chapitre 2.1) pour la plupart consacrés à l'étude des aspects perceptifs et sociolinguistiques. Comme nous l'avons déjà montré les spécificités méthodologiques (objectifs, protocoles), sociales (milieu d'apprentissage) et linguistiques (niveau des apprenants) de ces études ainsi que la forte variabilité inter-locuteurs limitent les possibilités de comparaison entre les résultats de ces différentes études. Ces aspects font qu'une identification d'un itinéraire acquisitionnel de la liaison n'est pas actuellement envisageable. Une telle analyse, comme nous l'avons vu dans le cas du projet ESF (cf. Partie 3.Chapitre 1.4.3.1.3), nécessite une large variété de données, possiblement multilingues, qui permettent de vérifier les généralisations formulées au sein de plusieurs langues (typologiquement différentes). Il existe aujourd'hui un certain nombre de corpus de français L2 (le corpus FLLOC, les corpus LANCOM et le corpus ESF) qui sont consacrés et conçus principalement pour l'analyse lexicale, syntaxique, morphosyntaxique et pragmatique de l'interlangue des apprenants. Dans ce cadre le projet « InterPhonologie du Français Contemporain » (IPFC) (Detey et Kawaguchi 2008 ; Racine et al. 2010 ; Racine, Detey et Kawaguchi 2012) repond aux besoins de disposer d'un corpus diversifié de français L2 multilingue (plusieurs L1) consacré à l'étude de la liaison et plus en général de l'interphonologie du français (Detey 2005).

Dans la quatrième partie de ce travail, nous allons donc brièvement parcourir les études de corpus linguistiques menées en français L2, ce qui nous conduira à présenter le projet « InterPhonologie du Français Contemporain » (IPFC), notamment ses objectifs et ses méthodologies. C'est dans ce cadre que s'inscrit le travail de l'équipe italienne du « projet InterPhonologie du Français Contemporain » (IPFC-italien) qui compte actuellement quatre points d'enquête (Milan, Trieste, Rome et Venise). Le présent travail est le résultat de l'étude longitudinale d'une année menée à l'Université Sapienza de Rome auprès de 17 étudiants italophones de FLE inscrits en première année de licence. Cette recherche a donné lieu à la constitution d'un corpus dont les caractéristiques, les méthodologies (cf. Paragraphe 2.3) et les résultats (cf. Partie 4.Chapitre 1.4) feront l'objet des prochains paragraphes.

1. Corpus oraux en L2

Les travaux sur la linguistique du corpus en L2 ont connu récemment un nouveau développement. Comme nous l'avons montré dans le chapitre (cf. Partie 3.Chapitre 1.4.3.1), plusieurs projets (ZISA, ESF) avaient abouti dans les années 80 à la constitution de larges corpus qui ont permis d'identifier les premiers itinéraires acquisitionnels dans les domaines de la syntaxe et de la morphosyntaxe de plusieurs langues.

Au cours de ces dernières années, de nouveaux corpus en L2 ont été élaborés notamment auprès de différentes populations d'apprenants (plusieurs L1) de français (Debrock et Flament-Boistrancourt 1996 ; Myles et Mitchell 2007), de néerlandais (Neri, Cucchiarini et Strik 2006), de polonais (Cylwik, Wagner et Demenko 2009), d'allemand (Gut 2009) et d'anglais (Neumeyer, Franco, Wintraub et Prince 1996 cité par Gut 2009 ; Gut 2009 ; Visceglia et al. 2009; Gilquin, De Cock et Granger 2010).

De nombreux corpus parmi ces travaux ont été recueillis pour servir de bases de données pour l'entraînement de système d'ingénierie de la parole (comme le corpus VITALS *Voice Interactive Language Training System*)[261] ou d'enseignement assisté de la prononciation. L'étude de Neri et al. (2006) se propose en particulier de définir un inventaire des erreurs de prononciation réalisées par des apprenants de néerlandais de plusieurs L1 en vue de l'élaboration d'un système CAPT (*Computer Assisted Pronunciation Training systems*). Il en va de même pour le travail de Cylwik et al. (2009) consacré à l'étude des erreurs de prononciation dans l'interlangue des apprenants germanophones de polonais L2 servant à alimenter un système d'apprentissage et d'examen assisté.

Des études plus systématiques ont été également réalisées sur une large échelle surtout en anglais L2. C'est le cas par exemple des projets internationaux *Louvain Internation Database of Spoken English Interlanguage* (LINDSEI) (Gilquin, De Cock et Granger 2010) ou *Asian English Speech cOrpus Project* (AESOP) (Visceglia et al. 2009) ayant pour objectif la constitution d'un corpus oral de variétés d'anglais parlées en Asie.

Dans le domaine du français L2, mis à part le corpus ESF (Perdue 1993), nous pouvons citer le corpus *French Learner Oral Corpus* (FLLOC) (Myles et Mitchell 2007) comprenant des transcriptions de conversations et de récits d'apprenants anglophones ou encore le corpus LANCOM3 (Debrock et Flament-Boistrancourt 1996). Cependant ces corpus présentent des

[261] Le corpus *Voice Interactive Language Training System* (VILTS) (Neumeyer et al. 1996 cité en Gut 2009) est un corpus recueilli auprès de 100 apprenants anglophones de français L2 qui a servi comme corpus pour l'entraînement de la reconnaissance vocale d'un système pour l'enseignement de la prononciation française.

transcriptions adaptées le plus souvent à des recherches sur les aspects lexicaux, syntaxiques et pragmatiques de l'acquisition de la L2. Ces ressources, malgré leur richesse, se prêtent difficilement à l'étude des spécificités phonétiques et phonologiques de l'interlangue de l'apprenant. Dans ce sens, le projet « InterPhonologie du Français Contemporain » (IPFC) (Detey et Kawaguchi 2008 ; Racine et al. 2010 ; Racine, Detey et Kawaguchi 2012) suivi par le corpus COREIL (Delais-Roussarie et Yoo 2010), le corpus de Pillot-Loiseau et al.(2010) et le corpus PhoDIiFLE (Landron et al. 2010) ont été élaborés afin de rendre compte des aspects phonético-phonologiques d'acquisition du français L2.

2. Le projet InterPhonologie du Français Conteporain (IPFC)

Le projet « InterPhonologie du Français Conteporain » (dorénavant IPFC) (Detey et Kawaguchi 2008 ; Racine et al. 2010; Racine, Detey et Kawaguchi 2012)[262], piloté par Sylvain Detey (Université Waseda), Isabelle Racine (Université de Genève) et Yuji Kawaguchi (Tokyo University of Foreign Studies) a donné lieu aujourd'hui à la première ressource de français L2 à visée phonologique.

2.1. Objectifs

L'objectif principal du projet IPFC est de constituer et de diffuser une base de données orales de français L2 issue de travaux menés auprès d'apprenants de multiples L1 (en contexte hétéroglotte ou homoglotte) et consacrée à l'analyse des aspects phonético-phonologiques de l'acquisition du français. L'objet d'étude de ce corpus est donc précisément l'interphonologie (Detey 2005) à savoir le système phonético-phonologique des apprenants d'une L2 en cours de construction ou dans un état stabilisé afin d'identifier les aspects qui doivent être traités en priorité au niveau de l'enseignement de la prononciation. Pour être plus précis, le projet IPFC s'intéresse à l'analyse des difficultés segmentales majeures dans l'acquisition du français L2, notamment l'apprentissage des voyelles nasales et des voyelles arrondies ainsi qu'à la réalisation de certains phénomènes prosodiques et phonologiques spécifiques du français (tels que la liaison, l'enchaînement et l'effacement du *schwa*, etc.). Le projet IPFC se propose également de développer des activités autour de la prononciation du français et de son apprentissage, et de promouvoir l'image du français comme langue internationale.

[262] (http://cblle.tufs.ac.jp/ipfc/).

2.2. Points d'enquête

Le projet IPFC est un projet international qui regroupe des chercheurs en phonétique et phonologie du français L2 issus de différents pays. Il compte à ce jour 15 L1 différentes :

- allemand (Osnabrück, Munich, Vienne, Zürich) ;
- anglais (Ontario) ;
- danois (Copenhague) ;
- espagnol (Genève et Madrid) ;
- grec (Poitiers et Chypre / Thessalonique) ;
- italien (Milan, Trieste, Rome, Venise et Tessin) ;
- japonais (Tokyo) ;
- néerlandais (Nijmegen et Groningen) ;
- norvégien (Tromsø et Oslo) ;
- portugais (Brésil, Santa Catarina) ;
- russe (Moscou) ;
- suédois (Dalarna) ;
- turc (Marmara);
- arabe (France-Maroc-Liban-Palestine);
- coréen (Séoul).

Les deux premiers corpus constitués sont d'une part le corpus d'apprenants japonophones (IPFC-japonais) et d'autre part le corpus d'apprenants hispanophones (IPFC-espagnol).

2.3. Protocole

La comparabilité des données entre les différentes populations d'apprenants est assurée par un protocole de recueil de données multitâches commun à toutes les L1[263] calqué sur celui du projet « Phonologie du Français Contemporain » (PFC) (cf. Partie 2.Chapitre 1.1.3.1) (Durand, Laks et Lyche 2009). Ce dernier aspect permet d'analyser l'interphonologie des apprenants en faisant référence non pas aux normes proposées par les manuels de prononciation (c.f. Partie 3.Chapitre 3) mais aux variétés sociolinguistiques issues d'une vaste base de données de production de natifs.

[263] cf. Paragraphe 2.2.

Le protocole IPFC comprend 6 tâches réparties en deux séances de passation : (a) répétition d'une liste spécifique de mots (lue par un locuteur francophone natif); (b) lecture de cette même liste spécifique ; (c) lecture de la liste de mots PFC ; (d) lecture du texte PFC (cf. Partie 2.Chapitre 1.1.3.1)[264]; (e) entretien avec un locuteur natif (entretien guidé) ; (f) interaction semi-contrainte entre deux apprenants (conversation libre).

La première session de passation, identique pour tous les niveaux d'apprenants, comprend la lecture de la liste de mots PFC (c) et du texte déjà utilisé dans le protocole PFC (d) (cf. Partie 2.Chapitre 1.1.3.1). Ces tâches sont suivies par la répétition et la lecture d'une liste spécifique de mots élaborée par les équipes de chaque population d'apprenants (a) (cf. Annexe 4).

La deuxième séance de passation comprend deux tâches de conversation. D'une part l'entretien guidé avec un natif (e) qui s'appuie sur une liste de questions visant à retracer la biographie linguistique de l'apprenant. Selon le niveau des apprenants (A1-B1 et B2-C1 du Cadre Européen Commun de Référence pour les Langues, CECRL) et le type d'apprentissage (homoglotte ou hétéroglotte), l'interviewer peut décider de poursuivre la conversation suivant des questions plus ouvertes.

D'autre part, si le niveau de l'apprenant le permet, une deuxième tâche de conversation libre (f) peut être réalisée entre deux apprenants portant sur un thème choisi par chacun d'entre eux (un film vu, un livre lu, un endroit visité, un souvenir, un thème d'actualité etc.).

2.4. Traitement des données

Les données issues de ce protocole multitâche sont transcrites par le biais du logiciel Praat (Boersma et Weenink 2015) sur une ligne principale orthographique accompagnée d'autres couches d'annotations et de codages en fonction du phénomène observé (liaison, voyelles nasales etc.).

2.4.1. Transcription

La transcription, comme le font observer Racine et al. (2011) à la suite de Cappeau (2008), est une étape extrêmement délicate du traitement d'un corpus oral pour différentes raisons. En effet, elle ne constitue pas seulement le moyen d'accès aux données orales mais

[264] Par rapport au projet PFC deux autres sites de liaison erratique (entre « région » et « en » ainsi qu'entre « opposition » et « aurait » (cf. Annexe 4) ont été codés dans le projet IPFC dans la mesure où « le codage vise à décrire sans *a priori* les réalisations des apprenants, nous avons pris le parti de considérer comme un site potentiel de liaison tout mot se terminant par une consonne graphique habituellement non prononcée situé devant un mot commençant par une voyelle ou un h muet » (Racine et Detey sous presse).

présuppose également une première analyse théorique et méthodologique des données recueillies.

> « En participant à la construction des faits analysés, la transcription agit comme révélateur de la distance qui peut intervenir entre les productions effectives et les formes langagières interprétées ou comprises. Toute transcription n'est ainsi qu'une représentation limitée de la production orale, qui en efface nombre de paramètres non segmentaux, gestuels, contextuels ». (Racine et al. 2011, p. 15)

La question de la fidélité de la transcription est d'autant plus importante lorsqu'il s'agit de traiter des données phonético-phonologiques issues des productions des apprenants. Dans ce cas Racine et al (2011) suggèrent d'adopter une notation orthographique standard au lieu d'une transcription phonétique manuelle basée sur l'Alphabet Phonétique International (API) qui se révèle d'ailleurs extrêmement coûteuse à long terme. Ce choix relève de la constatation de l'impossibilité d'éviter toute interprétation même inconsciente des réalisations « déviantes » des apprenants de la part du transcripteur[265]. En effet celui-ci aurait involontairement tendance à reconstruire perceptivement les productions des non-natifs en L2 à partir de son propre système phonético-phonologique et serait donc incapable de saisir les propriétés de l'interlangue de l'apprenant. Une analyse pertinente de l'interlangue, comme le soulignent Racine et al (2011) à la suite de Wester et al. (2001 cité par Racine et al. 2011), doit en outre faire appel à des transcripteurs bilingues ou qui tout au moins connaissent le système phonologique de la L1 de l'apprenant afin de limiter les effets de filtrage.

2.4.1.1 Traitement des formes ambiguës

Si l'adoption d'une transcription orthographique permet de limiter le degré d'interprétation de la part du transcripteur, il existe néanmoins des cas d'ambiguïté où les réalisations de l'apprenant, bien qu'elles soient phonétiquement identifiables, se prêtent à diverses interprétations. Le choix de la stratégie à adopter dans ces contextes doit être défini et explicité en tenant compte du type de tâche réalisée (lecture ou parole spontanés) (Racine et al. 2011). Dans les cas d'ambiguïté des séquences « déviantes » enregistrées dans des tâches de lecture, les auteurs soulignent que là où la connaissance préalable du système phonético-phonologique de l'apprenant permet d'exclure une interprétation possible, le transcripteur peut opérer un choix qui doit par conséquent rester cohérent dans l'ensemble du corpus.

En revanche, dans des tâches de production de parole spontanée, cette approche est plus audacieuse. Il serait donc plus prudent d'avoir recours à des multi-transcriptions de séquences

[265] L'étude de Zechner (2009 cité par Racine et al. 2011) montre un écart important entre le taux de désaccord entre les transcripteurs des données natives (5%) et ceux de données non natives (10-34%).

« déviantes » ou incertaines afin de préserver l'ambigüité entre encodage morphologiquement correct mais phonétiquement approximatif, et réalisation phonétiquement bien ciblée mais morphologiquement ou grammaticalement erronée[266].

2.4.2. Codage liaison

Après avoir été transcrites orthographiquement à l'aide du logiciel Praat (Boersma et Weenink 2015), les données sont accompagnées de plusieurs couches d'annotation et de codage en fonction du nombre de phénomènes analysés. En ce qui concerne l'annotation de la liaison, elle se fait par le biais d'un code alphanumérique qui s'articule sur différents champs d'annotation descriptive et perceptive. On dénombre un total de 7 champs du codage liaison.

Les quatre premiers sont purement descriptifs et permettent de décrire les caractéristiques phonético-syntaxiques du contexte de liaison.

- Champ 1: nature de la consonne de liaison cible ;
- Champ 2: catégorie syntaxique du mot liaisonnant (à savoir M1) ;
- Champ 3: catégorie syntaxique du mot qui suit (à savoir M2);
- Champ 4: nombre de syllabes et nature de la voyelle du mot liaisonnant (à savoir M1).

En revanche, les trois derniers champs portent sur une analyse perceptive des traits phonético-prosodiques de réalisation de la liaison :

- Champ 5: réalisation de la liaison ;
- Champ 6: nature de la consonne de liaison ;
- Champ 7: présence d'une pause / hésitation.

Le codage comporte donc une séquence d'un nombre minimum de 7 symboles alphanumériques dont l'ordre de codage est fixe comme dans l'exemple ci-dessous (cf. Exemple 11).

[266] Dans l'exemple proposé par Racine et al. (2011) « les gens […] ils habitent […] ils ne connaient pas / il ne connaît pas », la séquence phonétiquement non ambiguë « ils ne connaient pas / il ne connaît pas » peut recevoir une double interprétation. « Soit l'apprenant conserve le même référent pluriel (les gens… ils… ils…) et on peut considérer qu'il régularise le présent du verbe 'connaître' sur le modèle du verbe 'habiter' […]. Soit on envisage avec la deuxième transcription qu'il y a eu passage inopiné d'un référent pluriel à un référent singulier (correspondant à un collectif) » (Racine et al. 2011, p. 18).

« Qu'est-ce qui a donc valu à Beaulieu ce grand30_ADJ_NOM_100_11_2_0 honneur ? ».
(Corpus IPFC-italien Rome, Dbr S3_T4)

Exemple 11 : Exemple de séquence de liaison codée selon le codage alphanumérique liaison IPFC

Cette séquence de codage décrit un contexte de liaison en [t] au singulier (champ 1 : code 30) entre un adjectif (M1) (champ 2 : code ADJ) et un nom suivant (M2) (champ 3 : code NOM) où le M1 est monosyllabique (champ 4 : code 100). Quant aux aspects perceptifs, dans l'exemple ci-dessus (cf. Exemple 11) la liaison est réalisée avec enchaînement (champ 5 : code 11) mais en utilisant une CL non conforme à la cible avec différence phonétique, notamment [d] (champ 6 : code 2) et elle n'est ni suivie ni précédée de pause, coup de glotte ou hésitation (champ 7 code 0).

À l'intérieur de chaque champ, l'attribution des codes alphanumériques permet donc de préciser les caractéristiques des différents éléments susceptibles d'apparaître dans ce contexte. La richesse du codage de chaque champ varie ainsi en fonction de la complexité du phénomène codé.

2.4.2.1 Champs 1 : nature de la CL cible

Le premier champ définit la nature de la CL cible et s'articule en 9 cases

- 10 = [z] singulier ex. « gros insecte » ;
- 11 = [z] pluriel ex. « des insectes » ;
- 20 = [n] ex. « un ami » ;
- 30 = [t] singulier ex. « petit ours / est allé » ;
- 31 = [t] pluriel ex. « sont allés » ;
- 40 = [ʀ] ex. « premier élève » ;
- 50 = [p] ex. « trop intéressant » ;
- 60 = [g] ex. « long hiver » ;
- 70 = pas de consonne cible Ex. « si évident » = « si[t]évident ».

Ce dernier identifiant numérique, comme nous le verrons, permet de coder les cas de liaisons non attendues (cf. Partie 1.Chapitre 4.2.4).

Dans le cas de l'exemple ci-dessus (cf. Exemple 11), la nature de la CL cible du mot « grand » correspond à la consonne [t].

2.4.2.2 Champs 2 et Champs 3 : catégories syntaxiques du mot liaisonnant (M1) et du mot qui suit (M2)

Les codes alphabétiques qui suivent correspondent aux étiquettes morphosyntaxiques attribuées au mot liaisonnant (M1) et au mot qui suit (M2) et donc valables à la fois pour le champ 2 et pour le champ 3.

- ABR = abréviation ;
- ADJ = adjectif ;
- ADV = adverbe ;
- AUX = verbes auxiliaires ;
- CON = conjonction ;
- DET = déterminant (y compris adjectif démonstratif et possessif) ;
- EXF = élément appartenant à une expression figée (« de temps en temps », « de plus en plus », « Etats-Unis ») ;
- INF = verbe à l'infinitif ;
- INT = interjection ;
- NAM = nom propre ;
- NEG = négation ;
- NOM = nom ;
- NUM = numéral ;
- PDE = préposition - article (« au », « du », « aux », « des ») ;
- PPA = participe passé ;
- PPR = participe présent ;
- PRO = pronom ;
- PRP = préposition ;
- VER = autres verbes (sans distinction de temps ou de mode etc.) ;
- XXX = indéfinissable (faux départs, hésitations, etc.).

La séquence de liaison « grand émoi » que nous avons présentée dans l'Exemple 11 est décrite par les bais des étiquettes ADJ et NOM qui définissent respectivement le contexte morphosyntaxique gauche et droit de liaison à savoir le type de liaison.

2.4.2.3 Champ 4 : Nombre de syllabes et nature de la voyelle du mot liaisonnant (M1)

Les codes numériques qui suivent permettent de préciser le nombre de syllabes des tous les M1 ne se terminant pas par CL [n].

- 100 = monosyllabe ex. « de plus en plus », « un homme » ;
- 200 = polysyllabe ex. « petit ours », « nombreux amis ».

Dans les contextes de liaison en [n] des éléments alphabétiques intègrent les codes numériques ci-dessus afin de préciser la nature de la voyelle du M1 : VO si la voyelle est orale et VN si elle est nasale. Autrement dit, 6 combinaisons sont envisageables.

- 1VO = monosyllabe avec liaison possible en [n] (réalisée ou non) précédée d'une voyelle orale ex. « en effet » = « [a][n]effet » ou « [a]#effet » ;
- 1VN = monosyllabe avec liaison en [n] (réalisé ou non) précédée d'une voyelle nasale ex. « en effet » = « [ã][n]effet » ou « [ã]#effet » ;
- 1VX = monosyllabe avec liaison possible en [n] (réalisée ou non) précédée d'une voyelle orale - nasalisation ou incertain ;
- 2VO = polysyllabe avec liaison possible en [n] (réalisée ou non) précédée d'une voyelle orale ex. « aucun ami » = « auc[œ][n]ami » ou « auc[œ]#ami » ;
- 2VN = polysyllabe avec liaison possible en [n] (réalisée ou non) précédée d'une voyelle nasale ex. « aucun ami » = « auc[œ̃][n]ami » ou « auc[œ̃]#ami » ;
- 2VX = polysyllabe avec liaison possible en [n] (réalisée ou non) précédée d'une voyelle orale - nasalisation ou incertain.

2.4.2.4 Champ 5 : Réalisation de la liaison

En revanche, le champ 5 permet d'identifier l'absence de liaison ou la réalisation enchaînée ou non-enchaînée de la liaison sans préciser les aspects phonétiques de réalisation de la CL.

- 00 = pas de liaison ex. « grand ami »= grand#ami ;
- 11 = liaison réalisée avec enchaînement ex. « grand ami » = grand[t]ami ;
- 12 = liaison réalisée sans enchaînement ex. « grand ami » = grand[t]#ami ;
- 13 = liaison réalisée avec incertitude quant à l'enchaînement (si réalisée ou non).

2.4.2.5 Champ 6 : Type de CL réalisée

Cette partie du codage assure la définition de la typologie de la CL réalisée permettant d'identifier l'ensemble des productions que nous avons évoquées dans la troisième partie de ce travail (cf. Partie 3.Chapitre 2.1.4) notamment (en ordre d'apparition) : absence de liaison (0) ; réalisations conformes (1) ; réalisations orthographiques de la CL (2) ; réalisations d'une consonne orthographique présente dans le M1 (3); substitutions de la CL ou liaisons non attendues (4).

- 0 =si la liaison n'est pas réalisée ;
- 1 = CL conforme à la cible ex. « de temps en temps » = de temps [z] en temps (avec ou sans enchaînement) ;
- 2 = CL presque conforme à la cible mais avec différence phonétique ex. « de temps en temps » = de temps [s] en temps ;
- 3 = CL non conforme à la cible mais avec consonne présente dans le mot ex. « de temps en temps » = de temps [p] en temps ;
- 4 = CL non conforme à la cible avec consonne épenthétique (à savoir les substitutions et les liaisons non attendues) (cf. Partie 3.Chapitre 2.1.4) ex. « gros enfant » = gros [t] enfant

  ex. « il va à Paris » = il va [t] à Paris.

2.4.2.6 Champ 7 : Présence d'une pause / hésitation / coup de glotte

Ce dernier champ contient seulement deux codes numériques qui signalent l'absence ou la présence d'une pause / hésitation / coup de glotte entre le M1 et le M2 indépendamment de la place que ces discontinuités occupent par rapport à la CL.

- 0 = pas de pause / hésitation / coup de glotte entre le mot liaisonnant (M1) et le mot qui suit (M2) ;
- 1 = pause / hésitation / coup de glotte entre le mot liaisonnant (M1) et le mot qui suit (M2) (si la liaison est réalisée, peu importe où pause / hésitation / coup de glotte sont situés, avant ou après la CL) ; ex. « mais après » = mais#après ; mais[z]#après ; mais#[z]après.

En résumé, la combinaison de ces 7 champs du codage de la liaison assure l'analyse des principaux facteurs syntaxiques, phonologiques et prosodiques qui caractérisent les contextes de réalisation de la liaison.

3. L'équipe IPFC-italien

C'est dans le cadre du projet IPFC que s'inscrit le travail de l'équipe italienne dirigée par Enrica Galazzi et à laquelle participent les enquêteurs suivants : Giulia Barreca, Cristina Bosisio, Caterina Falbo, Oreste Floquet, Marie Christine Jamet, Pascale Janot, Chiara Molinari, Michela Murano et Roberto Paternostro.

3.1. Points d'enquête

Cinq terrains font à ce jour l'objet de différentes études menées auprès d'une population de locuteurs aux profils diversifiés.

Les premiers points d'enquête ont été établis à l'Université Catholique du Sacré Cœur de Milan, Faculté des Sciences du Langage et Littératures Étrangères (Enrica Galazzi, Cristina Bosisio, Chiara Molinari, Michela Murano et Roberto Paternostro) et à l'université de Trieste, département des Sciences Juridiques, du Langage, de l'Interprétation et de la Traduction, Section d'Études en langues Modernes pour Interprètes et Traducteurs (Caterina Falbo et Pascale Janot). Depuis décembre 2012, un troisième terrain de recherche a été développé à l'Université de Rome Sapienza, département d'études européennes, américaines et interculturelles (Giulia Barreca, Oreste Floquet et Carolina Lombardo), suivi en 2015 de celui de l'Université de Venise (Marie Christine Jamet) et Tessin (Roberto Paternostro).

Ces enquêtes portent dans la plupart des cas sur l'analyse des phénomènes spécifiques de la phonologie du français, notamment l'analyse des voyelles nasales et de la liaison.

Le travail de l'équipe de Milan (Galazzi et al. 2013 ; Falbo et al. sous presse), consacré à l'étude de la liaison, a été conduit auprès de 12 étudiants italophones de niveau débutant, intermédiaire et avancé inscrits en 1ère année de Licence en Sciences Linguistiques et Littératures Étrangères (11 filles et 1 garçon, de 19 à 21 ans). Pour ces sujets, l'apprentissage du français a toujours eu lieu en milieu institutionnel et hétéroglotte, hormis de courts séjours (1-2 semaines) dans des pays francophones pour des vacances ou des programmes d'échange de classes.

Les recherches de l'équipe de Trieste (Galazzi et al. 2013 ; Falbo et al. sous presse) portent également sur l'étude de la liaison mais visent une population différente de sujets à

savoir 12 étudiants italophones de niveau avancé inscrits en Master 2 d'Interprétation de conférence. Ces apprenants ont le français comme langue B, langue étrangère vers laquelle ils doivent traduire à partir de l'italien, en interprétation consécutive (IC) et en interprétation simultanée (IS). L'objet de ce travail est donc de comparer les liaisons réalisées / non réalisées pendant les tâches prévues par le protocole IPFC et les tâches d'IC et d'IS. Les chercheuses se proposent de vérifier l'hypothèse selon laquelle la production en français pendant l'IC et l'IS pourrait donner lieu à une présence plus importante d'erreurs de liaison.

Les résultats de ces deux études[267] (cf. Partie 3.Chapitre 2.2) ont permis de montrer que la réalisation de la liaison dans certains contextes catégoriques (ex. « déterminant - nom » ; « pronom personnel sujet - verbe ») semble être maitrisée par les locuteurs italophones alors que dans d'autres elle est encore sujette à variation. Dans le cas de l'étude de Trieste, aucune différence significative n'a été enregistrée entre les tâches de lecture et les deux entretiens d'une part, et les deux interprétations d'autre part. La variation stylistique requise par le passage aux interprétations de style plus soutenu n'est donc pas acquise.

Les deux recherches pilotes sont actuellement en cours sur un ensemble plus important de sujets et sur des populations diversifiées d'apprenants[268].

4. L'enquête IPFC-italien (Rome)

L'enquête IPFC-italien (Rome) (Barreca et Floquet 2015) a été lancée en décembre 2012 auprès de l'Université Sapienza de Rome et se propose de suivre l'évolution de l'acquisition de la liaison en FLE en milieu institutionnel en s'appuyant sur le protocole du projet IPFC. Nous avons ainsi mené une étude longitudinale en quatre temps espacés de quatre mois auprès de 17 étudiants italophones de niveau intermédiaire (A2-B1).

Le tableau ci-dessous (cf. Tableau 31) résume les informations concernant les dates des 4 séances d'enregistrement de l'étude longitudinale.

Séance	Code Séance	Dates
1ère séance	S1	9-17 mai 2013
2ème séance	S2	24-30 sept 2013
3ème séance	S3	20-24 janv 2014
4ème séance	S4	12-16 mai 2014

Tableau 31 : Séances de l'étude longitudinale IPFC-italien (Rome)

[267] L'équipe de Venise a intégré le projet en mars 2015.

[268] En particulier, l'équipe de Milan est en train de mener une nouvelle enquête sur un groupe de futurs enseignants de FLE.

L'approche longitudinale a été adoptée afin d'essayer d'esquisser les étapes évolutives de l'acquisition de la liaison chez les apprenants de FLE. En effet, comme nous l'avons déjà observé à plusieurs reprises, aucun itinéraire acquisitionnel ou stade de développement (Bartning et Schlyter 2004) n'a été proposé à ce jour pour décrire l'apprentissage de la liaison en français langue seconde ou étrangère.

4.1. Objectifs

Les objectifs de cette étude longitudinale sont multiples. Dans un premier temps, nous nous proposons de suivre l'évolution de l'acquisition de la liaison en FLE et vérifier les capacités des apprenants à établir des généralisations. Nous nous attacherons, dans un deuxième temps, à analyser l'influence de la L1 de l'apprenant italophone sur la capacité de resyllabation de la CL ainsi que l'impact de l'input écrit en milieu hétéroglotte sur la réalisation de la liaison et notamment sur la production des formes orthographiques de la CL. Il s'agira par la suite d'étudier dans quelle mesure l'apprenant de FLE est conscient de la variation de la liaison. Pour conclure, dans une dernière partie de ce travail (cf. Partie 5), en nous appuyant sur l'analyse de liaison issue des données des natifs (corpus PFC) (cf. Partie 2.Chapitre 2.2.2.1), nous proposerons des ressources pédagogiques pour l'enseignement de la liaison dont l'objectif est de fournir une riche description de la variation de la liaison.

4.1.1. Objectif 1 : Développement de représentations phonologiques de la CL

Comme nous l'avons déjà évoqué, nous proposons de suivre l'apprentissage de la liaison en FLE chez les apprenants italophones dans le but de définir des régularités d'acquisition (Ellis 1994, p. 111) (cf. Partie 3.Chapitre 1.4.3). Une telle approche nous permettra également d'évaluer l'évolution des erreurs de liaison et notamment de vérifier la présence de liaisons non-enchaînées chez les apprenants italophones. De plus à partir de l'étude de l'ensemble des réalisations, nous espérons pouvoir apporter des éléments de réponse au débat concernant la capacité des apprenants d'opérer des généralisations et donc de développer des représentations phonologiques de la CL.

4.1.1.1 Hypothèse 1

Les résultats des études précédentes sur l'acquisition de la liaison (cf. Partie 3.Chapitre 2.3) ne permettent pas d'exclure la capacité des apprenants à formuler des généralisations phonologiques. De plus le taux important de liaisons réalisées dans certains contextes de liaisons catégoriques (ex. liaison après les articles et entre le pronom sujet et le verbe) suggèrent

la présence chez l'apprenant de représentations phonologiques qui assurent la productivité de la liaison.

L'acquisition de la liaison ne se réduirait donc pas à une simple mise en place de stratégies lexicales à savoir à un traitement « au cas par cas » de la liaison. Contrairement à Wauquier (2009), nous avançons l'hypothèse que l'apprenant à partir d'un niveau intermédiaire est capable d'opérer des généralisations concernant la réalisation de la liaison dans des contextes de liaison catégorique.

4.1.2. Objectif 2 : Influence de la L1 dans l'acquisition de la liaison

4.1.2.1 Hypothèse 2

Comme nous l'avons vu précédemment (cf. Partie 3.Chapitre 2.4), selon des hypothèses formulées par les études précédentes (Thomas 2002 ; Tennant sous presse ; Wauquier 2009), les liaisons non-enchaînées seraient majoritairement attestées dans les études menées auprès d'apprenants ayant comme langue source une L1 pour lesquels « la congruence entre les frontières lexicales et les frontières syllabiques est beaucoup plus forte que pour le français » (Wauquier 2009), comme par exemple les langues germaniques (anglais, allemand etc.). Or l'italien est une langue caractérisée par des phénomènes de *sandhi* externe (cf. Partie 1.Chapitre 1.1) (tels que l'élision, l'apocope, la troncation des voyelles ou encore l'assimilation des consonnes[269]) (Serianni et Castelvecchi 1988 ; Nespor 1993) qui provoquent des resyllabations des éléments vocaliques et consonantiques à la frontière entre deux unités lexicales. Nous devrions donc nous attendre selon l'hypothèse proposée ci-dessus (Thomas 2002 ; Tennant sous presse ; Wauquier 2009) à ce que la faible rigidité entre frontières lexicales et syllabiques dans la L1 des apprenants italophones puisse faciliter la production d'une resyllabation de la CL à l'initiale du M2 et ainsi éviter la réalisation de liaisons non-enchaînées. Cependant, nous n'excluons pas que les sujets de notre étude réalisent ce type d'erreur dans des tâches de lecture comme résultat de l'interférence des représentations orthographiques des mots qui favorisent un traitement lexical de la liaison (Chevrot et Malderez 1999; Laks 2005b) (cf. Partie 1.Chapitre 2.2.1.1). En effet certaines recherches (Racine sous presse ; Falbo sous presse) relèvent la présence de productions de liaisons non-enchaînées aussi bien chez des apprenants ayant pour L1 des langues romanes (espagnol et italien) caractérisées par un système syllabique ouvert.

L'observation d'une plus grande présence de liaisons non-enchaînées en lecture qu'en parole semi-spontanée chez des apprenants ayant des L1 typologiquement différentes tel que

[269] Sans considérer les phénomènes de lénition, spirantisation et « raddoppiamento fonosintattico » qui intéressent certains dialectes de l'Italie.

l'anglais (Tennant sous presse), l'espagnol (Racine sous presse) et le norvégien (Andressen et Lyche sous presse) (cf. Partie 3.Chapitre 2.2) semble confirmer l'hypothèse de l'influence de l'input écrit sur la production des liaisons non-enchaînées.

C'est pour cette raison qu'à la suite des études précédentes nous nous attendons à observer la présence d'un écart important entre le taux des liaisons non-enchaînées réalisées dans des tâches de lecture et des tâches de conversation semi-spontanée[270].

4.1.3. Objectif 3 : Influence de l'input écrit dans l'acquisition de la liaison

4.1.3.1 Hypothèse 3

Comme nous l'avons observé dans la troisième partie de ce travail (cf. Partie 3.Chapitre 1.3.2.4.1), la forme orthographique des mots peut jouer un rôle positif dans l'apprentissage d'une langue étrangère (renforçant la représentation phonologique des mots) ou bien un rôle négatif qui donne lieu, comme dans le cas de l'acquisition de la liaison, à des productions calquées sur l'orthographe. À l'aide d'une comparaison entre les données issues de la lecture à haute voix, de l'oral semi-spontané et les résultats d'un test de jugement, nous espérons pouvoir étudier l'influence des représentations orthographiques (cf. Partie 1.Chapitre 6.3) sur l'acquisition de la liaison et donc sur la production de réalisations orthographiques de la CL (ex. [grɑ̃daʀbʀ]). Cette démarche nous permettra d'apporter des éléments afin d'évaluer le niveau d'interaction (lexical ou sub-lexical) entre orthographe et phonologie. Dans une dernière tâche de reformulation (cf. Paragraphe 4.3.2.3), nous effectuerons une comparaison entre les productions issues de deux groupes d'apprenants : l'un ayant été exposé à un support écrit et l'autre soumis au seul input oral.

4.1.4. Objectif 4 – Maîtrise et connaissance de la variabilité de la liaison

Cette étude se propose également de comparer les connaissances épilinguistiques et les compétences de production de la liaison catégorique et variable des apprenants à l'aide d'un protocole expérimental basé sur la méthodologie du projet IPFC et sur d'autres tâches que nous nous attacherons à décrire dans les pages qui suivent. En particulier, à travers le recours à un test de jugement (cf. Paragraphe 4.3.1.6), nous espérons pourvoir vérifier dans quelle mesure les sujets de cette étude sont conscients de la variabilité de réalisation de la liaison.

[270] Tennant (sous presse) observe également une différence importante entre liaisons non-enchaînées produites en lecture (12,9) et en parole semi-spontanée (4,6% dans les entretiens guidés et 6% dans les conversations libres).

4.1.4.1 Hypothèse 4

Étant donné le contexte d'apprentissage uniquement hétéroglotte et le niveau intermédiaire des apprenants[271] (cf. Paragraphe 4.2), nous supposons que les sujets de la présente étude ont une compétence limitée de la variabilité sociolinguistique de la liaison. Celle-ci a été plusieurs fois évoquée en production par les études précédentes menées auprès d'apprenants débutants et intermédiaires (Galazzi et al. 2013). Nous nous attendons donc, contrairement aux études menées chez certains apprenants avancés (Racine sous presse ; Detey sous presse), à observer une faible variation entre la production des liaisons variables issue des tâches de lecture et celle de parole semi-spontanée.

Nous supposons cependant que les connaissances épilinguistiques de la variabilité sociolinguistique sont plus développées que la variabilité attestée en production.

4.1.5. Objectif 5 : enjeux pédagogiques

Pour terminer, nous souhaitons développer une réflexion didactique, sur la base des productions des locuteurs natifs extraites du corpus PFC, afin de préciser la notion de variabilité dans les classifications traditionnelles de la liaison. Ce dernier objectif sera traité dans la dernière partie de ce travail.

4.2. Sujets

Les sujets ayant participé à cette étude longitudinale sont 17 étudiants italophones tous inscrits en Licence de Langues et Littératures Étrangères à l'Université Sapienza de Rome. Il s'agit de 14 filles et de 3 garçons, de 18 à 21 ans, ayant toujours pour la plupart vécu dans la région du *Latium*. Nous avons également recueilli, à l'aide d'une fiche signalétique et d'un questionnaire sociolinguistique (cf. Annexe 4), des informations concernant le sexe, l'âge, le lieux de naissance et le niveau socio-culturel des apprenants (notamment le niveau d'études, et les connaissances linguistiques des parents) qui apparaît dans l'ensemble assez homogène (cf. Tableau 32).

Informations socio-culturelles					
Locuteur				**Parents**	
Code locuteur	**Sexe**	**Âge**	**Lieu naissance**	**Niveau études parents**	**Langues des parents**
Chr	F	19	Latium	secondaire	aucune
Cld	F	18	Latium	secondaire	fr, en
Dbr	F	20	Autre	secondaire/supérieur	fr
Dnl	F	19	Latium	secondaire	aucune
Eln	F	19	Latium	secondaire	aucune

[271] Ainsi que le faible contact avec la langue française en dehors des cours universitaires.

Fdl	F	19	Latium	secondaire	anglais
Frc	F	19	Latium	secondaire	aucune
Frv	F	19	Latium	secondaire	aucune
Gbr	M	19	Latium	supérieur	aucune
Gla	F	19	Autre	secondaire	aucune
Glr	F	19	Latium	secondaire	fr, en
Gmm	M	19	Autre	secondaire	fr (père natif)
Gmp	M	19	Latium	secondaire	en
Irn	F	19	Latium	secondaire	en ; es; de ; el
Jsm	F	21	Latium	secondaire	fa
Nth	F	19	Autre	secondaire	aucune
Yly	F	21	Autre	secondaire	uk; ru

Tableau 32 : Profil socio-culturel des locuteurs du corpus IPFC-italien (Rome)

Afin d'évaluer le niveau de français des apprenants, nous avons pris en compte plusieurs facteurs : (a) nombre d'années d'étude du français ; (b) niveau CECRL selon le test de niveau de français réalisé en début des cours universitaires; (c) autoévaluation ; (d) obtention et niveau du « diplôme d'études en langue française » (dorénavant DELF) ; (e) réalisation d'un séjour ou d'un stage dans un pays francophone ; (g) fréquentation de cours particuliers ; (h) pratiques informelles de la langue en dehors des cours universitaires. En outre, nous avons complété le profil linguistique des apprenants en ajoutant également des données concernant les connaissances linguistiques autres que celles du français.

Les apprenants possèdent des compétences en FLE relativement homogènes correspondant au niveau intermédiaire selon le test de niveau de français réalisé en début des cours. Le français a été toujours appris en milieu institutionnel que ce soit depuis le collège (y compris le lycée) ou seulement au collège ou au lycée. Certains sujets ont obtenu un diplôme DELF de niveau A2 ou B1 pendant leur cursus d'études secondaires.

Bien que la plupart des sujets aient déjà appris le français à l'école secondaire, ils jugent leurs acquis antérieurs insuffisants. En effet les résultats du test d'auto-évaluation que nous avons soumis aux apprenants montrent que plus de la moitié d'entre eux estiment avoir un niveau de français « moyen ». Ces jugements rejoignent les considérations faites par Galazzi (2002b) quant aux attitudes des étudiants.

> « Tout se passe comme si chaque institution gérait son enseignement en vase clos, sans tenir compte de ce qui précède ni de ce qui suit, ceux qui suivent ayant toujours considéré que ce qui le précédait n'était pas valable ou pas suffisant et qu'il valait mieux tout recommencer de zéro ou presque. Cette attitude d'effacement des acquis antérieurs se retrouve au début des cursus universitaires et, qui plus est, elle marque aussi les représentations des étudiants qui demandent souvent un redémarrage sous prétexte que ce qu'ils ont fait à l'école est insuffisant ». (Galazzi 2002b, p. 48)

Quant au contexte d'apprentissage du français il s'agit d'un milieu uniquement hétéroglotte (cf. Partie 3.Chapitre 1.3.2.2) puisque les apprenants n'ont fait que de courts séjours (1-3 semaines) dans des pays francophones dans le cadre de vacances, de programmes d'échanges ou de stages. De plus la plupart des sujets ne pratiquent pas régulièrement le français en dehors de l'université bien qu'ils déclarent communiquer parfois en français sur Internet, écouter de la musique française[272] et avoir pris des cours particuliers pendant les études secondaires. En ce qui concerne les connaissances d'autres langues étrangères, tous les sujets ont étudié ou apprennent à l'université au moins une deuxième langue vivante (cf. Tableau 33).

Code locuteur	N. années études fr	Niveau	Autoévaluation	Diplôme (DELF)	Stage	Séjour	Cours particuliers	Pratique externe	Autres langues[273]
Chr	5	Intermédiaire	Moyen			court	X		en; es, ru
Cld	8	Intermédiaire	Moyen	X (B1)		court	X		en; es
Dbr	8	Intermédiaire	Bon	X (B1)		court		X	en; de; zh
Dnl	8	Intermédiaire	Bon	X (A2)		aucun			en; es
Eln	3	Intermédiaire	Moyen	X (B1)		aucun	X		en; es
Fdl	5	Intermédiaire	Bon	X (B1)	X	court	X		en; es
Frc	3	Intermédiaire	Mauvais			court			en
Frv	8	Intermédiaire	Moyen			court			en; es
Gbr	3	Intermédiaire	Bon			aucun	X	X	en; ru
Gla	5	Intermédiaire	Moyen	X (B1)	X	court			en; de
Glr	8	Intermédiaire	Moyen		X	court			en; de
Gmm	8	Intermédiaire	Moyen		X	court			en; de
Gmp	8	Intermédiaire	Bon	X		aucun		X	en; es; de; sv
Irn	8	Intermédiaire	Bon	X		court	X		ru
Jsm	5	Intermédiaire	Bon	X	X	court	X	X	en; es; jp
Nth	5	Intermédiaire	Moyen	X	X	court	X		en; es; ru
Yly	5	Intermédiaire	Moyen		X	court			en; es; uk; ru

Tableau 33 : Profil linguistique des locuteurs du corpus IPFC-italien (Rome)

Le cursus d'étude du français suivi par l'ensemble des apprenants s'articule en trois parties : (a) des exercices de lectorat oral et écrit ; (b) l'enseignement de la littérature française ; (c) une introduction à la phonétique, phonologie, graphémique et linguistique françaises. L'ensemble correspond à environ 16 heures hebdomadaires comprenant : 6 heures de langue;

[272] Certains déclarent également lire des textes (journaux, magazines, textes en ligne et romans) et regarder des chaînes TV ou des films français ou canadiens.

[273] Codes des langues ISO 639.

4 heures de littérature et 6 heures de linguistique. L'étude de la phonétique occupe donc une place assez importante dans le programme de langue française dès la 1ère année de licence. Les heures de laboratoire linguistique sont consacrées à l'enseignement de la prononciation à travers des activités de compréhension, de lecture et de transcription phonétique. Les exercices de lecture, notamment, se font à haute voix en classe et portent sur des textes assez longs extraits de la presse ou sur des textes littéraires.

Quant à l'enseignement spécifique de la liaison, les manuels proposés aux apprenants contiennent des classifications de la liaison basées sur les descriptions faites par Delattre (1947) et Encrevé (1988). Les supports didactiques utilisés sont la « Grammaire du français pour italophones » de Françoise Bidaud et le texte « Grammatica della lingua francese » de Madonia et Principato (2011).

Pendant les cours de littérature, l'apprenant est également confronté au phénomène de la liaison dans la lecture à haute voix des textes littéraires. Le cursus de linguistique offre en outre aux étudiants l'accès à la lecture d'études théoriques de la liaison comme dans le cas des textes « Phonétique et phonologie du français : théorie et pratique » de Eggs et Mordellet (1990) ou encore « Les variétés du français parlé dans l'espace francophone » de Detey, Durand, Laks et Lyche (2010).

4.3. Méthodologies

L'ensemble du protocole du projet IPFC[274] (cf. Paragraphe 2.3) a été au cœur des méthodologies de la présente étude longitudinale. Cependant, au sein du protocole IPFC, seuls la lecture du texte PFC et les entretiens guidés permettent de recueillir des données concernant la production de la liaison chez les apprenants. Pour cette raison, étant donnés les objectifs que nous nous sommes fixés, l'ajout de tâches supplémentaires a été nécessaire afin de disposer d'un échantillon plus important pour l'analyse du phénomène de la liaison. Parmi ces dernières, une deuxième tâche de répétition et deux tâches de lecture ont été proposées tout au long de l'étude longitudinale[275]: (a) la répétition d'une liste de séquences isolées ; (b) la lecture d'une deuxième liste contenant les mêmes types de liaison potentielles que la tâche de répétition (a) mais des items différents; (c) la lecture d'un texte tiré du journal « L'Express » adapté par la suite afin d'assurer le plus grand nombre de sites de liaisons potentielles ; (d) et un test de jugement basé sur de séquences extraites du texte journalistique précédent.

[274] Sauf la conversation entre les deux apprenants pour laquelle un niveau plus avancé est requis.

[275] Le choix d'ajouter d'autres tâches de lecture se justifie par le fait que le niveau des apprenants de la présente étude ne permet pas d'assurer la mise en place d'entretiens de longue durée.

Deux autres tâches ont fait l'objet d'une étude transversale, l'une au cours de la troisième séance d'enregistrement (S3) à savoir le test de dénomination d'image (cf. Paragraphe 4.3.2.1), l'autre lors de la quatrième et dernière session (S4) correspondant à une tâche de reformulation d'énoncés (cf. Paragraphe 4.3.2.3).

Le tableau ci-dessous (cf. Tableau 34) résume l'ensemble des tâches soumises aux apprenants au cours des quatre séances d'enregistrement de l'étude longitudinale et transversale selon l'ordre de passation adopté.

Séance	Code Séance	Dates	Temps de passation	Tâches
1ère séance	S1	9-17 mai 2013	1h	Répétition liste séquences liaison Liste PFC Lecture liste séquences liaison Lecture texte PFC Lecture texte journalistique Test de jugement Entretiens guidés
2ème séance	S2	24-30 sept 2013	1h	Répétition liste séquences liaison Liste PFC Lecture liste séquences liaison Lecture texte PFC Lecture texte journalistique Test de jugement Entretiens guidés
3ème séance	S3	20-24 janv 2014	1h15	Dénomination d'image Répétition liste séquences liaison Liste PFC Lecture liste séquences liaison Lecture texte PFC Lecture texte journalistique Test de jugement Entretiens guidés

4ème séance	S4	12-16 mai 2014	1h30	Tâche de reformulation Répétition liste IPFC-italien Lecture liste IPFC - italien Répétition liste séquences liaison Liste PFC Lecture liste séquences liaison Lecture texte PFC Lecture texte journalistique Test de jugement

Tableau 34 : Protocole enquête IPFC-italien (Rome)

Dans les pages qui suivent, nous présenterons en détail l'ensemble de ces tâches en les répartissant entre celles qui ont fait l'objet d'une étude longitudinale et celles qui n'ont été proposées qu'au cours d'une seule séance d'enregistrement et donc d'une étude transversale. Les enregistrements ont été réalisés à l'aide d'un enregisteur Tascam DR-100 au format wav à 22050 Hz et en modalité Mono.

Toutes les tâches utilisées dans l'étude longitudinale et transversale sont illustrées dans à la quatrième annexe (cf. Annexe 4).

4.3.1. Étude longitudinale

4.3.1.1 Répétition des séquences isolées de liaison

Dans cette tâche nous avons demandé à l'apprenant d'écouter et de répéter une liste de séquences isolées contenant 22 contextes potentiels de liaison (catégorique et variable), 1 contexte de liaison erratique et 15 distracteurs. Une pause de 5 secondes séparait les séquences permettant au sujet de répéter les cibles. Ceux-ci ont été enregistrés par un locuteur natif. Parmi les contextes de liaison on compte 9 contextes de liaison catégorique, 13 contextes de liaison variable (Durand et al. 2011) et 1 contexte de liaison erratique.

Le tableau ci-dessous (cf. Tableau 35) résume les types ainsi que les séquences de liaison proposées dans la tâche de répétition.

	Type	**Tokens**
Catégoriques	DET-NOM	Des amis
		Des ânons
		Des événements
		Des envoûtements
	PRO-VER	Vous êtes
		Vous encrez
	VER-PRO	Allons-y

		Donnez-en
	EXF-EXF	De plus en plus
Variables	ADJ-NOM	Un petit / grand[276] ami
		Un petit / grand ânon
		Un petit / grand événement
		Un petit / grand envoûtement
	AUX-PPA	Ils ont appris
		Ils ont acquis
	NOM-ADJ	Des efforts inutiles
		Des essais intéressants
	VER-ADV	Nous sommes ici
		Il sort aussi
	CON-PRO	Quand il neige
	PRP-NAM	En Inde
	PRP-NAM	En Aude
Erratique	EXF-EXF	Au fur et // à mesure

Tableau 35 : Types et séquences de liaison issues de la tâche de répétition des séquences isolées

Quant aux séquences distractrices nous avons sélectionné des formes contenant des aspects complexes dans l'apprentissage de la phonologie française tels que les voyelles nasales (ex. l'opposition entre « des teintes » et « des tantes ») (cf. Annexe 4).

La passation de la tâche était automatique, réalisée à l'aide du logiciel Praat (Boersma et Weenink 2015) et précédée d'une séance d'entraînement contenant 3 séquences utilisées uniquement pendant cette phase. Les ordres de proposition des séquences et des distracteurs variaient pour chaque groupe de deux locuteurs et d'une session à l'autre afin de limiter les effets de fatigue dans l'analyse des productions des apprenants. Les sujets écoutaient la cible à l'aide d'un casque et n'avaient pas la possibilité de revenir en arrière. La consigne donnée aux apprenants consistait à répéter tout simplement ce qu'ils entendaient[277].

[276] La séquence « un petit ami » a été soumise lors de la première et de la deuxième séance (S1, S2) alors que la séquence « un grand ami » a été proposée aux apprenants lors de la troisième et de la quatrième séance (S3, S4). Cela nous a permis de comparer respectivement les productions de la liaison après les adjectifs « petit » et « grand » issues de la répetition et de la lecture avec celles réalisées en parole semi-spontanée notamment dans la tâche de reformulation (cf. Partie 4.Chapitre 1.4.3.2.3).

[277] Aucune précision n'était donnée pour les cas d'incertitude ou de non-compréhension laissant au sujet le choix de reformuler ou de ne pas donner de réponse.

4.3.1.2 Lecture des séquences isolées de liaison

Les types de liaison retenus correspondent à ceux de la tâche de répétition, bien que les items sélectionnés soient différents[278]. Nous avons donc soumis une liste de séquences isolées contenant 22 contextes potentiels de liaison, 1 contexte de liaison erratique et 21 distracteurs. Parmi les contextes de liaison, on compte 9 contextes de liaison catégorique, 13 contextes de liaison variable (Durand et al. 2011) et 1 contexte de liaison erratique. Nous avons essayé de créer, dans les contextes de liaison catégorique, des couples de séquences ayant les mêmes caractéristiques morphosyntaxiques (à savoir le même type de liaison) mais une fréquence de cooccurrence différente. Les indices de fréquence de cooccurrence ont été tirés de la *freqfilm2* de « Lexique 3 » (New 2006) correspondant à la moyenne des fréquences par million des quatre sous-corpus de sous titres de films qui constituent cette base de donnée lexicale. Cette valeur est représentée de manière absolue dans le tableau ci-dessous. Nous espérons à travers cette approche démontrer que la production de liaison dans des contextes catégoriques n'est pas sensible aux facteurs lexicaux notamment de la fréquence de cooccurrence (cf. Partie 1.Chapitre 2.2.4.1). Ces résultats nous permettraient d'apporter des éléments de réponse à l'hypothèse (hypothèse 1) formulée dans les paragraphes précédents en montrant que l'acquisition de liaison dans ces contextes ne serait pas limitée à un traitement « au cas par cas ». Une acquisition de ce type présupposerait, en effet, que les apprenants produisent un taux plus important de liaison au sein des séquences auxquelles ils sont majoritairement exposés à savoir les plus fréquentes. En revanche, nous supposons que les apprenants sont capables de développer des généralisations, au moins dans des contextes de liaison catégorique, qui dépassent de simples stratégies d'acquisition lexicale. Le tableau ci-dessous (cf. Tableau 36) résume les types et les séquences proposés pendant cette tâche.

	Type	**Tokens**	**Fréq**	**Tokens**	**Fréq**
Catégoriques	DET-NOM	Des arbres	341	Des âtres	0
		Des appartements	61	Des appontements	0
	PRO -VER	Nous avons	> 1000	Nous amarrons	0
	VER-PRO	Allez-y	> 1000	Prenez-en	194
	EXF-EXF	De temps en temps			
Variables	ADJ-NOM	Un petit/grand[279] arbre			
		Un petit/grand âtre			
		Un petit/grand appartement			

[278] Sauf dans les cas du contexte de liaison erratique et de deux séquences « prépositions – nom propre ».

[279] « Petit » lors de la première et de la deuxième séance (S1, S2) et « grand » lors de la troisième et de la quatrième séance (S3, S4).

		Un petit/grand appontement
	AUX-PPA	Ils ont acheté
		Ils ont achevé
	NOM-ADJ	Des enfants intelligents
		Des conseils idiots
	VER-ADV	Ils sont ici
		Il vient aussi
	CON-PRO	Quand il pleut
	PRP-NAM	En Inde
	PRP-NAM	En Aude
Erratique	EXF-EXF	Au fur et // à mesure

Tableau 36 : Types et séquences de liaison issus de la tâche de lecture des séquences isolées

Les distracteurs utilisés correspondent aux mêmes que ceux déjà utilisés dans la tâche de répétition auxquels nous avons ajouté 6 autres distracteurs afin de vérifier les réalisations des mots « grand », « quand » , « ces », « sans », « premier » et « vient » devant des mots à initiale consonantique (ex. « un grand cheval ») (cf. Annexe 4). Comme nous le verrons dans le prochain chapitre (cf. Partie 4.Chapitre 2.1.3.2) cette demarche nous a permis d'écarter l'hypothèse d'après laquelle les réalisations ortographiques des CL des ces mots en contexte de liaison pourraient correspondre à des prononciations erronées de ces items en contexte de non-liaison à savoir devant un mot à initiale consonantique (ex. « un grand cheval » prononcé [œ̃gʀɑ̃dʃ(ə)val] au lieu de [œ̃gʀɑ̃ʃ(ə)val]).

La passation était automatique, réalisée à l'aide des diapositives et précédée par une séance d'entraînement contenant 3 séquences utilisées uniquement pendant cette phase. Une pause de 5 secondes séparait les diapositives pour permettre au sujet de lire à haute voix les cibles. L'ordre de distribution des séquences et des distracteurs variait d'un locuteur sur deux et d'une session à l'autre afin de pouvoir limiter les effets de fatigue pendant la lecture des séquences. Les sujets ne pouvaient pas revenir en arrière pour relire les séquences. La consigne donnée aux sujets consistait tout simplement à lire à haute voix les séquences qui s'affichaient à l'écran.

4.3.1.3 Lecture de la liste PFC

Cette tâche correspond à une lecture à haute voix d'une liste de 94 mots issue du protocole du corpus PFC (Durand, Laks et Lyche 2009) (cf. Partie 2.Chapitre 1) qui comprend d'une part

une liste où l'ordre des mots est aléatoire (1-84) et d'autre part une dizaine de paires minimales (85-94) (cf. Annexe 4). Cette lecture s'est avérée extrêmement coûteuse en temps et en énergie pour les apprenants qui cependant n'ont pas demandé de pause alors que nous avions signalé la possibilité d'interrompre la lecture tous les 25 mots.

4.3.1.4 Lecture texte PFC

La lecture ici porte sur un passage d'un texte journalistique construit et tiré du protocole du projet PFC (Durand, Laks et Lyche 2009) (cf. Annexe 4). La passation était manuelle. Le texte présenté par l'enquêteur était reproduit avec un double espace et une police de taille égale à 14 pour faciliter la lecture. L'alignement à droite du texte ne coïncidait pas avec des sites de liaison potentielle. L'enregistrement était précédé d'une phase d'environ 5 minutes de familiarisation avec le texte qui permettait au sujet de réaliser une lecture silencieuse avant la lecture à haute voix. Le texte d'environ 400 mots comprend 12 sites potentiels de liaison catégorique, 19 de liaison variable et 6 de liaison erratique[280] (Durand et al 2011). Le tableau ci-dessous (cf. Tableau 37) résume l'ensemble des séquences de liaisons potentielles ainsi que les types de liaison.

	Type	Tokens
Catégoriques	**DET-NOM**	Son usine
		Les élections
		Les opposants
		Quelques articles
		Des activistes
	DET-ADJ	Un autre côté
	PRO-AUX	En a vu
	PRO -VER	On est
		Nous avons
	PRO-PRO	On en
	EXF-EXF	En effet
		Jeux olympiques
Variables	**ADJ-NOM**	Grand émoi
		Grand honneur
	ADV-ADJ	Très inquiet
	ADV-PRP	Toujours autour
	AUX-PPA	Ont eu
	AUX-PRP	S'est en

[280] Cf. Note 264.

	INF-DET	Provoquer une
	INF-PRP	Trouver au
	ADJ-NOM	Pâtes italiennes
		Circuits habituels
		Visites officielles
	NOM-AUX	Fanatiques auraient
	NOM-PRP	Chemises en soie
	PRO-AUX	Tout est
	PRP-DET	Dans une
	VER-PRP	Est en (x2)
	ADV-PRP	Plus à quel
	VER-DET	Préparent une
Erratiques	**ADV-DET**	Vraiment // une
	ADV-PRP	Comment // en plus
	NAM-PRP	Berlin // en
	NOM-PRO	Dans le coin //on
	NOM-PRP	Région // en
	NOM-VER	Opposition // aurait

Tableau 37 : Types et séquences de liaison issus de la tâche de lecture du texte PFC

Comme le font observer Galazzi et al. (2013), ce texte présente un certain nombre de difficultés car il contient des mots qui sont probablement inconnus d'un public d'étudiants de niveau intermédiaire (« émoi », « entamé », « hasard » et « vouer ») ou encore des mots difficiles à prononcer (« Beaulieu » et « 1936 ») ou pouvant donner lieu à des confusions (« Beaulieu / banlieue » ; « cote / côté »). De plus ces auteurs signalent la présence de structures complexes telles que « comme on en a vu ». Malgré la familiarisation avec le texte et l'habitude des sujets à pratiquer une lecture à haute voix pendant leurs cours, nous pouvons nous attendre à ce que les apprenants rencontrent des difficultés ou fassent tout au moins preuve de quelques hésitations dans la réalisation de cette tâche. Néanmoins, le recours à ce texte nous permet de confronter les sujets à des types de liaison que nous aurions difficilement pu observer dans une conversation semi-spontanée (cf. De Moras 2011).

De plus, cette approche nous permettrait de comparer le taux de réalisation des liaisons non-enchaînées produites en lecture avec celui issu de la tâche de parole semi-spontanée. Il s'agira notamment de vérifier l'hypothèse (voir hypothèse 2) concernant l'influence de la L1

sur la réalisation du non-enchaînement de la liaison et sur l'origine de ce type d'erreurs de liaison chez des apprenants italophones.

4.3.1.5 Lecture du texte journalistique adapté

Les modalités de passation et les hypothèses de cette deuxième tâche de lecture que nous avons intégrée au protocole sont les mêmes que celles de la tâche précédente. De plus, à l'aide de cette tâche nous souhaiterons vérifier l'hypothèse (hypothèse 4) évoquée dans les paragraphes précédents (cf. Paragraphe 4.1.4.1) et comparer ainsi les productions de la liaison variable avec les connaissances épilinguistiques des apprenants (Gombert 1990) observées dans la tâche de jugement (cf. Paragraphe 4.3.1.6).

Le texte employé dans cette tâche compte environ 580 mots. Il est tiré de l'article « Ghettos, sweet ghettos » du journal « l'Express » et a ensuite été adapté afin de créer le plus grand nombre de sites de liaison possibles. Il comprend 31 sites potentiels de liaison catégorique, 30 de liaison variable et 1 de liaison erratique. Le tableau ci-dessous (cf. Tableau 38) résume l'ensemble des séquences de liaisons ainsi que les types de liaison.

	Type	Tokens
Catégoriques	**DET-NOM**	Les images
		Les intrus
		Les appartements
		Les appareils
		Les élus
		Ces enclaves
		Son interphone
		Son alarme
		Ces arsenaux
		Leurs intérêts
	DET-ADJ	Un autre
	NUM-NOM	Deux étages
		Trente-huit ans
		Deux exposants
		Dix ans
		Les deux cents acquéreurs
		Huit cents euros
		Les deux cents habitants
	PDE-NOM	Des espaces verts
		Des émeutes
		Des alentours
	PRO-AUX	Vous êtes encore filmés

		Nous avons retiré
		On a commencé
	PRO-VER	Ils arrivent
		Ils abiment
	PRO-PRO	Nous en installions
		Ils y ont
	EXF-EXF	De temps en temps
	VER-PRO	Sont-elles
		Dit-elle
Variables	**ADJ-NOM**	Premier octobre
		Grand appartement
		Petit enfant
		Petit abri
		Grand enclos
	ADV-PRP	Plus à
	ADV-DET	Pas un
	ADV-ADJ	Les plus aisés
	ADV-PRO	Mais on
		Quand on
	AUX-ADV	Ont aussi dû
	AUX-PPA	Sont équipés
		Ont été rayées
	NOM-ADJ	Intérêts économiques
		Cocons ultraprotégés
		Émeutes urbaines
	PRO-PRP	Le premier à
	PRP-NOM	Sans autorisation
	VER-ADJ	Est imprenable
		Est impossible
	VER-ADV	Habitent ici
		Sont aujourd'hui
		Vivent ici
	VER-DET	Il faut une
		Est une
	VER-PRP	Est à
		Fleurissent en

		Sont en
		Sent à
Erratiques	**ADV-ADJ**	Plus // haut

Tableau 38 : Types et séquences de liaison issus de la tâche de lecture du texte journalistique

4.3.1.6 Test de jugement

Une partie des contextes de liaison contenus dans la tâche de lecture journalistique ont été extraits et utilisés dans la tâche de jugement afin de disposer des mêmes séquences et de pouvoir ainsi comparer les productions et les jugements des apprenants. Pour être plus précis, nous avons soumis aux sujets 14 distracteurs et 13 sites de liaison tirés du texte journalistique (dont 6 induisant une liaison catégorique et 7 introduisant une liaison variable), chacun possédant trois différentes prononciations. En effet, pour chaque énoncé contenant la séquence de liaison, trois différentes prononciations ont été enregistrées par un locuteur natif, dont une seule était correcte dans le cas des contextes de liaison catégorique et deux dans le cas des contextes de liaison variable. Les réponses restantes correspondaient à des prononciations orthographiques (cf. Partie 3.Chapitre 2.1.4), des substitutions de la CL ou à des liaisons non réalisées. À titre d'exemple, dans le cas de l'énoncé « ils abiment des voitures », nous avons enregistré : (a) une prononciation orthographique [ilsabimdevwatyʀ][281] ; (b) une réalisation conforme à la CL cible [ilzabimdevwatyʀ] ; (c) une réalisation sans liaison [ilabimdevwatyʀ].

[281] Dans les autres cas nous avons inséré une prononciation correspondant à une substitution de la CL cible (ex. [dizɛl] au lieu de [ditɛl]).

L'apprenant devait écouter, à l'aide d'un casque, les trois différentes réalisations de la même séquence et choisir entre une ou deux réponses correctes. La passation était automatique et réalisée à l'aide d'une interface Java en italien (cf. Figure 27).



Figure 27 : Interface pour la passation automatique du test de jugement

La transcription de l'énoncé était affichée au centre de la page (onglet *Frasi* « énoncés ») suivie de trois boutons placés en dessous, qui permettaient de lancer l'écoute des trois enregistrements (boutons *Ascolta* « écoute »). Aucune limite de temps n'était donnée pour la réalisation de ce test ni pour la répétition des écoutes. Le sujet devait ensuite cocher la case (ou les deux cases au-dessous des boutons « écoute ») pour les réponses qu'il jugeait correctes et appuyer sur le bouton « suivant » (*successivo*) pour passer à la séquence successive.

Ce test expérimental pouvait donner lieu à 8 combinaisons de réponses possibles : 4 pour les contextes de liaison catégorique et 4 pour les contextes de liaison variable (cf. Tableau 39).

- Schéma de réponses possibles pour les contextes de liaison catégorique (ex. « ils abiment des voitures ») :
  - *Good* « correct » : 1 réponse attendue = 1 réponse correcte (ex. [ilzabimdevwatyʀ]) ;
  - *Invalid plus* « nul plus » : 1 réponse attendue = 1 réponse correcte et 1 autre réponse non attendue et fausse (ex. [ilzabimdevwatyʀ] et [ilsabimdevwatyʀ]) ;
  - *Error* « erreur » : 1 réponse attendue = 1 réponse fausse (ex. [ilsabimdevwatyʀ]) ;
  - *Error plus* « erreur plus » : 1 réponse attendue = 2 réponses fausses (ex. [ilabimdevwatyʀ] et [ilsabimdevwatyʀ]).

- Schéma de réponses possibles pour les contextes de liaison variable (ex. « grand enclos ») :
    - *Good* « correct » : 2 réponses attendues = 2 réponses correctes (ex. [gʀãtãklo] et [ləgʀããklo]);
    - *Missing* « manqué » : 2 réponses attendues = 1 réponse correcte et 1 réponse attendue manquée (ex. [ləgʀããklo]);
    - *Invalid* « nul » : 2 réponse attendue = 1 réponse correcte et 1 réponse attendue fausse (ex. [gʀãtãklo] et [ləgʀãdãklo]);
    - *Error* « erreur » : 2 réponses attendues = 1 réponse fausse (ex. [ləgʀãdãklo]).

Nous espérons qu'une comparaison de ces résultats avec les données tirées de la lecture du texte journalistique nous permettra d'accéder aux représentations épilinguistiques (Gombert 1990) que les apprenants ont de la liaison. Si l'hypothèse que nous avons formulée ci-dessus (hypothèse 4) est correcte (à savoir les apprenants observés en milieu institutionnel et hétéroglotte ne sont pas conscients de la variation de la liaison), nous pouvons nous attendre à avoir une majorité de réponse de type *missing* où la variante correspondant à la réalisation de la liaison ou à la non-réalisation de la liaison n'est pas sélectionnée.

De plus, comme nous l'avons déjà observé (hypothèse 3), les résultats de cette expérience nous permettront d'analyser le niveau d'interaction (lexical ou sub-lexical) entre phonologie et orthographe grâce à l'analyse des réponses contenant des réalisations orthographiques de la CL.

	Type	Tokens
Catégoriques	**DET-NOM**	Ces arsenaux a. [sesaʀsəno] b. [sezaʀsəno] c. [seaʀsəno]
	NUM-NOM	Deux exposants a. [døɛkspozã] b. [døzɛkspozã] c. [døsɛkspozã]
		Dix ans a. [diã] b. [dizã] c. [disã]
	PRO -VER	Ils abiment a. [ilsabim] b. [ilzabim] c. [ilabim]
	EXF-EXF	De temps en temps a. [dətããtã] b. [dətãsãtã] c. [dətãzãtã]

	VER-PRO	Dit-elle a. [ditɛl] b. [diɛl] c. [dizɛl]
Variables	**ADJ-NOM**	Petit abri a. [pətitabʀi] b. [pətabʀi] c. [pətiabʀi]
		Grand enclos a. [ləgʀɑ̃ɑ̃klo] b. [ləgʀɑ̃tɑ̃klo] c. [ləgʀɑ̃dɑ̃klo]
	ADV-PRO	Quand on a. [kɑ̃dɔ̃] b. [kɑ̃tɔ̃] c. [kɑ̃ɔ̃]
	AUX-ADV	Ont aussi a. [ilzɔ̃osi] b. [ilzɔnosi] c. [ilzɔ̃tosi]
	NOM-ADJ	De cocons ultraprotégés a. [dəkɔkɔ̃zyltraprɔteʒe] b. [dəkɔkɔ̃syltraprɔteʒe] c. [dəkɔkɔ̃yltraprɔteʒe]
	VER-ADJ	Est imprenable a. [ɛtɛ̃pʀənabl] b. [ɛɛ̃pʀənabl] c. [ɛsɛ̃pʀənabl]
	VER-PRP	Fleurissent en a. [flœʀisɑ̃tɑ̃] b. [flœʀisɑ̃ɑ̃] c. [flœʀisɑ̃nɑ̃]

Tableau 39 : Types et séquences de liaison issus du test de jugement

4.3.1.7 Entretien guidé

L'entretien guidé a été réalisé avec un natif. Cette tâche s'appuie sur une liste de questions fermées visant à retracer la biographie linguistique de l'apprenant avant de passer à des questions plus ouvertes ayant pour objet des thèmes tels que les vacances, les examens etc. Étant donné le niveau et les difficultés de production rencontrées par les sujets observés, la durée moyenne de chaque conversation dépasse rarement dix minutes.

4.3.2. Étude transversale

L'étude transversale comprend quatre tâches. D'une part un test de dénomination d'image soumis au cours de la troisième séance d'enregistrement (S3) et d'autre part les tâches de répétition et de lecture de la liste spécifique IPFC (cf. Paragraphe 2.3) ainsi qu'une tâche de reformulation d'énoncé, réalisées lors de la quatrième et dernière session (S4).

4.3.2.1 Dénomination d'images

Dans le but de disposer d'un plus grand nombre de données issues de l'oral semi-spontané[282], nous avons soumis au cours de la troisième séance d'enregistrements (S3) une tâche de dénomination d'image contenant cinq images cibles (cf. Tableau 40). La tâche consiste à présenter aux apprenants deux images représentant le même objet (ex. l'image d'un arbre) mais sous deux formats différents afin d'amener le locuteur à produire des séquences cibles contenant les adjectifs « grand » et « petit ». Chaque image était accompagnée de deux types d'indices. D'une part un indice écrit indiquant le mot italien correspondant à l'image (ex. *un albero* « un arbre » en français) et d'autre part un autre indice oral signalant le mot français correspondant (ex. « un arbre »). Cette démarche nous permet d'éviter que le sujet ait accès à l'input écrit en français tout en assurant la compréhension de la cible grâce au recours à la transcription du mot italien. De plus les images cibles ont été tirées de la tâche de lecture des séquences isolées de liaison (cf. Paragraphe 4.3.1.2) de l'étude longitudinale qui a suivi le test de dénomination dans la troisième séance (S3). Cela nous permettra de comparer les données issues de la lecture avec les productions de parole semi-spontanée du test de dénomination d'images où les sujets n'ont pas accès à des supports écrits en français. À travers cette démarche nous espérons pouvoir analyser l'influence de l'orthographe sur les représentations phonologiques de la liaison (cf. Partie 1.Chapitre 6.3) dans la production orale (hypothèse 3 cf. Paragraphe 4.1.3.1).

La passation de ce test était automatique et réalisée à l'aide de diapositives qui contenaient deux images du même objet avec deux formats différents (l'un plus petit que l'autre, cf. Figure 28). Celles-ci étaient associées à un titre correspondant au terme italien et à un bouton permettant d'écouter l'enregistrement du mot correspondant en français (réalisé par un locuteur natif). Le test était précédé d'une séance d'entraînement. Les sujets devaient écouter l'indice oral avant de décrire les images selon leurs caractéristiques de grandeur et produire ainsi des énoncés cibles tels que « un petit arbre » et « un grand arbre ».

Images (2 formats)	**Séquences cibles**	**Input écrit (it)**	**Input oral (fr)**
Arbre	Un petit arbre Un grand arbre	*Un albero*	Un arbre
Etang	Un petit étang Un grand étang	*Uno stagno*	Un étang

[282] Vue la faible durée des entretiens guidés (cf. Paragraphe 4.3.1.7).

Appontement	Un petit appontement Un grand appontement	*Un pontile*	Un appontement
Âtre	Un petit âtre Un grand âtre	*Un focolare*	Un âtre
Appartement	Un petit appartement Un grand appartement	*Un appartamento*	Un appartement

Tableau 40 : Séquences de liaison de la tâche de dénomination d'images

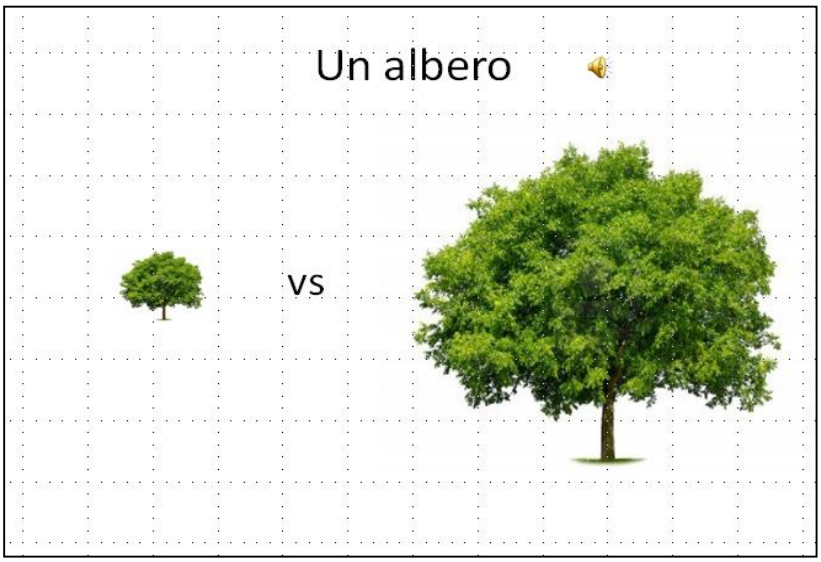


Figure 28 : Exemple d'une diapositive de la tâche de dénomination d'images

4.3.2.2 Répétition et lecture d'une liste spécifique de mots (IPFC-italien)

La liste utilisée pour la répétition et la lecture comprend une base commune de 34 mots et une liste de 38 mots spécifiques visant à vérifier les prononciations d'items contenant des caractéristiques phonétiques complexes pour l'apprenant italophone (à savoir l'opposition /u/ ~/y/) (cf. Annexe 4). La liste de mots obtenue a été enregistrée par un locuteur natif. Ces deux tâches n'ont évidemment pas permis de recueillir des données concernant la réalisation de la liaison et elles ont été uniquement soumises lors de la quatrième séance (S4) afin de mettre en évidence les difficultés de prononciation rencontrées par les apprenants.

4.3.2.3 Tâche de reformulation

Cette dernière tâche consiste en une épreuve de reformulation d'une série de 15 séquences (cf. Tableau 42) isolées que nous avons réalisée au moment de la dernière séance d'enregistrements (S4). L'objectif du test était de faire produire un énoncé cible contenant un site de liaison ayant un sens opposé à celui de l'énoncé source. La reformulation demandait la modification du premier élément (ex. « sur» dans la séquence « sur un arbre ») ou du dernier

élément (ex. « femmes » dans la séquence « plusieurs femmes ») de la séquence source afin de produire des séquences cibles de sens opposé contenant des sites de liaison potentielle (ex. « sous un arbre » et « plusieurs hommes »). Dans le cadre de ce test, nous avons séparé les sujets en deux groupes selon le type d'input écrit ou oral sous lequel les séquences sources étaient présentées (cf. Tableau 41).

La passation de la tâche était automatique et réalisée à l'aide de diapositives (cf. Figure 29). La tâche était précédée d'une session d'entraînement pendant laquelle il était expliqué que le test était divisé en deux parties selon la position du mot faisant l'objet de la reformulation. Une diapositive signalait ainsi si l'apprenant devait modifier le premier ou le dernier item des séquences sources. Celles-ci étaient ensuite présentées sous forme écrite ou orale selon le groupe d'appartenance du sujet[283]. La consigne présentait l'épreuve comme un test psycholinguistique dans lequel le sujet devait trouver, à l'aide d'un indice, le contraire du mot (premier ou dernier) appartenant à la séquence source afin de produire à haute voix un énoncé cible ayant un sens opposé à celui d'origine. Les mots de sens contraire étaient en outre suggérés par un indice oral (voir « Aide » Figure 29) que l'apprenant devait écouter avant de produire la séquence cible attendue.

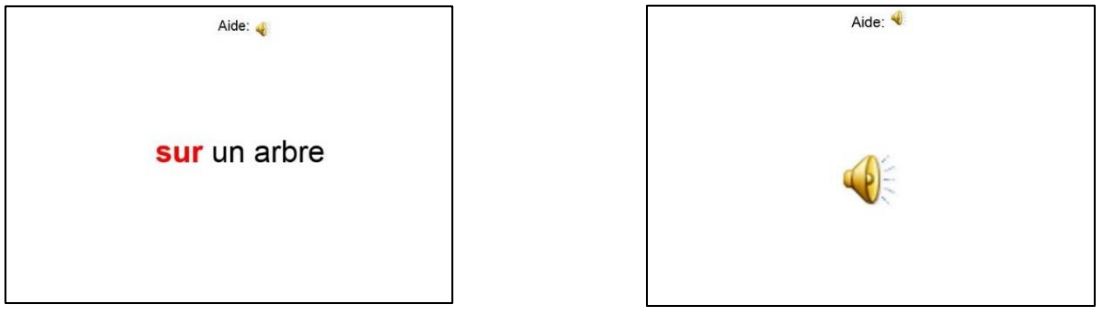


Figure 29 : Exemple d'une diapositive de la tâche de reformulation

[283] Pour les sujets ayant accès à l'input écrit, les mots pour lesquels une modification était requise étaient signalés en gras et en rouge afin d'éviter des confusions éventuelles qui cependant n'ont jamais étaient observées dans aucun des deux groupes d'apprenants.

	Groupe 1- Input écrit	Groupe 2- Input oral
Sujets	Dnl	Chr
	Eln	Cld
	Gla	Dbr
	Gmm	Fdl
	Irn	Frc
	Nth	Frv
	Jsm	Gbr
	Yly	Glr
		Gmp

Tableau 41 : Répartition des sujets en deux groupes (écrit et oral) pendant la tâche de reformulation

	Séquence source	Indice	Séquences cibles	Type de liaison
Première partie : reformulation du premier mot	Sur un arbre	Sous	Sous un arbre	PRP-DET
	Sur une plante	Sous	Sous une plante	PRP-DET
	Derrière un arbre	Devant	Devant un arbre	PRP-DET
	Derrière une plante	Devant	Devant une plante	PRP-DET
	Nouvel ami	Ancien	Ancien ami	ADJ-NOM
	Mal écrit	Bien	Bien écrit	ADV-PPA
Deuxième partie : reformulation du dernier mot	Des femmes	Hommes	Des hommes	DET-NOM
	Cette femme	Homme	Cet homme	DET-NOM
	Très malheureux	Heureux	Très heureux	ADV-ADJ
	Moins malheureux	Heureux	Moins heureux	ADV-ADJ
	Plus malheureux	Heureux	Plus heureux	ADV-ADJ
	Toujours malheureux	Heureux	Toujours heureux	DET-NOM
	Il est malheureux	Heureux	Il est heureux	VER-ADJ
	C'est correct	Incorrect	C'est incorrect	VER-ADJ
	Je suis malheureux	Heureux	Je suis heureux	VER-ADJ

Tableau 42 : Types et séquences de liaison de la tâche de reformulation

4.4. Traitement des données

L'ensemble des tâches de deux types d'étude (transversale et longitudinale) que nous venons de décrire nous ont permis de recueillir un corpus diversifié (répétition, lecture, parole semi-spontanée et jugements) correspondant à environ 35 heures d'enregistrement. Ceux-ci ont été intégralement transcrits selon les conventions du protocole IPFC à l'aide du logiciel Praat (Boersma et Weenink 2015).

4.4.1. Traitement des données de liaison

Comme nous l'avons déjà observé, une partie seulement des tâches du protocole IPFC a été exploitée pour l'étude de l'acquisition de la liaison. À ces données s'ajoutent les résultats des expériences que nous avons menées dans le cadre de l'étude longitudinale et transversale. De l'ensemble de ces tâches nous avons extrait un total de 11302 sites de liaison[284] (toutes liaisons confondues y comprises les liaisons erratiques réalisées et les liaisons inattendues) qui ont été codés par le biais du codage alphanumérique adopté par le projet IPFC. La transcription orthographique ainsi que le codage de la liaison ont été réalisés à l'aide du logiciel Praat (Boersma et Weenink 2015) dans deux couches d'annotation différentes. Ce traitement des données assure une analyse perceptive et une description fine de l'ensemble des contextes de liaison par le biais des 7 champs de codage que nous avons décrits dans les paragraphes précédents (cf. Paragraphe 2.4.2). De plus, ce codage de la liaison présente l'avantage de permettre de nuancer la typologie de réalisation de la liaison (champ 5 de codage cf. Paragraphe 2.4.2.4) sans avoir à trancher entre l'absence de liaison et sa réalisation (tout court). En outre, le codage du type de CL offre la possibilité d'une analyse perceptive adaptée aux productions interlinguistiques propres à l'apprenant de FLE telles que les réalisations orthographiques des CL ou encore les réalisations d'une consonne orthographique présente dans le mot (cf. Partie 3.Chapitre 2.1.4).

Quant à l'analyse quantitative des données de liaison, elle a été réalisée à l'aide du logiciel Dolmen[285] (plug-in IPFC) (Eychenne et Paternostro sous presse) qui nous a permis d'interroger le corpus sur la base du système de codage de la liaison. Cet outil présente l'avantage d'assurer une intégration avec le logiciel Praat ce qui facilite le traitement et l'analyse acoustique des données du corpus.

[284] Parmi lesquels 10857 sites issus de l'étude longitudinale et 445 sites issus de l'étude transversale.
[285] Outil pour l'analyse, la gestion de corpus de langue (www.julieneychenne.info/dolmen).

Chapitre 2. Résultats de l'enquête IPFC-italien (Rome)

Dans ce chapitre, nous nous attacherons à présenter les données issues de l'enquête IPFC-italien afin de vérifier les hypothèses formulées dans le chapitre précédent.

Dans un premier temps, nous décrirons les résultats totaux de la seule étude longitudinale afin d'établir un premier bilan de l'évolution de l'acquisition de la liaison. Il s'agira plus précisément de présenter les résultats concernant le type de réalisation de la liaison (champ 4 et 5 du codage liaison IPFC cf. Partie 4.Chapitre 1.2.4.2.4), la nature de la CL réalisée (champ 6 du codage liaison IPFC cf. Partie 4.Chapitre 1.2.4.2.5) selon les typologies de liaison (cf. Partie 1.Chapitre 4.2) toutes tâches confondues de l'étude longitudinale[286]. Ces premières analyses des résultats nous permettront d'apporter des éléments de réponse à l'hypothèse (hypothèse 1) (cf. Partie 4.Chapitre 1.4.1.1.1) que nous avons formulée dans le chapitre précédent, selon laquelle l'apprenant à partir d'un niveau intermédiaire serait capable d'opérer des généralisations concernant la réalisation de la liaison dans des cas de liaisons catégoriques. Nous espérons également à travers ces résultats pouvoir étudier l'influence de la L1 sur l'acquisition de liaison et notamment vérifier la deuxième hypothèse (hypothèse 2) selon laquelle les apprenants italophones, contrairement aux apprenants anglophones (Mastromonaco 1999 ; Howard 2005 ; Howard 2013 ; Thomas 2002 ; De Moras 2011), produisent un nombre plus faible de liaisons non-enchaînées.

Dans un deuxième temps, nous présenterons les résultats des tâches proposées au cours de l'étude transversale (cf. Partie 4.Chapitre 1.4.3.2).

Pour terminer, l'ensemble de ces données issues de l'étude longitudinale et transversale seront reparties selon les trois types de production orale que nous avons recueillis dans notre corpus à savoir la répétition, la lecture oralisée et la parole semi-spontanée. Une analyse comparative du taux de production des réalisations orthographiques de la CL (ex. [gʀɑ̃daʀbʀ]) entre les différents types de production nous permettra de recueillir des données pour étudier les interactions entre phonologie et orthographe (cf. Partie 1.Chapitre 6.3) (hypothèse 3) dans

[286] Malgré des limites inévitables (cf. Partie 1.Chapitre 4.4), cette modalité de classification des données s'avère nécessaire afin d'éviter l'attribution *a priori* d'étiquettes issues des classifications qui, comme nous l'avons vu, ne font pas l'objet d'un consensus. Dans chaque sous-chapitre, nous présenterons donc tout d'abord les contextes morphosyntaxiques de réalisation de la liaison (type de liaison) (voir étiquettes morphosyntaxiques proposées par les champs 2 et 3 du codage liaison IPFC cf. Partie 4.Chapitre 1.2.4.2.2) qui nous serviront ensuite comme bases pour établir un bilan de l'évolution des différentes typologies de liaison (catégorique, variable et erratique) d'un point de vue de la réalisation de la liaison (champ 5 du codage liaison IPFC) et de la nature de la CL réalisée (champ 6 du codage liaison IPFC).

l'acquisition de la liaison. Ces données seront en outre comparées avec les résultats issus du test de jugement, ce qui nous permettra également de vérifier la maîtrise ainsi que les connaissances épilinguistiques que les apprenants FLE possèdent de la variation de la liaison (hypothèse 4).

1. Résultats de l'étude longitudinale

Les résultats présentés dans ce paragraphe s'appuient sur les données issues de l'ensemble des tâches de production des quatre sessions de l'étude longitudinale IPFC-italien (Rome). Les données longitudinales de quatre sessions d'enregistrement (S1, S2 ; S3 ; S4) correspondent à 10857 occurrences totales de liaison (y comprises les liaisons erratiques réalisées et les liaisons non attendues)[287] que nous allons repartir selon les types de liaison. Cette analyse préalable des contextes morphosyntaxiques de liaison servira d'appui pour l'identification des typologies de liaison (catégorique, variable et erratique) selon la classification proposée par Durand et al. 2011. Une telle démarche nous permettra d'assurer à tout moment une comparaison avec les résultats des études ayant utilisé des classifications différentes (Delattre 1947 ; Encrevé 1988).

1.1. Résultats totaux

Le tableau ci-dessous (cf. Tableau 43) résume l'ensemble des contextes de liaison potentielle catégorique et variable enregistrés au cours des 4 séances de l'étude longitudinale. Ces sites de liaison sont identifiés sur la base des catégories morphosyntaxiques d'appartenance des M1 et M2 qui ont été assignées à partir du jeu d'étiquettes morphosyntaxiques (séparées par un trait d'union ex. DET-NOM) adopté par le projet IPFC (cf. Partie 4.Chapitre 1.2.4.2.2). Les types de liaison ainsi caractérisés sont répartis selon la classification de Durand et al. (2011) en liaisons catégorique et variable (cf. CAT et VAR Tableau 43).

	Type de liaison	**Session**				
		S1	**S2**	**S3**	**S4**	**Total**
	Total CAT	**1271**	**1321**	**1313**	**1275**	**5180**
CATEGORIQUE	DET-ADJ	33	36	38	37	144
	DET-NOM	511	526	516	518	2071
	DET-PRO	1	2	0	0	3
	EXF-EXF	93	92	92	88	365
	NUM-NOM	107	119	114	106	446
	PDE-ADJ	0	1	0	0	1

[287] Parmi lesquelles nous pouvons distinguer: 5180 sites potentiels de liaison catégorique, 5560 sites potentiels de liaison variable, 90 occurrences de liaison erratique réalisées et 27 cas attestés de liaison non attendue.

	PDE-NOM	52	53	56	54	215
	PRO-AUX	171	186	201	183	741
	PRO-PRO	57	54	51	53	215
	PRO-VER	148	148	141	136	573
	VER-PRO	98	104	104	100	406
VARIABLE	**Total VAR**	**1367**	**1418**	**1386**	**1389**	**5560**
	ADJ-NOM	299	299	307	293	1198
	ADV-ADJ	34	42	35	42	153
	ADV-ADV	2	2	12	5	21
	ADV-DET	17	17	17	18	69
	ADV-INF	1	1	1	1	4
	ADV-NOM	1	0	1	2	4
	ADV-PPA	1	2	1	2	6
	ADV-PRO	16	17	17	18	68
	ADV-PRP	47	50	51	55	203
	AUX-ADV	27	18	18	17	80
	AUX-INF	0	0	0	1	1
	AUX-PPA	106	112	113	106	437
	AUX-PRP	41	38	37	35	151
	CON-PRO	56	51	52	58	217
	INF-ADV	0	1	0	1	2
	INF-DET	17	17	16	17	67
	INF-PRP	17	17	17	17	68
	NOM-ADJ	169	176	174	171	690
	NOM-ADV	0	0	1	0	1
	NOM-AUX	18	17	17	16	68
	NOM-PRP	17	17	17	17	68
	NOM-VER	1	1	0	0	2
	PPA-ADV	1	0	0	0	1
	PRO-INF	0	0	0	1	1
	PRO-PRP	18	17	17	18	70
	PRP-DET	30	22	23	26	101
	PRP-NAM	79	78	72	70	299
	PRP-NOM	19	19	19	19	76
	PRP-PRO	1	0	0	0	1
	VER-ADJ	52	62	53	53	220
	VER-ADV	131	144	138	139	552
	VER-DET	54	62	56	61	233
	VER-INF	24	21	16	22	83
	VER-NOM	0	2	0	0	2
	VER-PDE	0	2	0	0	2
	VER-PRP	71	94	88	88	341
	Total général	**2638**	**2739**	**2699**	**2664**	**10740**

Tableau 43 : Types de liaison potentielle catégorique et variable (étude longitudinale)

Les types de liaison les plus attestés dans notre corpus sont bien évidemment les contextes catégoriques tels que la liaison entre un déterminant[288] et le nom suivant (DET-NOM), entre des numéraux et le nom suivant (NUM-NOM), entre un verbe et le pronom clitique postposé (VER-PRO), entre le pronom personnel sujet et le verbe (PRO-VER) ou l'auxiliaire suivant (PRO-AUX) ou encore la liaison à l'intérieur des expressions figées (EXF-EXF). Néanmoins, le protocole de l'étude longitudinale (cf. Partie 4.Chapitre 1.4.3.1) nous a également permis d'enregistrer un nombre relativement important de sites potentiels de liaison variable (5560 occurrences). Parmi ces derniers, nous pouvons citer la liaison entre l'adjectif épithète et le nom qui suit (ADJ-NOM) ou entre un nom pluriel suivi d'un adjectif (NOM-ADJ).

Parmi l'ensemble des liaisons réalisées enchaînées, comme nous le verrons dans les pages qui suivent, seule une partie a été produite de manière conforme à la cible, à savoir avec la forme correcte de la CL cible. Celles-ci sont résumées dans le tableau ci-dessous (cf. Tableau 44) et reparties par contexte morphosyntaxique (type) de liaison et par typologie de liaison (catégorique et variable) (Durand et al. 2011). Les taux des réalisations sont calculés sur la base du nombre total des types de liaison potentielle catégorique et variable (cf. Tableau 43) (Durand et al. 2011).

Réalisation	**Session**								
	S1	**%**	**S2**	**%**	**S3**	**%**	**S4**	**%**	**Total**
Total CAT	888	**70%**	987	**75%**	1025	**78%**	1039	**81%**	3939
DET-ADJ	31	**94%**	34	**94%**	37	**97%**	32	**86%**	134
DET-NOM	406	**79%**	441	**84%**	435	**84%**	468	**90%**	1750
DET-PRO	1	**100%**	2	**100%**	0	**0%**	0	**0%**	3
EXF-EXF	46	**49%**	50	**54%**	55	**60%**	55	**63%**	206
NUM-NOM	41	**38%**	44	**37%**	47	**41%**	44	**42%**	176
PDE-ADJ	0	**0%**	1	**100%**	0	**0%**	0	**0%**	1
PDE-NOM	34	**65%**	35	**66%**	43	**77%**	52	**96%**	164
PRO-AUX	105	**61%**	138	**74%**	156	**78%**	147	**80%**	546
PRO-PRO	34	**60%**	41	**76%**	38	**75%**	38	**72%**	151
PRO-VER	116	**78%**	117	**79%**	124	**88%**	113	**83%**	470
VER-PRO	74	**76%**	84	**81%**	90	**87%**	90	**90%**	338
Total VAR	439	**32%**	527	**37%**	511	**37%**	587	**42%**	2064
ADJ-NOM	144	**48%**	176	**59%**	139	**45%**	154	**53%**	613
ADV-ADJ	22	**65%**	29	**69%**	25	**71%**	37	**88%**	113
ADV-ADV	1	**50%**	0	**0%**	2	**17%**	2	**40%**	5
ADV-DET	4	**24%**	5	**29%**	5	**29%**	4	**22%**	18
ADV-INF	0	**0%**	1	**100%**	1	**100%**	0	**0%**	2
ADV-NOM	0	**0%**	0	**0%**	0	**0%**	1	**50%**	1
ADV-PPA	0	**0%**	1	**50%**	1	**100%**	0	**0%**	2

[288] Nous rappelons que pour le déterminant, nous considérons à la fois les articles définis (ex. « les ») et indéfinis (ex. « des »), les adjectifs possessifs (ex. « son »), les adjectifs démonstratifs (ex. « ces ») et les adejctifs indéfinis (ex. « quelques »).

ADV-PRO	2	**13%**	2	**12%**	1	**6%**	7	**39%**	12
ADV-PRP	2	**4%**	4	**8%**	2	**4%**	3	**5%**	11
AUX-ADV	9	**33%**	7	**39%**	10	**56%**	8	**47%**	34
AUX-PPA	49	**46%**	63	**56%**	71	**63%**	71	**67%**	254
AUX-PRP	9	**22%**	6	**16%**	6	**16%**	7	**20%**	28
CON-PRO	16	**29%**	28	**55%**	28	**54%**	32	**55%**	104
INF-DET	0	**0%**	1	**6%**	0	**0%**	1	**6%**	2
INF-PRP	1	**6%**	1	**6%**	0	**0%**	0	**0%**	2
NOM-ADJ	18	**11%**	14	**8%**	16	**9%**	27	**16%**	75
NOM-VER	0	**0%**	1	**100%**	0	**0%**	0	**0%**	1
PRO-INF	0	**0%**	0	**0%**	0	**0%**	1	**100%**	1
PRO-PRP	4	**22%**	5	**29%**	4	**24%**	7	**39%**	20
PRP-DET	13	**43%**	8	**36%**	14	**61%**	12	**46%**	47
PRP-NAM	59	**75%**	69	**88%**	61	**85%**	66	**94%**	255
PRP-NOM	5	**26%**	7	**37%**	14	**74%**	12	**63%**	38
PRP-PRO	1	**100%**	0	**0%**	0	**0%**	0	**0%**	1
VER-ADJ	9	**17%**	14	**23%**	14	**26%**	24	**45%**	61
VER-ADV	49	**37%**	59	**41%**	62	**45%**	71	**51%**	241
VER-DET	8	**15%**	10	**16%**	12	**21%**	12	**20%**	42
VER-INF	0	**0%**	0	**0%**	1	**6%**	0	**0%**	1
VER-PRP	14	**20%**	16	**17%**	22	**25%**	28	**32%**	80

Tableau 44 : Comparaison longitudinale des taux de production des liaisons catégoriques et variables conformes à la CL cible par type de liaison

Aucune différence significative (test Kruskal-Wallis $p > 0,05$) n'a été enregistrée entre les taux de réalisation des liaisons catégoriques conformes à la cible issus des quatre séances de l'étude longitudinale ($p = 0,9$) (cf. Tableau 45, Figure 30).

Kruskal-Wallis chi carré	**Degré de liberté**	**p-value**
$x^2 = 0,49$	$df = 3$	$p = 0,92$ ($p > 0,05$)

Tableau 45 : Résultats du test statistique non paramétrique Kruskal-Wallis mené sur les données de production des liaisons catégoriques conformes à la CL cible (étude longitudinale)

De manière générale, le taux de production de liaison augmente de 70% à 81% entre la première et la dernière séance. Il existe néanmoins des contextes dans lesquels les apprenants semblent progresser davantage. C'est le cas par exemple des types EXF-EXF (ex. « de temps en temps ») ou encore PDE-NOM (ex. « la prise en charge des espaces »). En effet, ces contextes font état dès les premières séances de pourcentages de réalisation assez faibles (environ 50%-60%) qui augmenteront progressivement au cours de la période d'observation.

En revanche, d'autres types comme DET-NOM (ex. « les examens ») présentent dès le début des pourcentages élevés de productions (79%) jusqu'à atteindre 90% de réalisations à la fin de l'étude longitudinale.

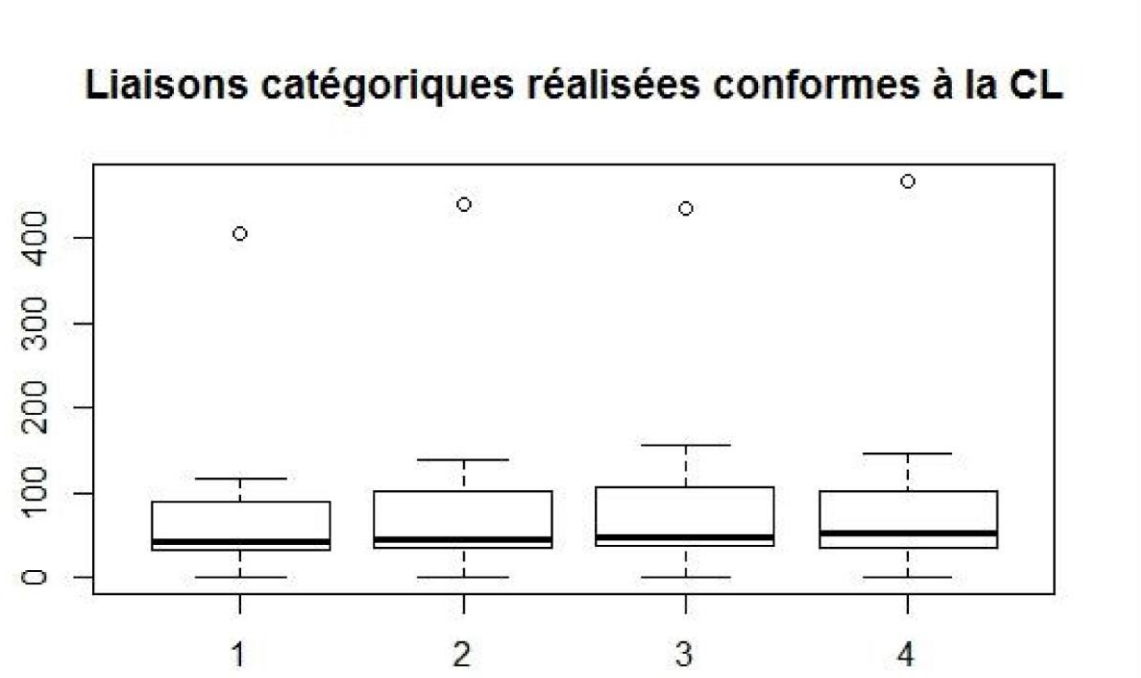


Figure 30 : Comparaison des taux de production des liaisons catégoriques conformes à la CL cible entre les quatre séances de l'étude longitudinale

Quant aux résultats des liaisons variables, l'augmentation des productions conformes à la CL cible de 31% (S1) à 42% (S4) des réalisations semble également ne pas être significative (p = 0,8) (cf. Tableau 46 ; cf. Figure 31).

Kruskal-Wallis khi carré	Degré de liberté	p-value
$x^2 = 0,63$	df = 3	p = 0,88 (p > 0,05)

Tableau 46 : Résultats du test statistique non paramétrique Kruskal-Wallis mené sur les données de production des liaisons variables conformes à la CL cible (étude longitudinale)

Cependant, comme nous le verrons (cf. Paragraphe 1.2.1.3), certains types de liaison tels que ADV-ADJ (ex. « très inquiet »), AUX-PPA (ex. « ils ont appris »), CON-PRO (ex. « quand on ») et VER-ADJ (« il est impossible ») sont caractérisés par une augmentation plus importante par rapport aux autres contextes variables (ex. « NOM-ADJ » « pâtes italiennes ») qui présentent tout au long des quatre séances de faibles pourcentages de réalisations conformes à la CL cible (11% -16%).

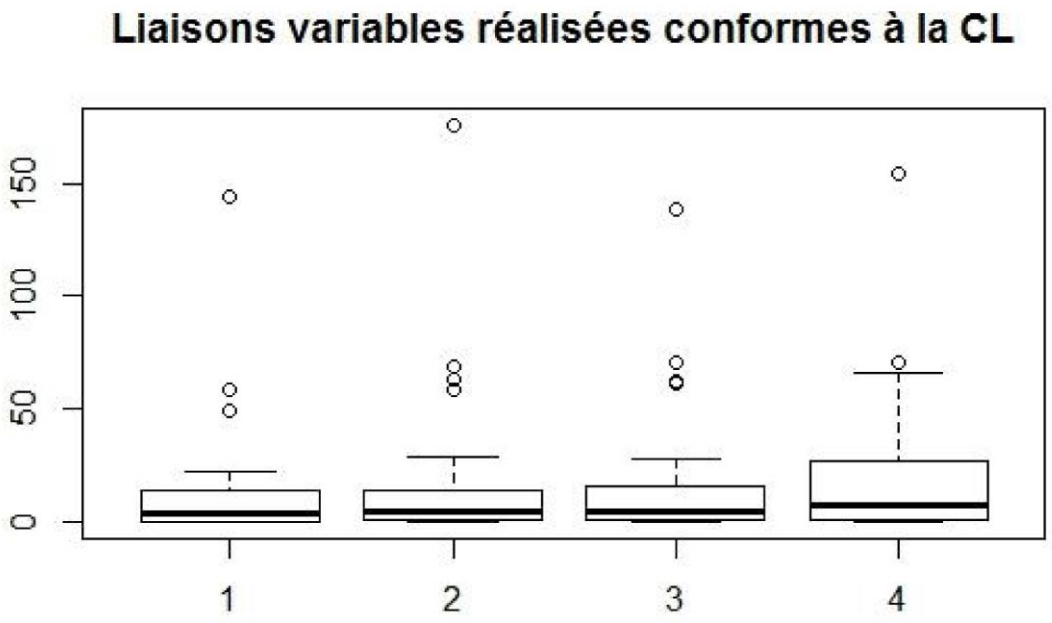


Figure 31 : Comparaison des taux de production des liaisons variables conformes à la CL cible entre les quatre séances de l'étude longitudinale

De manière générale, les résultats totaux montrent un écart non significatif entre les réalisations des liaisons catégoriques et variables conformes à la CL cible observées lors de la première (S1) et la dernière séance d'enregistrement (S4). Cependant, ce bilan général mérite une analyse plus approfondie des productions au niveau de la « micro-planification » (Racine 2014). Dans les pages qui suivent, nous nous attacherons à analyser plus en détail les types des réalisations de la liaison (enchaînée et non-enchaînée) ainsi que les formes des CL (conformes et non conformes à la CL cible) produites par les apprenants au cours de l'étude longitudinale.

1.2. Réalisation de la liaison

En ce qui concerne la réalisation de la liaison, il convient de rappeler que les données seront réparties en quatre catégories définies par le champ 5 du codage liaison IPFC (cf. Partie 4.Chapitre 1.2.4.2.4) : (a) absence de liaison (code 00) (dorénavant « Non-réalisation ») ; (b) liaison enchaînée (code 11) (dorénavant « Enchaînement ») ; (c) liaison non-enchaînée (code

12) (dorénavant « Non-enchaînement ») ; (d) des cas d'enchaînement incertain (code 13) (dorénavant « Enchaîn. incertain »)[289]. Dans les pages qui suivent nous séparerons du reste des réalisations les cas de CL précédées d'une « pause-hésitation » (Ågren 1973) temporelle ou intonationnelle (dorénavant « Pause-hésitation »)[290]. Le tableau ci-dessous (cf. Tableau 47) présente les résultats concernant les types de réalisation de la liaison observés tout au long de l'étude longitudinale. Les taux de productions ont été calculés sur la base du nombre de sites de liaisons potentielles catégoriques et variables (Durand et al. 2011).

Réalisation	**Session**								
	S1	**%**	**S2**	**%**	**S3**	**%**	**S4**	**%**	**Total**
Total CAT	**1271**	**100%**	**1321**	**100%**	**1313**	**100%**	**1275**	**100%**	**5180**
Non-réalisation	199	**16%**	184	**14%**	138	**11%**	132	**10%**	653
Enchaînement	1042	**82%**	1123	**85%**	1155	**88%**	1134	**89%**	4454
Non-enchaînement	0	**0%**	1	**0%**	1	**0%**	1	**0%**	3
Pause-hésitation	13	**1%**	13	**1%**	19	**1%**	8	**1%**	53
Enchaîn. incertain	0	**0%**	0	**0%**	0	**0%**	0	**0%**	0
NC	17	**1%**	0	**0%**	0	**0%**	0	**0%**	17
Total VAR	**1367**	**100%**	**1418**	**100%**	**1386**	**100%**	**1389**	**100%**	**5560**
Non-réalisation	738	**54%**	747	**53%**	689	**50%**	654	**47%**	2828
Enchaînement	546	**40%**	635	**45%**	640	**46%**	705	**51%**	2526
Non-enchaînement	6	**0%**	9	**1%**	3	**0%**	6	**0%**	24
Pause-hésitation	10	**1%**	6	**0%**	24	**2%**	12	**1%**	52
Enchaîn. incertain	0	**0%**	1	**0%**	1	**0%**	0	**0%**	2
NC	67	**5%**	20	**1%**	29	**2%**	12	**1%**	128

Tableau 47 : Résultats totaux de l'étude longitudinale – réalisation des liaisons catégoriques et variables

1.2.1. Non-réalisations de la liaison

Le tableau ci-dessous (cf. Tableau 48) montre que, toutes typologies de liaison confondues (catégorique et variable)[291], le taux de non-réalisation reste relativement stable bien que la plupart des non-réalisations correspondent à des contextes de liaison variable.

[289] Nous avons ajouté à ces différentes réalisations la catégorie des « non-compréhensions » (dorénavant NC) afin de rendre compte des cas de non-compréhensions des séquences cibles, enregistrées dans la tâche de répétition.

[290] Cette distinction a été adoptée par les auteurs en raison des considérations faites dans la note 37 (cf. Partie 1.Chapitre 2.2.1.1).

[291] Les résultats des types de liaison erratique feront l'objet du paragraphe 2.4.2.

Non-réalisation	Session								
	S1	%	S2	%	S3	%	S4	%	Total
Total CAT	**1271**	**100%**	**1321**	**100%**	**1313**	**100%**	**1275**	**100%**	**5180**
Non-réal.	199	**16%**	184	**14%**	138	**11%**	132	**10%**	653
Total VAR	**1367**	**100%**	**1418**	**100%**	**1386**	**100%**	**1389**	**100%**	**5560**
Non réal.	738	**54%**	747	**53%**	689	**50%**	654	**47%**	2828

Tableau 48 : Résultats totaux - taux de non-réalisation des liaisons catégoriques et variables (étude longitudinale)

Quant aux pourcentages de non-réalisations par typologie de liaison (catégorique et variable), aucune baisse significative n'a été enregistrée dans les cas des liaisons catégoriques et variables tout au long de la période d'observation. Les apprenants ne sembleraient pas produire un taux significativement plus important de liaisons catégoriques ou de liaisons variables à la fin de la période d'observation.

Néanmoins, nous pouvons constater que le taux de non-réalisations de la liaison catégorique (16%-10%) est moins important que celui de la liaison variable (54-47%). Cet écart ne semble pourtant pas augmenter au cours de la période d'observation. Autrement dit, si les apprenants omettent un plus grand nombre de liaisons dans des contextes de liaison variable que dans des cas de liaison catégorique, ils ne semblent pas améliorer davantage leur maîtrise de la liaison catégorique par rapport à celle de la liaison variable (cf. Tableau 48).

Cependant, si nous analysons plus en détail les types de liaison, nous pouvons identifier certains contextes au sein de l'ensemble des liaisons catégoriques et variables qui semblent le plus souvent donner lieu à des non-réalisations (cf. Tableau 49).

	Non-réalisation							
	Type de liaison	**S1**	**S2**	**S3**	**S4**	**Total Non réal.**	**Total**	**% Non réal.**
CATEGORIQUE	**Total CAT**	**199**	**184**	**138**	**132**	**653**	**5180**	**13%**
	DET-ADJ	2	2	0	5	9	144	**6%**
	DET-NOM	44	35	24	21	124	2071	**6%**
	EXF-EXF	29	28	18	22	97	365	**27%**
	NUM-NOM	47	51	43	37	178	446	**40%**
	PDE-NOM	5	8	6	1	20	215	**9%**
	PRO-AUX	23	17	18	13	71	741	**10%**
	PRO-PRO	15	11	9	13	48	215	**22%**
	PRO-VER	19	18	9	10	56	573	**10%**
	VER-PRO	15	14	11	10	50	406	**12%**
VARIABLE	**Total VAR**	**738**	**747**	**689**	**654**	**2828**	**5560**	**51%**
	ADJ-NOM	99	82	87	74	342	1198	**29%**
	ADV-ADJ	8	8	6	4	26	153	**17%**

	ADV-ADV	1	2	9	3	15	21	**71%**
	ADV-DET	13	12	12	13	50	69	**72%**
	ADV-INF	1	0	0	1	2	4	**50%**
	ADV-NOM	1	0	1	1	3	4	**75%**
	ADV-PPA	1	1	0	2	4	6	**67%**
	ADV-PRO	14	15	16	11	56	68	**82%**
	ADV-PRP	45	46	49	52	192	203	**95%**
	AUX-ADV	18	11	8	8	45	80	**56%**
	AUX-INF	0	0	0	1	1	1	**100%**
	AUX-PPA	51	48	37	32	168	437	**38%**
	AUX-PRP	32	32	31	28	123	151	**81%**
	CON-PRO	11	9	7	12	39	217	**18%**
	INF-ADV	0	1	0	1	2	2	**100%**
	INF-DET	17	16	16	16	65	67	**97%**
	INF-PRP	16	16	17	17	66	68	**97%**
	NOM-ADJ	127	131	125	117	500	690	**72%**
	NOM-ADV	0	0	1	0	1	1	**100%**
	NOM-AUX	18	17	17	16	68	68	**100%**
	NOM-PRP	17	17	17	17	68	68	**100%**
	PPA-ADV	1	0	0	0	1	1	**100%**
	PRO-PRP	14	12	13	11	50	70	**71%**
	PRP-DET	4	4	3	7	18	101	**18%**
	PRP-NAM	1	0	1	2	4	299	**1%**
	PRP-NOM	3	2	0	1	6	76	**8%**
	VER-ADJ	41	48	37	27	154	220	**70%**
	VER-ADV	59	62	56	50	227	552	**41%**
	VER-DET	45	52	43	49	189	233	**81%**
	VER-INF	24	21	14	22	81	83	**98%**
	VER-NOM	0	2	0	0	2	2	**100%**
	VER-PDE	0	2	0	0	2	2	**100%**
	VER-PRP	55	78	66	59	258	341	**76%**

Tableau 49 : Taux de non-réalisation des liaisons catégoriques et variables par type de liaison (étude longitudinale)

1.2.1.1 Contextes catégoriques

Après déterminant

Dans le cas des contextes catégoriques, le taux de non-réalisations est très faible (6%) dans les types DET-ADJ et DET-NOM (cf. Tableau 49). De plus, l'analyse des non-réalisations menée parmi les occurrences du type DET-NOM (cf. Tableau 50) montre qu'au sein de ce contexte les non-réalisations apparaissent le plus souvent après l'adjectif possessif « leurs » et après l'adjectif indéfini « quelques », même lors des dernières sessions de la période d'observation.

En revanche, l'étude longitudinale des non-réalisations au sein du type DET-NOM révèle une évolution qui conduit à une maîtrise presque totale de la liaison après les articles et les adjectifs démonstratifs.

DET-NOM						
M1	**Non-réalisation**	**Enchaînement**	**Non-enchaînement**	**Pause-hésitation**	**NC**	**Total**
Ces	6	138	1	2	0	147
Cet	1	7	0	0	0	8
Des	7	983	0	2	2	994
Les	21	515	0	6	0	542
Leurs	32	10	0	3	0	45
Mes	0	27	0	0	0	27
Mon	1	1	0	0	0	2
Plusieurs	1	0	0	0	0	1
Quelques	32	43	0	0	0	75
Ses	0	1	0	0	0	1
Son	23	183	0	0	0	206
Un	0	22	0	0	1	23
Total général	**124**	**1930**	**1**	**13**	**1**	**2071**

Tableau 50 : Taux de non-réalisation et des réalisations de la liaison du type DET-NOM par M1 (étude longitudinale)

Avant et après pronoms personnels et clitiques

En ce qui concerne les autres types de liaison catégorique, la liaison avant ou après pronom présente un taux de non-réalisations d'environ 10% pour les types PRO-AUX, PRO-VER et VER-PRO et de 22% pour le type PRO-PRO (cf. Tableau 51).

Au sein du type PRO-AUX, la liaison est presque entièrement maîtrisée après des pronoms personnels sujets tels que « ils », « nous », « on » et « vous ». En revanche, elle est moins fréquente après le pronom indéfini « tout » dans la séquence « tout est fait» issue de la lecture du texte PFC (cf. Annexe 4). Quant au pronom clitique « en », nous pouvons supposer que les non-réalisations sont liées d'un côté à la difficulté de la séquence dans laquelle ce pronom se trouve (« on en a vu », toujours issue de la lecture du texte PFC) qui contient une suite de clitiques, et de l'autre à des raisons d'ordre d'apprentissage liées aux itinéraires acquisitionnels (cf. Partie 3.Chapitre 1.4.3.1.4).

Quant au type PRO-VER les apprenants produisent presque toujours des liaisons après les pronoms « nous », « on » et « vous ». Cependant, dans le cas du pronom « ils », nous observons un plus grand nombre de non-réalisations par rapport au type précédent (PRO-AUX) sans doute en raison de la cohésion syntaxique qui lie le pronom « ils » et le verbe auxiliaire suivant. Comme dans les cas du type PRO-AUX, la liaison est moins fréquente après des

pronoms clitiques tels que « en » qu'après des pronoms personnels sujets. Autrement dit, la liaison avant ou après les pronoms clitiques semble donner lieu à un nombre important de non-réalisations comme le montre le taux de non-réalisations des types PRO-PRO (22%) et VER-PRO (12%). Le taux de non-réalisations augmente sensiblement (22%) devant un pronom clitique tel que « en » et « y » au sein du type PRO-PRO (« comme on en a vu », « nous en installions » « ils y ont vu» issues des textes lus). Ces résultats sont d'autant plus intéressants si nous considérons le fort degré de cohésion syntaxique qui lie les pronoms clitiques aux constituants dont ils dépendent[292].

En revanche, en ce qui concerne le cas du type VER-PRO, la production devant des pronoms clitiques est moins régulière. Si les apprenants semblent presque toujours réaliser la liaison dans les séquences « allez-y », « allons-y », « dit-elle » « donnez-en », ils ont en revanche tendance à produire moins de réalisations dans des séquences telles que « prenez-en », « sont-elles ». Il convient toutefois de souligner, comme nous le verrons dans les pages qui suivent, que les séquences « donnez-en » et « allons-y », contrairement à « prenez-en » et « allez-y » (issues d'une tâche de lecture), étaient présentées au cours de la tâche de répétition. Il est donc possible de supposer que l'écart du taux de non-réalisations enregistré entre ces séquences relève d'un effet de tâche. Cependant, l'absence de différences significatives entre le taux de non-réalisations de « allons-y » (en répétition) et « allez-y » (en lecture) suggère que la plus grande fréquence de ces séquences par rapport aux précédentes pourrait jouer un rôle dans la production des liaisons chez des apprenants FLE[293].

M1	Non-réalisation	Enchaînement	Non-enchaînement	Pause-hésitation	NC	Total
PRO-AUX	**71**	**655**	**0**	**9**	**6**	**741**
En	15	30	0	4	0	49
Ils	5	284	0	1	6	296
Les	2	2	0	0	0	4
Nous	5	157	0	3	0	165
On	1	88	0	1	0	90
Tout	42	31	0	0	0	73
Vous	1	63	0	0	0	64
PRO-PRO	**48**	**161**	**0**	**6**	**0**	**215**
En	1	0	0	1	0	2
Ils	30	34	0	5	0	69
Nous	8	76	0	0	0	84
On	9	51	0	0	0	60
PRO-VER	**56**	**509**	**0**	**5**	**3**	**573**
Elles	0	1	0	0	0	1

[292] Nous nous attacherons dans le prochain chapitre (cf. Partie 4.Chapitre 3.1.1.1) à apporter une réponse à ces résultats apparemment contradictoires.

[293] Nous développerons ce sujet dans les pages qui suivent (cf. Paragraphe 1.2.1.2).

En	19	48	0	2	0	69
Ils	28	112	0	2	1	143
Les	1	0	0	0	0	1
Nous	1	149	0	1	0	151
On	7	66	0	0	0	73
Vous	0	133	0	0	2	135
VER-PRO	**50**	**350**	**0**	**4**	**2**	**406**
Allez	4	64	0	0	0	68
Allons	0	67	0	0	0	67
Dit	4	61	0	0	0	65
Donnez	0	67	0	0	2	69
Prenez	17	49	0	2	0	68
Sont	25	42	0	2	0	69
Total	**650**	**4451**	**3**	**53**	**15**	**5172**

Tableau 51:Taux de non-réalisation et des réalisationd de la liaison des types PRO-AUX, PRO-PRO, PRO-VER et VER-PRO par M1 (étude longitudinale)

Dans des expressions figées

L'analyse des occurrences du type de liaison EXF-EXF (cf. Tableau 52) montre des écarts très importants entre les taux de réalisation des différentes expressions figées. Si aucune non-réalisation n'a été enregistrée dans le cas de la séquence « de plus en plus », la production de la liaison est moins stable dans le groupe figé « jeux olympiques » auquel les apprenants sont sans doute moins exposés. Au sein de l'adverbe composé « de temps en temps », comme nous le verrons dans les pages qui suivent, nous avons enregistré de nombreux cas de prononciations de la consonne /p/ enchaînée. Les non-réalisations dans ce contexte pourraient donc s'expliquer par une insécurité de l'apprenant face à la prononciation de cette consonne.

En effet, si certaines unités polylexicales semblent être acquises (notamment les plus fréquentes comme « de plus en plus », « en effet » et « plus ou moins »), d'autres n'ont pas atteint un degré suffisant de figement chez les apprenants italophones (« jeux olympiques »). Bien que les apprenants soient sans doute conscients du caractère « obligatoire » de la liaison à l'intérieur des expressions figées, ils ne sont pas toujours en mesure de saisir le niveau de figement qui les caractérise et d'acquérir en conséquence le fonctionnement de la liaison à l'intérieur de ces séquences.

	EXF-EXF					
M1	**Non-réalisation**	**Enchaînement**	**Non-enchaînement**	**Pause-hésitation**	**NC**	**Total**
En effet	5	71	0	0	0	76
Jeux olympiques	54	13	0	4	0	71
De plus en plus	0	73	0	0	3	76
De temps en temps	38	100	0	0	0	138
Tout à fait	0	2	0	0	0	2

Vingt et un	0	2	0	0	0	2
Total général	**97**	**261**	**0**	**4**	**3**	**365**

Tableau 52 : Taux de non-réalisation et des réalisations de la liaison des types EXF-EXF (étude longitudinale)

Après les adjectifs numéraux

Dans le cas du type de liaison (NUM-NOM) (cf. Tableau 53), le plus grand nombre de non-réalisations a été observé après le numéral « cents » (134 non-réalisations / 208 occurrences totales) dans les séquences « deux cents acquéreurs » « huit cents euros » et « deux cents habitants », issues de la lecture du texte journalistique (cf. Partie 4.Chapitre 1.4.3.1.5). Ce type de liaison s'avère sans doute difficile pour l'apprenant italophone à cause du transfert du mot italien correspondant *cento* (« cent » en français) qui d'une part est invariable et d'autre part contient la consonne /t/[294]. Dans d'autres séquences issues du type NUM-NOM, telles que la liaison après le numéral « deux », les non-réalisations de liaison enregistrent une baisse significative et arrivent à 0% dans tous les autres contextes. En effet, aucune absence de liaison n'a été enregistrée après les autres numéraux, notamment « dix », « trente-huit », « trente-six », « trois » et « un ».

NUM-NOM						
M1	**Non-réalisation**	**Enchaînement**	**Non-enchaînement**	**Pause-hésitation**	**NC**	**Total**
Cents	134	68	1	5	0	208
Deux	43	106	0	0	0	149
Dix	0	68	0	0	0	68
Quatre-vingts	0	0	0	1	0	1
Trente-huit	0	3	0	0	0	3
Trente-six	0	1	0	0	0	1
Trois	0	5	0	0	0	5
Un	0	1	0	0	0	1
Vingt	1	9	0	0	0	10
Total général	**178**	**261**	**1**	**6**	**0**	**446**

Tableau 53 : Taux de non-réalisation et des réalisations de la liaison du type NUM-NOM par M1 (étude longitudinale)

1.2.1.2 Le rôle de la fréquence de cooccurrence

Malgré la persistance de non-réalisations dans les contextes catégoriques au cours de l'étude longitudinale, nous avons recueilli des résultats intéressants issus de la tâche de lecture des séquences isolées de liaison. Il convient de rappeler qu'au sein de cette tâche, nous avons établi des paires de sites de liaison potentielle relevant des types DET-NOM, PRO-VER et

[294] En effet, les apprenants ont tendance à enchaîner la consonne /t/.

VER-PRO constitués par des séquences M1-M2 ayant des indices de cooccurrence différents. L'analyse de ces résultats (cf. Tableau 54) nous a permis de mettre en évidence deux points importants.

	Lecture			
	Non-réalisation	**Enchaînement**	**Pause-hésitation**	**Total**
DET-NOM	**2**	**407**	**0**	**409**
Des	2	407	0	409
arbres	0	69	0	69
âtres	0	68	0	68
appartements	0	67	0	67
appontements	1	67	0	68
PRO-VER	**1**	**136**	**0**	**137**
Nous	1	136	0	137
amarrons	1	68	0	69
avons	0	68	0	68
VER-PRO	**21**	**113**	**2**	**136**
Allez-	4	64	0	68
y	4	64	0	68
Prenez-	17	49	2	68
en	17	49	0	66
Total général	**24**	**656**	**2**	**682**

Tableau 54 : Taux de non-réalisation et des réalisations de la liaison des paires de séquences ayant de indices de cooccurrence différents (lecture-étude longitudinale)

D'un côté la liaison après l'article indéfini « des » et après le pronom « nous » n'est pas sensible à des effets de fréquence de coccurrence des séquences M1-M2. Autrement dit, les apprenants ne réalisent pas plus de liaisons devant le mot « arbres » que devant le mot « âtres ». Il en va de même pour le cas de la liaison après les pronoms personnels sujets « nous » et « vous ». D'autre part, en revanche, des écarts de réalisation ont été observés au sein des types VER-PRO (différence de réalisations entre « allez-y » et « prenez-en » où la fréquence de cooccurrence de la séquence semble jouer un rôle crucial dans la réalisation de la liaison (cf. Paragraphe 1.2.1.1).

1.2.1.3 Contextes variables

Afin d'évaluer la maîtrise de la liaison variable chez les apprenants italophones, il nous a paru intéressant, en ligne avec l'approche proposée par le projet IPFC (cf. Partie 4.Chapitre 1.1), de comparer les productions de notre corpus d'apprenants aux réalisations de la liaison observées chez les natifs et issues de l'analyse fréquentielle menée sur le corpus PFC (Durand et al. 2009) (cf. Partie 2.Chapitre 2.2.2.1). Une telle démarche permet d'éviter le recours à des modèles de référence normatifs (Fouché 1959).

Le tableau ci-dessous (cf. Tableau 55) montre les taux de non-réalisations observés pour chaque type de liaison variable chez les apprenants FLE d'une part et chez les sujets natifs de l'autre[295].

Type	% Non-réalisation apprenants FLE	% Non-réalisation natifs (PFC)
ADJ-NOM	29%	18%
ADV-ADJ	17%	49%
ADV-ADV	71%	92%
ADV-DET	72%	92%
ADV-INF	50%	/
ADV-NOM	75%	95%
ADV-PPA	67%	/
ADV-PRO	82%	99%
ADV-PRP	95%	90%
AUX-ADV	56%	100%
AUX-INF	100%	/
AUX-PPA	38%	69%
AUX-PRP	81%	57%
INF-ADV	100%	100%
INF-DET	97%	100%
INF-PRP	97%	100%
NOM-ADJ	72%	91%
NOM-ADV	100%	/
NOM-AUX	100%	/
NOM-PRP	100%	/
PPA-ADV	100%	99%
PRO-PRP	71%	33%
PRP-DET	18%	27%
PRP-NAM	1%	/
PRP-NOM	8%	5%
VER-ADJ	70%	82%
VER-ADV	41%	95%
VER-DET	81%	87%
VER-INF	98%	/
VER-NOM	100%	88%
VER-PDE	100%	/
VER-PRP	76%	93%

Tableau 55 : Comparaison entre le taux de non-réalisation de la liaison variable chez les apprenants FLE (corpus IPFC-italien Rome) et chez les natifs (corpus PFC)

Cette analyse comparative montre que le taux de non-réalisation des liaisons variables chez les apprenants ne s'éloigne pas de manière significative de celui des natifs. En revanche, les apprenants italophones semblent réaliser dans certains types plus de liaisons par rapport aux

[295] Pour plus de détails sur les résultats de l'analyse fréquentielle menée sur le corpus des natifs voir la deuxième annexe (cf. Annexe 2).

locuteurs natifs. Comme dans le cas de l'analyse des liaisons catégoriques, il nous paraît nécessaire d'étudier de plus près l'ensemble des types de liaison variable afin d'identifier les occurrences susceptibles d'entraîner le plus grand nombre de non-réalisations.

Après adjectif

La liaison après adjectif épithète, comme nous l'avons vu (cf. Partie 2.Chapitre 2.2.2.3.3), est caractérisée par un nombre restreint d'adjectifs susceptibles d'être antéposés en particulier au singulier. Nous avons cité parmi ceux-ci les adjectifs « bon », « petit », « grand » etc. La liste d'adjectifs est bien évidement plus large au pluriel étant donné que la liaison constitue dans ce contexte une marque morphologique.

Les résultats de notre corpus semblent indiquer que les apprenants présentent un taux de non-réalisations relativement proche de celui des natifs (18%). Toutefois, la plupart des occurrences de ce type de liaison correspondent à des séquences contenant les adjectifs « petit » et « grand » (cf. Tableau 56).

	ADJ-NOM					
M1	**Non-réalisation**	**Enchaînement**	**Non-enchaînement**	**Pause-hésitation**	**NC**	**Total**
Autres	1	0	0	0	0	1
Bon	0	1	0	0	0	1
Grand	200	403	3	8	20	634
Mêmes	1	0	0	0	0	1
Petit	109	341	7	3	33	493
Petits	1	0	0	0	0	1
Premier	30	37	0	0	0	67
Total général	**342**	**782**	**10**	**11**	**53**	**1198**

Tableau 56 : Taux de non-réalisation et des réalisations de la liaison du type ADJ-NOM par M1 (étude longitudinale)

Malgré ces limites d'ordre méthodologique, les résultats de l'étude longitudinale nous montrent une réduction progressive des non-réalisations de la liaison (cf. Tableau 56) dans ce contexte qui cependant s'accompagne de l'émergence, comme nous le verrons, de formes orthographiques de la liaison dans le cas de l'adjectif « grand », l'un des plus représentés dans notre corpus.

Après adverbe

Quant à la réalisation de la liaison après adverbe, étant donné le faible nombre d'occurrences enregistrées, il est difficile de formuler des conclusions pour chaque type de liaison décrit dans le tableau ci-dessous (cf. Tableau 57). Néanmoins, nous pouvons observer que la liaison est souvent produite après les adverbes monosyllabiques « très » (12%) et « plus »

(25%) devant un adjectif (type ADV-ADJ) et qu'elle est en revanche non réalisée entre un adverbe et une préposition (type ADV-PRP) dans la plupart des cas (94%). L'écart entre ces deux contextes de liaison a été également relevé chez les natifs (cf. Tableau 55).

	Non-réalisation	Enchaînement	Non-enchaînement	Pause-hésitatation	NC	Total
ADV-ADJ	**26**	**126**	**0**	**1**	**0**	**153**
Assez	1	0	0	0	0	1
Plus	14	55	0	1	0	70
Très	9	71	0	0	0	80
Vraiment	2	0	0	0	0	2
ADV-ADV	**15**	**6**	**0**	**0**	**0**	**21**
Mais	7	3	0	0	0	10
Pas	6	2	0	0	0	8
Plus	0	1	0	0	0	1
Puis	2	0	0	0	0	2
ADV-DET	**50**	**18**	**1**	**0**	**0**	**69**
Pas	50	17	1	0	0	68
Seulement	0	1	0	0	0	1
ADV-INF	**2**	**2**	**0**	**0**	**0**	**4**
Bien	0	1	0	0	0	1
Pas	2	0	0	0	0	2
Plus	0	1	0	0	0	1
ADV-NOM	**3**	**1**	**0**	**0**	**0**	**4**
Moins	1	0	0	0	0	1
Pas	1	0	0	0	0	1
Plus	0	1	0	0	0	1
Seulement	1	0	0	0	0	1
ADV-PPA	**4**	**2**	**0**	**0**	**0**	**6**
Bien	2	0	0	0	0	2
Jamais	0	1	0	0	0	1
Pas	0	1	0	0	0	1
Toujours	2	0	0	0	0	2
ADV-PRO	**56**	**12**	**0**	**0**	**0**	**68**
Mais	56	12	0	0	0	68
ADV-PRP	**192**	**11**	**0**	**0**	**0**	**203**
Mais	1	0	0	0	0	1
Pas	3	1	0	0	0	4
Plus	120	10	0	0	0	130
Toujours	68	0	0	0	0	68
Total général	**348**	**178**	**1**	**1**	**0**	**528**

Tableau 57 : Taux de non-réalisation et des réalisations de la liaison du type ADV-ADJ, ADV-ADV, ADV-DET, ADV-INF, ADV-NOM, ADV-PPA, ADV-PRO, ADV-PRP par M1 (étude longitudinale)

Après nom

La liaison après noms est caractérisée par un pourcentage très élevé de non-réalisations, une tendance qui avait déjà émergé de l'analyse des données des natifs extraites du corpus PFC (Durand, Laks et Lyche. 2009) (91% de non-réalisations cf. Tableau 55). Les occurrences enregistrées chez les apprenants sont issues principalement des tâches de lecture. Cependant, bien que les apprenants soient exposés aux formes écrites des séquences, ils réalisent un nombre très faible de liaisons dans ce contexte.

Pour ce qui est des données extraites d'autres tâches que celles de lecture, nous avons remarqué la présence de non-réalisations même dans la tâche de répétition. Ce résultat pourrait constituer un indice de la représentation que les apprenants ont de la liaison de ce contexte qui fait apparaître des faiblesses même au niveau perceptif. Les séquences « des essais intéressants » et « des efforts inutiles » proposées dans la tâche de répétition possèdent un taux de non-réalisations respectivement de 23% et 8% (cf. Tableau 58).

Cet aspect, que nous aurons l'occasion d'approfondir dans les pages qui suivent (cf. Paragraphe 3), suggère que même en répétition la liaison après nom est une forme marquée à laquelle l'apprenant préfère le plus souvent une variante sans liaison.

	Non-réalisation	Enchaînement	Non-enchaînement	Enchain. Incertain	Pause-hésitation	NC	Total
NOM-ADJ	**500**	**157**	**7**	**1**	**19**	**6**	**690**
Circuits	63	5	0	0	0	0	68
Cocons	63	7	1	0	1	0	72
Conseils	50	14	0	0	4	0	68
Cultures	1	0	0	0	0	0	1
Efforts	6	51	5	1	0	6	69
Émeutes	64	0	0	0	0	0	64
Enfants	36	23	1	0	8	0	68
Essais	16	48	0	0	3	0	67
Intérêts	61	6	0	0	3	0	70
Langues	4	1	0	0	0	0	5
Pâtes	68	0	0	0	0	0	68
Relations	1	0	0	0	0	0	1
Visites	67	2	0	0	0	0	69
NOM-ADV	**1**	**0**	**0**	**0**	**0**	**0**	**1**
Jours	1	0	0	0	0	0	1
NOM-AUX	**68**	**4**	**0**	**0**	**0**	**0**	**68**
Fanatiques	68	0	0	0	0	0	68
NOM-PRP	**68**	**2**	**0**	**0**	**0**	**0**	**68**
Amies	1	0	0	0	0	0	1
Chemises	67	0	0	0	0	0	67
NOM-VER	**0**	**1**	**0**	**0**	**1**	**0**	**2**
Résidents	0	1	0	0	1	0	2
Total général	**637**	**158**	**7**	**1**	**20**	**6**	**829**

Tableau 58 : Taux de non-réalisation et des réalisations de la liaison des types NOM-ADJ, NOM-ADV, NOM-AUX, NOM-PRP et NOM-VER par M1 (étude longitudinale)

Après préposition

En ce qui concerne la liaison après préposition, nous n'avons enregistré que des prépositions monosyllabiques (le plus souvent contenues dans les tâches de lecture oralisée) après lesquelles le taux de non-réalisations apparaît assez faible (cf. Tableau 59). Autrement dit les productions de la liaison après les prépositions monosyllabiques ne s'éloignent pas de celles des natifs (cf. Tableau 55). Toutefois, les cas de réalisations après la préposition « dans », comme nous le verrons dans les pages qui suivent (cf. Paragraphe 1.3.2), sont caractérisés par un nombre important de productions orthographiques. Quant à la liaison après la préposition « en », il convient de procéder avec prudence dans l'analyse des non-réalisations dans ce contexte. En effet, il est à notre avis difficile d'établir dans quelle mesure la réalisation de la liaison en [n] relève d'une volonté réelle de produire la liaison ou de la dénasalisation de la voyelle nasale qui précède la CL et donc de l'enchaînement de son appendice nasal.

	Non-réalisation	Enchaînement	Non-echaînement	Pause-hésitation	NC	Total
PRP-DET	**18**	**76**	**3**	**4**	**0**	**101**
Chez	1	2	0	0	0	3
Dans	17	74	3	4	0	98
PRP-NAM	**4**	**255**	**0**	**1**	**39**	**299**
En	4	255	0	1	39	298
PRP-NOM	**6**	**68**	**0**	**2**	**0**	**76**
Dans	0	0	0	1	0	1
En	0	4	0	0	0	4
Sans	6	64	0	0	0	70
PRP-PRO	**0**	**1**	**0**	**0**	**0**	**1**
Chez	0	1	0	0	0	1
Total général	**28**	**400**	**3**	**7**	**39**	**477**

Tableau 59 : Taux de non-réalisation et des réalisations de la liaison des types PRP-DET, PRP-NAM, PRP-NOM et PRP-PRO par M1 (étude longitudinale)

Après verbe

Les résultats concernant la liaison après verbe rejoignent les données des différentes analyses menées sur les corpus des natifs (De Jong 1988 ; Mallet 2008) ainsi que les résultats de notre analyse fréquentielle (cf. Partie 2.Chapitre 2.2.2.2) qui indiquent que la liaison est plus souvent réalisée après la forme verbale (non auxiliaire) « est » qu'après les formes « sont » et « suis » du verbe « être ». En revanche, l'absence de non-réalisations après le verbe « sommes » s'explique encore une fois par le fait que cette séquence est contenue dans la tâche de répétition (cf. Tableau 60). L'analyse de l'ensemble des données longitudinales et transversales nous permettra dans les pages qui suivent de mieux évaluer cet effet de la tâche (cf. Paragraphe 3).

	Non-réalisation	Enchaînement	Non-enchaînement	Enchaîn. incertain	Pause-hésitation	NC	Total
VER-ADJ	**154**	**63**	**1**	**0**	**2**	**0**	**220**
Est	95	45	0	0	1	0	141
Était	5	0	0	0	0	0	5
Sont	50	17	1	0	0	0	68
Suis	3	1	0	0	0	0	4
Vient	0	0	0	0	1	0	1
VER-ADV	**227**	**318**	**1**	**1**	**1**	**4**	**552**
Avait	1	0	0	0	0	0	1
Descend	7	60	1	0	0	0	68
Devons	0	1	0	0	0	0	1
Dois	2	1	0	0	0	0	3
Est	2	0	0	0	0	0	2
Était	1	0	0	0	0	0	1
Habitent	68	0	0	0	0	0	68
Parlent	1	0	0	0	0	0	1
Pouvais	1	0	0	0	0	0	1
Sais	1	0	0	0	0	0	1

Sommes	0	68	0	0	0	0	68
Sont	54	75	0	1	0	0	130
Sort	0	65	0	0	0	4	69
Suis	1	0	0	0	0	0	1
Vient	19	47	0	0	0	1	67
Vivent	65	1	0	0	0	0	66
Voudrais	4	0	0	0	0	0	4
Sert	0	0	0	0	0	2	2
VER-DET	**189**	**42**	**0**	**0**	**2**	**0**	**233**
Avait	1	0	0	0	0	0	1
Est	49	39	0	0	2	0	90
Étais	1	0	0	0	0	0	1
Était	3	0	0	0	0	0	3
Faut	65	3	0	0	0	0	68
Préparent	68	0	0	0	0	0	68
Sont	1	0	0	0	0	0	1
Suis	1	0	0	0	0	0	1
VER-INF	**81**	**1**	**0**	**0**	**1**	**0**	**83**
Devais	1	0	0	0	0	0	1
Devrais	4	0	0	0	0	0	4
Dois	4	0	0	0	0	0	4
Fait	1	0	0	0	0	0	1
Faut	2	0	0	0	0	0	2
Peux	3	0	0	0	0	0	3
Préférais	1	0	0	0	0	0	1
Serais	2	0	0	0	0	0	2
Voudrais	7	0	0	0	0	0	7
Voulais	2	0	0	0	0	0	2
Voulons	54	1	0	0	1	0	56
VER-NOM	**2**	**0**	**0**	**0**	**0**	**0**	**2**
Suis	2	0	0	0	0	0	2
VER-PDE	**2**	**0**	**0**	**0**	**0**	**0**	**2**
Suis	2	0	0	0	0	0	2
VER-PRP	**258**	**82**	**0**	**0**	**1**	**0**	**341**
Est	79	49	0	0	1	0	129
Étais	1	0	0	0	0	0	1
Fleurissent	62	5	0	0	0	0	67
Sent	59	8	0	0	0	0	67
Sommes	2	0	0	0	0	0	2
Sont	49	19	0	0	0	0	68
Suis	5	1	0	0	0	0	6
Vais	1	0	0	0	0	0	1
Total général	**913**	**506**	**2**	**1**	**7**	**4**	**1433**

Tableau 60 : Taux de non-réalisation et des réalisations de la liaison des types VER-ADJ, VER-ADV, VER-DET, VER-INF, VER-NOM, VER-PDE et VER-PRP par M1 (étude longitudinale)

Après verbe auxiliaire

Quant à la liaison après verbe auxiliaire, nous avons notamment recueilli des formes verbales « ont », « est » et « sont » après lesquelles la liaison semble être fortement influencée, même chez l'apprenant de FLE, par des factures syntaxiques. La fréquence des non-réalisations est en effet plus faible entre le verbe auxiliaire et le participe passé (AUX-PPA) (38%) qu'entre le verbe auxiliaire et la préposition (AUX-PRP) (81%) et le verbe auxiliaire et l'adverbe suivant (AUX-ADV) (56%) (cf. Tableau 61). De plus dans le contexte AUX-PPA les apprenants semblent produire plus de liaisons que les locuteurs natifs (cf. Tableau 55).

	Non-réalisation	Enchaînement	Non-enchaînement	Pause-hésitation	NC	Total
AUX-ADV	**45**	**35**	**0**	**0**	**0**	**80**
Ont	37	32	0	0	0	69
Sont	8	3	0	0	0	11
AUX-INF	**1**	**0**	**0**	**0**	**0**	**1**
Vais	1	0	0	0	0	1
AUX-PPA	**168**	**257**	**1**	**6**	**5**	**437**
Est	1	0	0	0	0	1
Ont	155	250	1	5	4	415
Sommes	1	2	0	1	0	4
Sont	4	2	0	0	0	6
Suis	7	3	0	0	0	10
AUX-PRP	**123**	**28**	**0**	**0**	**0**	**151**
Est	120	21	0	0	0	141
Sommes	0	1	0	0	0	1
Suis	3	6	0	0	0	9
Total général	**337**	**320**	**1**	**6**	**5**	**669**

Tableau 61 : Taux de non-réalisation et des réalisations de la liaison des types AUX-ADV, AUX-INF, AUX-PPA et AUX-PRP par M1 (étude longitudinale)

Après verbe infinitif

La liaison après verbe infinitif n'est quant à elle presque jamais réalisée (cf. Tableau 62) par les apprenants comme d'ailleurs par les natifs (cf. Partie 2.Chapitre 2.2.2.2).

	Non-réalisation	Enchaînement	Non-enchaînement	Pause-hésitation	NC	Total
INF-ADV	**2**	**0**	**0**	**0**	**0**	**2**
Commencer	1	0	0	0	0	1
Fréquenter	1	0	0	0	0	1
INF-DET	**65**	**2**	**0**	**0**	**0**	**67**
Provoquer	65	2	0	0	0	67
INF-PRP	**66**	**2**	**0**	**0**	**0**	**68**
Trouver	66	2	0	0	0	68
Total général	**133**	**4**	**0**	**0**	**0**	**137**

Tableau 62 : Taux de non-réalisation et des réalisations de la liaison des types INF-ADV, INF-DET et INF-PRP par M1 (étude longitudinale)

1.2.2. Réalisations

La comparaison des données de notre corpus avec les productions des natifs a montré des similarités concernant les taux de non-réalisation des différents types de liaison variable. Il reste cependant à définir la typologie des réalisations de la liaison chez les apprenants de FLE. Si la plupart des liaisons sont produites avec enchaînement, il existe néanmoins des exemples de liaisons non-enchainées selon la définition donnée par Encrevé (1988). Comme nous l'avons déjà évoqué dans les paragraphes précédents (cf. Paragraphe 1.2), nous séparons du reste des réalisations les cas des CL précédées d'une pause temporelle ou intonationnelle (cf. « Pause-hésitation ») qui donnent la perception que la CL appartient au M2 (Ågren 1973).

1.2.2.1 Enchaînement

Les liaisons enchaînées correspondent à la presque totalité des liaisons réalisées. Les apprenants produisent des liaisons enchaînées à hauteur de 82%-89% dans des contextes catégoriques et de 39%-50% dans des contextes variables (cf. Tableau 63).

	Session								
	S1	**%**	**S2**	**%**	**S3**	**%**	**S4**	**%**	**Total**
Total CAT	**1271**	**100%**	**1321**	**100%**	**1313**	**100%**	**1275**	**100%**	**5180**
Enchaînement	1042	**82%**	1123	**85%**	1155	**88%**	1134	**89%**	4454
Total VAR	**1367**	**100%**	**1418**	**100%**	**1386**	**100%**	**1389**	**100%**	**5560**
Enchaînement	546	**40%**	635	**45%**	640	**46%**	705	**51%**	2526

Tableau 63 : Résultats totaux - taux d'enchaînement des liaisons catégoriques et variables (étude longitudinale)

1.2.2.2 Non-enchaînement

Contrairement aux études précédentes (Mastromonaco 1999 ; Howard 2005 ; Howard 2013 ; Thomas 2002 ; De Moras 2011), nous avons enregistré un taux très faible de liaisons non-enchaînées (« Non-enchaînement ») dans les productions des apprenants italophones. Ces formes de liaison semblent être presque uniquement présentes dans des types de liaison variable et tout particulièrement dans les contextes ADJ-NOM (ex. « petit appartement »), NOM-ADJ (ex. « des efforts inutiles ») et PRP-DET (ex. « dans un ») (cf. Tableau 64).

	Session								
	S1	**%**	**S2**	**%**	**S3**	**%**	**S4**	**%**	**Tota**
Total CAT	**1271**	**100**	**1321**	**100**	**1313**	**100**	**1275**	**100**	**5180**
Non-enchaînement	0	**0%**	1	**0%**	1	**0%**	1	**0%**	3
Total VAR	**1367**	**100**	**1418**	**100**	**1386**	**100**	**1389**	**100**	**5560**
Non-enchaînement	6	**0%**	9	**1%**	3	**0%**	6	**0%**	24

Tableau 64 : Résultats totaux - taux de non-enchaînement des liaisons catégoriques et variables (étude longitudinale)

1.2.2.3 Pauses-hésitations

En revanche, des formes de pause-hésitation (« Pause-hésitation ») sont bien plus fréquentes dans les productions des apprenants italophones et sembleraient d'ailleurs être majoritairement présentes au cours de la troisième session de l'étude longitudinale (S3) pour ensuite se réduire lors de la dernière session caractérisée par une augmentation sensible des liaisons enchaînées. Le tableau ci-dessous (cf. Tableau 65) montre la distribution de ces pauses-hésitations. Contrairement aux formes de non-enchaînement précédentes, celles-ci sont produites (bien qu'en faible mesure) non seulement dans les contextes de liaison variable mais aussi catégorique.

	Session								
	S1	**%**	**S2**	**%**	**S3**	**%**	**S4**	**%**	**Tota**
Total CAT	**127**	**100**	**1321**	**100**	**1313**	**100**	**1275**	**100**	**5180**
Pause-hésitation	13	**1%**	13	**1%**	19	**1%**	8	**1%**	53
Total VAR	**136**	**100**	**1418**	**100**	**1386**	**100**	**1389**	**100**	**5560**
Pause-hésitation	10	**1%**	6	**0%**	24	**2%**	12	**1%**	52

Tableau 65 : Résultats totaux - taux de pause-hésitation des liaisons catégoriques et variables (étude longitudinale)

1.2.2.4 Non-compréhensions

Quant aux formes de non-compréhension (NC) (cf. Tableau 66), elles relèvent uniquement des résultats de la tâche de répétition. Dans ce cas, les sujets étaient incapables de

reproduire les séquences input. La fréquence de ces cas est plus importante dans des types de liaison variable et se réduit progressivement au cours de l'étude longitudinale.

	Session								
	S1	**%**	**S2**	**%**	**S3**	**%**	**S4**	**%**	**Total**
Total CAT	**1271**	**100%**	**1321**	**100%**	**1313**	**100%**	**1275**	**100%**	**5180**
NC	17	**1%**	0	**0%**	0	**0%**	0	**0%**	17
Total VAR	**1367**	**100%**	**1418**	**100%**	**1386**	**100%**	**1389**	**100%**	**5560**
NC	67	**5%**	20	**1%**	29	**2%**	12	**1%**	128

Tableau 66 : Résultats totaux - taux de non-compréhensions des liaisons catégoriques et variables (étude longitudinale)

1.3. Nature de la CL réalisée

L'analyse de la réalisation de la liaison comporte bien évidemment une étude de la nature de la CL réalisée. Rappelons que celle-ci est décrite par le champ 6 du codage IPFC (cf. Partie 4.Chapitre 1.2.4.2.5) et correspond aux cas suivants : (a) CL conforme à la CL cible (dorénavant « Conforme ») ; (b) CL presque conforme à la cible mais avec des différences phonétiques à savoir correspondant à une prononciation orthographique de la CL (dorénavant « Orthographique ») ; (c) CL non conforme à la cible mais présente dans le mot (dorénavant « Présente dans M1 »); (d) CL non conforme à la CL et de nature épenthétique (à savoir des subtitutions des CL)[296] (cf. Partie 3.Chapitre 2.2.2) (dorénavant « Substitution »)[297]. Les tableaux résument les différentes réalisations de la CL enregistrées pendant l'étude longitudinale, calculées respectivement sur la base du total des liaisons réalisées par séance d'observation (cf. Tableau 67) et par typologie de liaison (cf. Tableau 68).

Nature CL	**S1**	**%**	**S2**	**%**	**S3**	**%**	**S4**	**%**	**Total**	**%**
Conforme	1350	**79%**	1539	**85%**	1571	**84%**	1648	**88%**	6108	**84%**
Orthographique	236	**14%**	202	**11%**	216	**12%**	169	**9%**	823	**11%**
Présente dans M1	24	**1%**	35	**2%**	44	**2%**	35	**2%**	138	**2%**
Substitution	7	**1%**	12	**1%**	12	**1%**	14	**1%**	45	**1%**
NC	84	**5%**	20	**1%**	29	**2%**	12	**1%**	145	**2%**
Total général	**1701**	**100%**	**1808**	**100%**	**1872**	**100%**	**1878**	**100%**	**7259**	**100%**

Tableau 67 : Résultats totaux (liaison catégorique et variable) - nature de la CL (étude longitudinale)

[296] Nous avons décidé de présenter séparemment les cas de substitution de la CL de ceux de réalisation d'une liaison inattendue (codés sous la même valeur dans le codage IPFC cf. Partie 4.Chapitre 1.2.4.2.5) afin d'isoler ces derniers du nombre total de liaisons potentielles (cf. Tableau 43).

[297] Nous avons ajouté à ces différentes typologies des réalisations de la CL la catégorie des « non-compréhensions » (dorénavant NC) afin de rendre compte des cas de non-compréhensions des séquences cibles enregistrées dans la tâche de répétition.

Nature CL	S1	%	S2	%	S3	%	S4	%
Total CAT	1072	**100%**	1137	**100%**	1175	**100%**	1143	**100%**
Conforme	900	**84%**	1001	**88%**	1041	**89%**	1046	**92%**
Orthographique	133	**12%**	103	**9%**	93	**8%**	59	**5%**
Présente dans M1	18	**2%**	28	**2%**	34	**3%**	31	**3%**
Substitution	4	**0%**	5	**0%**	7	**1%**	7	**1%**
NC	17	**2%**	0	**0%**	0	**0%**	0	**0%**
Total VAR	629	**100%**	671	**100%**	697	**100%**	735	**100%**
Conforme	450	**72%**	538	**80%**	530	**76%**	602	**82%**
Orthographique	103	**16%**	99	**15%**	123	**18%**	110	**15%**
Présente dans M1	6	**1%**	7	**1%**	10	**1%**	4	**1%**
Substitution	3	**0%**	7	**1%**	5	**1%**	7	**1%**
NC	67	**11%**	20	**3%**	29	**4%**	12	**2%**
Total général	**1701**	**100%**	**1808**	**100%**	**1872**	**100%**	**1878**	**100%**

Tableau 68 : Résultats totaux - nature de la CL par typologie de liaison (étude longitudinale)

1.3.1. Liaison conforme à la CL cible

Les résultats de l'analyse concernant la nature de la CL montrent que lorsque les apprenants italophones réalisent la liaison, ils produisent le plus souvent des CL conformes à la CL cible. Cette tendance est renforcée au cours des dernières séances d'enregistrement de l'étude longitudinale. En effet, le taux des réalisations de liaisons réalisées conformément à la CL augmente de 79% à 88% de la première (S1) à la dernière session d'observation (S4). De manière générale, on observe que les liaisons conformes à la CL cible sont plus fréquentes dans les contextes catégoriques que dans les contextes variables. Le taux de liaisons conformes à la CL cible augmente de 84% à 92% dans les contextes de liaison catégorique et de 72% à 82% dans les types de liaison variable.

1.3.2. CL presque conforme à la cible mais avec différence phonétique

Les réalisations non conformes à la CL correspondent à des productions orthographiques (cf. Partie 3.Chapitre 2.2.2). Celles-ci consistent plus précisément en des réalisations des CL /z/ et /t/ calquées sur les formes des consonnes graphiques finales présentes dans les M1 correspondants. Ces productions des CL apparaissent dans les cas où la production de la liaison demande un changement phonétique[298] à partir de la consonne présente dans l'orthographe. En effet, dans le cas des prononciations orthographiques, les consonnes graphiques < s > et < x >

[298] Nous rappelons que les consonnes graphiques < s > et < x > sont prononcées [z] ex. « les amis » [lezami] et « faux amis » [fozami] alors que la consonne graphique < d > donne lieu à la CL [t] ex. « grand ami » [grɑ̃tami] (cf. Partie 1.Chapitre 1.4).

sont prononcées [s] (ex. « les amis » [lesami] et « faux amis » [fosami]) et la consonne graphique < d > donne lieu à une prononcation en [d] de la CL (ex. « grand ami » [gʀɑ̃dami]).

Si nous analysons la distribution de ce type de réalisations, nous pouvons observer qu'elles sont plus importantes dans des séquences induisant une liaison variable (environ 15-16%) qu'une liaison catégorique (5%-12%). Toutefois, contrairement aux types de réalisations précédents, si ces réalisations orthographiques se réduisent progressivement dans les contextes catégoriques au cours de la période d'observation (12%-5%), elles restent relativement stables dans les contextes de liaison variable (16%-15%). Autrement dit, les apprenants produisent moins de prononciations orthographiques des liaisons catégoriques pendant les dernières séances d'enregistrement mais ils continuent à les réaliser dans des contextes de liaisons variables.

Contrairement aux études précédentes (Mastromonaco 1999 ; Howard 2005 ; Howard 2013 ; Thomas 2002 ; De Moras 2011) ayant enregistré une majorité de liaisons orthographiques de la CL /d/, nous avons observé à côté de celles-ci de nombreuses prononciations orthographiques de la CL /z/[299]. L'ensemble des occurrences intéressées par ce type de réalisations sont reparties dans le tableau ci-dessous (cf. Tableau 69) selon le type de liaison.

	Orth	**Total**	**%**
Total CAT	**388**	**4527**	**9%**
DET-NOM	**179**	**1947**	**9%**
Ces	17	141	**12%**
Des	105	987	**11%**
Les	42	521	**8%**
Leurs	2	13	**15%**
Mes	3	27	**11%**
Quelques	10	75	**13%**
EXF-EXF	**16**	**268**	**6%**
Plus	5	76	**7%**
Temps	11	100	**11%**
NUM-NOM	**15**	**446**	**3%**
Deux	10	106	**9%**
Dix	5	68	**7%**
PDE-NOM	**23**	**195**	**12%**
Des	23	195	**12%**
PRO-AUX	**103**	**670**	**15%**
Ils	97	291	**33%**
Nous	2	165	**1%**
Vous	4	63	**6%**
PRO-PRO	**10**	**167**	**6%**

[299] Cet aspect sera approfondi dans le prochain chapitre (cf. Partie 4.Chapitre 3.1.3.1).

Ils	7	39	**18%**
Nous	3	76	**4%**
PRO-VER	**33**	**517**	**6%**
Ils	19	115	**17%**
Nous	6	150	**4%**
Vous	8	135	**6%**
VER-PRO	**9**	**356**	**3%**
Allons	7	67	**10%**
Donnez	2	69	**3%**
Total VAR	**435**	**2732**	**16%**
ADJ-NOM	**158**	**856**	**18%**
Grand	158	434	**36%**
ADV-ADJ	**13**	**127**	**10%**
Plus	3	56	**5%**
Très	10	71	**14%**
ADV-ADV	**1**	**6**	**17%**
Pas	1	2	**50%**
AUX-PPA	**1**	**269**	**0%**
Sommes	1	3	**33%**
CON-PRO	**53**	**178**	**30%**
Quand	53	178	**30%**
NOM-ADJ	**69**	**190**	**36%**
Cocons	3	9	**33%**
Conseils	4	18	**22%**
Efforts	37	6	**16%**
Enfants	9	32	**28%**
Essais	12	51	**24%**
Intérêts	3	9	**33%**
Visites	1	2	**50%**
PRP-DET	**30**	**81**	**37%**
Dans	30	81	**37%**
PRP-NOM	**29**	**70**	**41%**
Dans	1	1	**100%**
Sans	28	64	**44%**
VER-ADJ	**1**	**66**	**2%**
Est	1	66	**2%**
VER-ADV	**78**	**325**	**24%**
Descend	50	61	**82%**
Sommes	27	68	**40%**
Vient	1	1	**100%**
VER-PRP	**2**	**83**	**2%**
Est	2	50	**4%**
Total général	**823**	**7259**	**11%**

Tableau 69 : Taux des réalisations orthographiques de la CL par type de liaison (étude longitudinale)

Dans le paragraphe 3.2, nous nous attacherons à analyser l'effet de tâche notamment de la lecture sur la réalisation de ces formes. Nous nous limitons pour l'instant à proposer une

analyse purement descriptive des occurrences les plus sujettes à ce type de réalisation. Parmi les contextes de liaison catégorique, les types DET-NOM et PRO-AUX présentent des taux relativement importants de réalisations orthographiques. Plus précisément, le déterminant « mes », « ces » et « quelques » ainsi que le pronom personnel sujet « ils » donnent souvent lieu respectivement à des réalisations telles que [mes] (11%), [ses] (12%), [kɛlk(ə)s] (13%) et [ils] (33%).

Nous pourrions supposer que les prononciations orthographiques de la CL /z/ du pronom personnel « ils » relèvent d'une confusion de l'apprenant entre les formes des verbes « être » et « avoir ». Toutefois, l'analyse des occurrences de ces liaisons orthographiques a montré que la presque totalité des formes orthographiques du pronom « ils » sont issues de séquences telles que « ils y ont ».

En ce qui concerne les contextes de liaison variable, un nombre important de types semble être intéressé par ce type de production. Parmi ceux-ci, nous pouvons signaler le type ADJ-NOM caractérisé par une présence massive (36%) de prononciations orthographiques de la CL de l'adjectif « grand ». Il en va de même pour le type NOM-ADJ dans lequel des M1 tels que « efforts » et « essais » sont souvent prononcées avec une CL orthographique.

1.3.3. CL non conforme à la CL présente dans le M1

Quant aux liaisons prononcées sur la base des CL non conformes à la cible mais présentes dans le M1, elles semblent ne représenter qu'une petite partie des réalisations totales (1%-2%) (cf. Tableau 67). De plus, l'analyse des occurrences des liaisons présentant ce type de productions de la CL montre qu'un groupe très restreint de M1 est sujet à ces réalisations. Ceux-ci correspondent le plus souvent à la prononciation des consonnes [t] (ex. « enfants ») et [p] (ex. « de temps en temps ») précédant la CL cible[300]. En revanche, compte tenu des difficultés liées à l'articulation des voyelles nasales (et non pas à la connaissance des règles de prononciation comme dans le cas précédent) pour les apprenants de FLE, nous avons choisi de traiter, contrairement à d'autres études précédentes (Racine sous presse ; Valetopoulos sous presse), les formes telles que « grand honneur » [gʁɑ̃nɔnoeʁ] comme des cas de non-réalisation de la liaison. La réalisation [gʁɑ̃n] représente sans doute le résultat d'une dénasalisation de la

[300] Nous pourrions également supposer que ces réalisations relèvent tout simplement de la prononciation erronée des consonnes fixes qui sont ensuite enchaînées au M2. Cependant nos données ne nous permettent pas d'étayer cette hypothèse. C'est la raison pour laquelle, dans le cadre des réflexions sur la transcription et le codage (cf. Partie 4.Chapitre 1.2.4.1.1) proposées par le projet IPFC, nous avons décidé de considérer ces formes comme des réalisations non conformes de la liaison, et d'assurer la cohérence de ce choix tout au long de l'analyse de notre corpus.

voyelle nasale[301] également observée devant des mots à initiale consonantique (ex. « grand dessin » [gʀɑ̃desɛ̃]).

Type de liaison	CL présente dans le M1	Total	%
Total CAT	**111**	**4527**	**2%**
EXF-EXF	**38**	**268**	**14%**
Temps	38	100	**38%**
NUM-NOM	**73**	**268**	**27%**
Cents	72	74	**97%**
Quatre-vingts	1	1	**100%**
Total VAR	**27**	**2738**	**1%**
NOM-ADJ	**26**	**190**	**14%**
Circuits	4	5	**80%**
Enfants	18	32	**56%**
Intérêts	5	9	**44%**
Total général	**138**	**7374**	**2%**

Tableau 70 : Taux des réalisations non conformes de la CL mais présentes dans le M1 par type de liaison (étude longitudinale)

1.3.4. CL non conforme avec substitution de la CL

Les exemples de substitutions de la CL cible sont rares (1%) dans notre corpus. Ces résultats rejoignent ceux des études précédentes qui décrivent ce type de réalisations comme caractéristique de l'acquisition L1. En effet, parmi les travaux menés en acquisition L2, les études de Mastromonaco (1999) et Harnois-Delpiano (2012) ont été les seules à relever la présence de ce type de réalisations. Bien que dans notre corpus, le taux de substitutions soit très faible, il semble toutefois augmenter au cours de l'étude longitudinale (cf. Tableau 71) et comme nous le verrons (cf. Paragraphe 3.2.3), presque uniquement en lecture (45 / 49 occurrences)[302].

Type de liaison	S1	S2	S3	S4	Total
Total CAT	**4**	**5**	**7**	**7**	**23**
DET-ADJ	**0**	**0**	**1**	**0**	**1**
Un autre	0	0	1	0	1
DET-NOM	**0**	**0**	**1**	**1**	**2**
Son alarme	0	0	1	0	1
Son interphone	0	0	0	1	1
EXF-EXF	**1**	**0**	**0**	**0**	**1**
Jeux olympiques	1	0	0	0	1
NUM-NOM	0	**1**	**1**	**1**	3

[301] Surtout dans le cas d'une langue comme l'italien où la correspondance entre phonie et graphie est très forte.

[302] Les 45 occurrences de l'étude longitudinale ont été enregistrées dans les tâches de lecture oralisée. En revanche, les 4 occurrences restantes relèvent de la parole spontanée issue de l'étude transversale.

Trente-huit ans	0	1	1	1	3
PDE-NOM	0	**1**	0	0	1
Des alentours	0	1	0	0	1
PRO-AUX	**2**	**1**	**2**	**1**	6
Nous avons	0	0	1	1	2
Tout a	0	0	1	0	1
Tout est	1	1	0	0	2
Vous êtes	1	0	0	0	1
PRO-VER	**0**	**1**	**1**	**4**	**6**
En installions	0	1	1	4	6
VER-PRO	**1**	**1**	**1**	**0**	**3**
Donnez-en	0	1	0	0	1
Prenez-en	1	0	0	0	1
Sont-elles	0	0	1	0	1
Total VAR	**3**	**7**	**5**	**7**	**22**
ADJ-NOM	**1**	**6**	**5**	**4**	**16**
Grand enclos	1	5	4	3	13
Petit enfant	0	1	1	1	3
AUX-ADV	**0**	**0**	**0**	**1**	**1**
Ont aussi	0	0	0	1	1
AUX-PPA	**1**	**0**	**0**	**1**	**2**
Ont acheté	1	0	0	1	2
PRP-NOM	**1**	**1**	**0**	**0**	**2**
Sans autorisation	1	1	0	0	2
VER-ADJ	**0**	**0**	**0**	**1**	**1**
Est impossible	0	0	0	1	1
Total général	**7**	**12**	**12**	**14**	**45**

Tableau 71 : Réalisations non conformes avec substitution de la CL par type (étude longitudinale)

Les consonnes utilisées dans ces substitutions de la CL cible sont [d], [g], [k], [l], [n], [t] et [z] (cf. Tableau 72). La consonne [z] semble être la plus souvent employée (dans 71% des cas). On retrouve des substitutions en [z] dans les contextes catégoriques tels que « trente-huit [z] ans », « tout [z] est fait», « nous en [z] installions », mais également dans les contextes variables comme « grand [z] enclos », « petit [z] enfant », « ils ont [z] aussi », « ils ont [z] acheté » ou « il est [z] impossible ».

Type de liaison	[d]	[g]	[l]	[n]	[t]	[z]	Total
Total CAT	**0**	**0**	**1**	**4**	**3**	**12**	**20**
DET-ADJ	**0**	**0**	**0**	**0**	**1**	**0**	**1**
Un autre	0	0	0	0	1	0	1
DET-NOM	**0**	**0**	**0**	**0**	**2**	**0**	**2**
Son alarme	0	0	0	0	1	0	1
Son interphone	0	0	0	0	1	0	1
EXF-EXF	**0**	**0**	**0**	**1**	**0**	**0**	**1**
Jeux olympiques	0	0	0	1	0	0	1
NUM-NOM	**0**	**0**	**0**	**0**	**0**	**3**	**3**
Trente-huit ans	0	0	0	0	0	3	3
PDE-NOM	**0**	**0**	**1**	**0**	**0**	**0**	**1**
Des alentours	0	0	1	0	0	0	1
PRO-AUX	**0**	**0**	**0**	**3**	**0**	**3**	**6**
Nous avons	0	0	0	2	0	0	2
Tout a	0	0	0	0	0	1	1
Tout est	0	0	0	0	0	2	2
Vous êtes	0	0	0	1	0	0	1
PRO-VER	**0**	**0**	**0**	**0**	**0**	**6**	**6**
En installions	0	0	0	0	0	6	6
Total VAR	**1**	**1**	**0**	**1**	**2**	**20**	**25**
ADJ-NOM	**0**	**0**	**0**	**0**	**0**	**16**	**16**
Grand enclos	0	0	0	0	0	13	13
Petit enfant	0	0	0	0	0	3	3
AUX-ADV	**0**	**0**	**0**	**0**	**0**	**1**	**1**
Ont aussi	0	0	0	0	0	1	1
AUX-PPA	**1**	**0**	**0**	**0**	**0**	**1**	**2**
Ont acheté	1	0	0	0	0	1	2
PRP-NOM	**0**	**0**	**0**	**0**	**2**	**0**	**2**
Sans autorisation	0	0	0	0	2	0	2
VER-ADJ	**0**	**0**	**0**	**0**	**0**	**1**	**1**
Est impossible	0	0	0	0	0	1	1
VER-PRO	**0**	**1**	**0**	**1**	**0**	**1**	**3**
Donnez-en	0	1	0	0	0	0	1
Prenez-en	0	0	0	1	0	0	1
Sont-elles	0	0	0	0	0	1	1
Total général	**1**	**1**	**1**	**5**	**5**	**32**	**45**

Tableau 72 : Réalisations non conformes avec substitution de la CL par type de consonne (étude longitudinale)

De façon générale, nous n'avons pas enregistré de différences entre le taux de production des substitutions en contexte de liaison catégorique et variable. Néanmoins, le type de liaison ADJ-NOM rend compte de 33% des substitutions totales. Celles-ci comportent, dans la presque totalité des cas, des remplacements de la CL cible par la consonne [z].

1.4. Liaisons erratiques

De nombreux cas de réalisations des liaisons erratiques devant « h » aspiré ont été produits dans la séquence « plus // haut » et « au fur et // à mesure », et plus généralement après la conjonction « et » (cf. Tableau 73). Leur production assez importante reste relativement stable dans les quatre séances d'enregistrement et le plus souvent attestée en lecture.

	Sessions				
	S1	**S2**	**S3**	**S4**	**Total**
Enchaînement	**21**	**26**	**18**	**22**	**87**
ADV-ADJ	**11**	**15**	**12**	**13**	**51**
Plus // haut	11	15	12	13	51
ADV-DET	**1**	**1**	**0**	**1**	**3**
Vraiment // une	1	1	0	1	3
ADV-PRP	**1**	**0**	**0**	**0**	**1**
Comment // en	1	0	0	0	1
CON-VER	**3**	**2**	**2**	**2**	**9**
Et // VER	3	2	2	2	9
EXF-EXF	**1**	**7**	**2**	**1**	**11**
Au *fur et // à mesure*	1	7	2	1	11
NAM-PRP	**2**	**1**	**0**	**0**	**3**
Berlin // en	2	1	0	0	3
NOM-PRO	**1**	**0**	**0**	**3**	**4**
Opposition // aurait	1	0	0	3	4
NOM-PRO	**0**	**0**	**1**	**1**	**2**
Dans le coin //on	0	0	1	1	2
PRP-VER	**1**	**0**	**0**	**0**	**1**
Et // VER	1	0	0	0	1
Orth	**1**	**0**	**1**	**1**	**3**
ADV-ADJ	**1**	**0**	**1**	**1**	**3**
Plus // haut	1	0	1	1	3
Total général	**22**	**26**	**19**	**23**	**90**

Tableau 73 : Liaisons erratiques par type et token de liaison (étude longitudinale)

1.5. Liaisons non attendues

Nous avons également enregistré quelques exemples de liaison non attendue après un mot à finale vocalique (ex. « va [t] à Paris »). À notre connaissance, ce type de réalisation, n'a jamais été attesté par les études précédentes menées en FLE (sauf dans le cas de l'étude de Falbo et al. sous presse). Ces formes sont très peu fréquentes dans notre corpus mais, comme dans le cas des substitutions, elles sembleraient apparaître moins souvent au cours de la première session pour ensuite augmenter au cours des sessions suivantes de l'étude longitudinale. De plus, plusieurs cas de liaisons non attendues ont été observées dans des contextes de liaison catégorique tels que DET-NOM, NUM-NOM, PRO-AUX et PRO-VER (cf. Tableau 74).

Type de liaison	Sessions				Total général
	S1	S2	S3	S4	
Total CAT	**2**	**1**	**3**	**4**	**10**
DET-NOM	0	0	1	1	2
NUM-NOM	0	0	1	0	1
NUM-PDE	0	1	0	0	1
PRO-AUX	1	0	1	2	4
PRO-VER	1	0	0	1	2
Total ERR	**0**	**3**	**3**	**1**	**7**
DET-NOM	0	3	3	1	7
Total VAR	**2**	**5**	**1**	**2**	**9**
ADJ-NOM	1	0	0	0	1
INF-NOM	1	0	0	0	1
PRO-INF	0	2	1	2	5
PDE-NOM	0	1	0	0	1
VER-DET	0	2	0	0	2
Total général	**4**	**9**	**7**	**7**	**27**

Tableau 74 : Liaisons non attendues par type de liaison (étude longitudinale)

Parmi l'ensemble des formes de liaison non attendue, nous pouvons distinguer entre celles qui apparaissent devant un mot à initiale consonantique (ex. « quelques fanatiques ») et celles qui apparaissent après un mot se terminant par voyelle (ex. « qui ont »). Les premières, dont nous n'avons enregistré que trois occurrences lors de la première et de la deuxième séance, relèvent sans doute d'une généralisation morpho-phonologique de la liaison devant des mots pluriels à initiale consonantique. Le deuxième type de liaisons non attendue est en revanche majoritaire et est le résultat de régularisations et de généralisations opérées par les apprenants. C'est le cas par exemple de réalisations telles que « le [z] halls » au lieu de « le hall », « cinq [z] ans », « quatrième [z] aux » ou encore « qui [z] ont » et « y [z] entrer » (cf. Tableau 75).

Type de liaison	Total général
Total CAT	**10**
DET-NOM	**2**
D'électeurs	1
Quelques fanatiques	1
NUM-NOM	**2**
Cinq ans	1
NUM-PDE	**1**
Quatrième aux	1
PRO-AUX	**4**
Qui ont	1
Y ont	3
PRO-VER	**2**

Elle est	1
Ils sont	1
Total ERR	**7**
DET-NOM	**7**
Le hall	5
Le hasard	2
Total VAR	**9**
ADJ-NOM	**1**
Joli interphone	1
INF-NOM	**1**
Faire étape	1
PRO-INF	**5**
Y entrer	5
PDE-NOM	**1**
Aux jeux	1
VER-DET	**2**
Eu aucun	2
Total général	**27**

Tableau 75 : Liaisons non attendues par type et token de liaison (étude longitudinale)

1.6. Résultats du test de jugement

Les résultats totaux issus du test de jugement montrent de façon générale l'absence d'une évolution significative du choix des réponses correctes (correspondant à la catégorie *good* cf. Partie 4.Chapitre 1.4.3.1.6) aussi bien dans les cas des liaisons catégoriques (cf. Figure 32) que variables (cf. Figure 33). Autrement dit, l'apprenant ne parvient pas à sélectionner la variante correspondant à la CL cible[303] dans le cas des liaisons catégoriques (ex. [ilzabimdevwatyʀ] pour l'énoncé « ils abiment des voitures ») ou les deux réponses sans et avec CL correspondant à la CL cible dans le cas des liaisons variables (ex. [ləgʀɑ̃tɑ̃klo] et [ləgʀɑ̃ɑ̃klo] « le grand enclos »)[304].

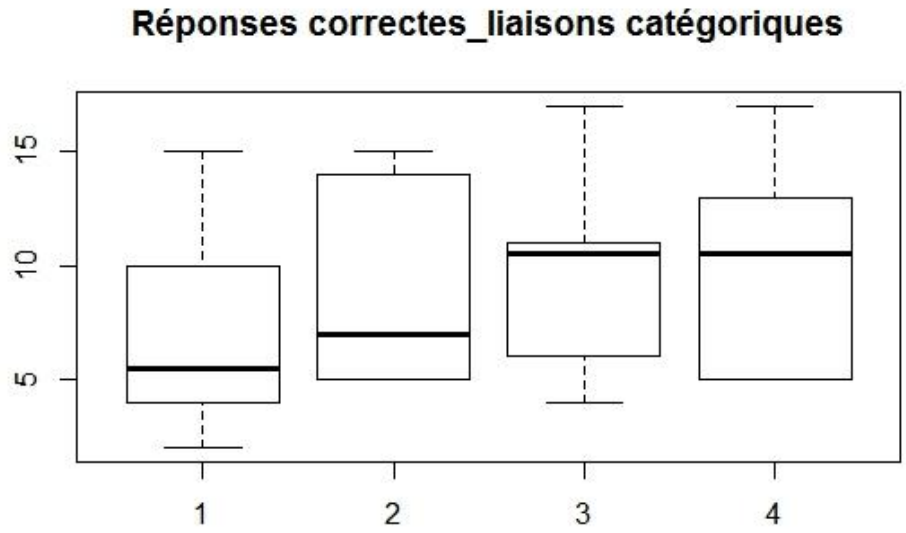


Figure 32 : Comparaison longitudinale des taux de réponses correctes des liaisons catégoriques issus du test de jugement

Les jugements corrects concernant les liaisons variables semblent en outre se réduire au cours de la troisième séance de l'étude longitudinale (x = 3 cf. Figure 33) pour ensuite augmenter lors de la dernière séance (x = 4 cf.Figure 33). Comme nous le verrons dans les pages qui suivent, à cette variation correspond une augmentation de réponses incomplètes (*missing* « manqué » cf. Partie 4.Chapitre 1.4.3.1.6).

[303] Les hypothèses formulées pour expliquer cette défaillance seront présentées dans le prochain chapitre (cf. Partie 4.Chapitre 3.1).

[304] Il convient de rappeler que pendant le test de jugement, le sujet avait la possibilité de choisir entre une ou deux réponses correctes parmi les trois proposées. Celles-ci correspondaient respectivement à : (a) une prononciation orthographique [ilsabimdevwatyʀ] ou avec une CL différente de la cible [dizɛl] au lieu de [ditɛl]; (b) une réalisation avec CL conforme à la cible [ilzabimdevwatyʀ] ; (c) une réalisation sans liaison [ilabimdevwatyʀ].

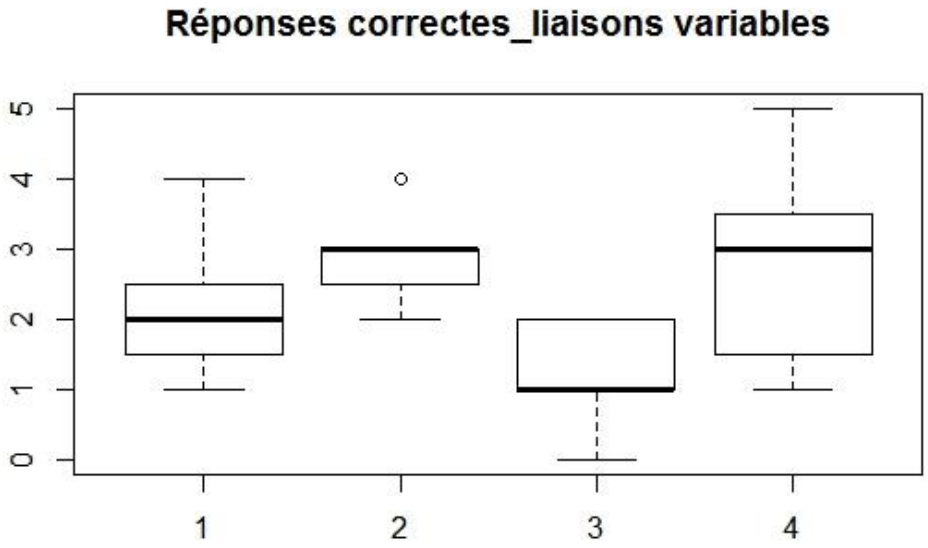


Figure 33 : Comparaison longitudinale des taux de réponses correctes des liaisons variables issus du test de jugement

La plupart des réponses correspondent en effet au choix de :

1) la seule variante orthographique tant en contexte de liaison catégorique (correspondant à la catégorie *error* « erreur ») (ex. « ils abiment des voitures» réalisé [ilsabimdevwatyʀ] au lieu de [ilzabimdevwatyʀ]) (1bis) que variable (ex. « le grand enclos » réalisé [ləgʀɑ̃dɑ̃klo] au lieu de [ləgʀɑ̃tɑ̃klo]) (correspondant à la catégorie *error* « erreur») ;
2) la variante correspondant à la CL cible à côté de la variante orthographique en contexte de liaison catégorique ([ilzabimdevwatyʀ] et [ilsabimdevwatyʀ]) (correspondant à la catégorie *invalid plus* « nul plus ») et (2bis) également en contexte de liaison variable ([ləgʀɑ̃tɑ̃klo] et [ləgʀɑ̃dɑ̃klo]) (correspondant à la catégorie *invalid* « nul ») ;
3) la variante orthographique à côté de la variante avec non-réalisation en contexte variable ([ləgʀɑ̃dɑ̃klo] et [ləgʀɑ̃ɑ̃klo]) (correspondant à la catégorie *invalid* « nul ») ;
4) la seule variante avec non-réalisation en contexte de liaison variable ([ləgʀɑ̃ɑ̃klo]) (correspondant à la catégorie *missing* « manqué ») ;
5) la seule variante avec CL correspondant à la cible en contexte de liaison variable et ([ləgʀɑ̃tɑ̃klo]) (correspondant à la catégorie *missing* « manqué ») ;

6) deux seuls cas de choix de la variante orthographique à côté de la variante avec non-réalisation en contexte de liaison catégorique ont été enregistrés (ex. [ilsabimdevwatyʀ] et [ilabimdevwatyʀ]) (correspondant à la catégorie *error plus* « erreur plus »).

Dans les prochaines pages, nous nous attacherons à présenter les taux de réponses pour chacune des catégories décrites ci-dessus tout au long de la période d'observation. Les résultats seront répartis par typologie de liaison : catégorique (cf. Paragraphe 1.6.1) et variable (cf. Paragraphe 1.6.2).

1.6.1. Jugements liaisons catégoriques

Les données issues des réponses totales des liaisons catégoriques montrent clairement une augmentation des réponses correctes (41% - 60%) (cf. Tableau 76) qui n'est cependant pas significative (Paragraphe cf. 1.6). Le taux des réponses *error* (correspondant au choix de la seule variante orthographique) (réponse 1) est relativement stable et affecte particulièrement la liaison après l'adjectif démonstratif « ces » (variante sélectionnée [ses]) (29% - 12%) et la liaison dans l'expression figée « de temps en temps » (variante sélectionnée [dətɑ̃sɑ̃tɑ̃] (35% - 41%).

Total CAT	**S1**	**%**	**S2**	**%**	**S3**	**%**	**S4**	**%**
good	42/102	**41%**	53/102	**52%**	59/102	**58%**	61/102	**60%**
error	16/102	**16%**	17/102	**17%**	18/102	**18%**	12/102	**12%**
error plus	1/102	**1%**	1/102	**1%**	0/102	**0%**	0/102	**0%**
invalid plus	43/102	**42%**	31/102	**30%**	25/102	**25%**	29/102	**28%**

Tableau 76 : Comparaison longitudinale des taux de réponses totales des liaisons catégoriques issus du test de jugement

En revanche, la réponse correspondant à la CL cible est sélectionnée le plus souvent dans les cas de la liaison après le pronom personnel sujet « ils » (59% - 76%), après les adjectifs numéraux « deux » (29% - 53%) et « dix » (24% - 71%) et devant le pronom clitique postposé « elle » (88% - 100%).

Quant aux choix des réponses *invalid plus*, la presque totalité (91%) correspond au choix de la variante avec CL conforme à la cible à côté de la variante orthographique (réponse 2) (cf. Paragraphe 1.6) ([ilzabimdevwatyʀ] et [ilsabimdevwatyʀ]). 8% seulement des réponses (la plupart attestés dans les premières séances) correspondent au choix de la variante avec CL conforme à la cible à côté de la variante avec non-réalisation (réponse 6) ([ilzabimdevwatyʀ] et [ilabimdevwatyʀ]) (cf. Paragraphe 1.6).

Comme dans le cas précédent, le taux de réponses *invalid plus* est plus important et en plus relativement stable dans les contextes de liaison après l'adjectif démonstratif « ces » (variante sélectionnée [ses]) (59% - 59%) et dans le cas de la liaison dans l'expression figée « de temps en temps » (variante sélectionnée [dətɑ̃sɑ̃tɑ̃] (29% - 29%) qu'après le pronom personnel sujet « ils » (35% - 12%). De plus, contrairement aux deux premiers contextes (à savoir « ces » et « de temps en temps ») pour lesquels le taux de réponses *invalid plus* est relativement stable, nous pouvons constater dans le contexte de liaison après « ils » une réduction progressive de ce type de réponse en faveur des réponses *good* correspondant au choix de la seule variante correcte (ex. [ilzabimdevwatyʀ]).

Ces résultats, combinés à l'absence presque totale (deux seules occurrences) de réponses *error plus* (réponse 6) impliquant le choix de la variante orthographique à côté de la variante avec non-réalisation (7) ([ilsabimdevwatyʀ] et [ilabimdevwatyʀ]) (cf. Paragraphe 1.6), indiquent que l'apprenant est conscient du caractère catégorique de ces contextes de liaison mais qu'il choisira cependant deux différentes variantes de réalisation, l'une avec CL conforme à la cible [ilzabimdevwatyʀ] et l'autre orthographique [ilsabimdevwatyʀ].

1.6.2. Jugements liaisons variables

En ce qui concerne les liaisons variables, aucune évolution significative n'a été enregistrée dans le taux de choix des variantes conformes à la CL cible (correspondant à la catégorie *good*). De plus, une réduction nette de ces réponses est observée lors de la troisième séance de la période d'observation accompagnée d'une augmentation de réponses *missing* (réponses 4 et 5) (cf. Tableau 77).

Total VAR	**S1**	**%**	**S2**	**%**	**S3**	**%**	**S4**	**%**
good	15/119	**13%**	20/119	**17%**	9/119	**8%**	19/119	**16%**
error	16/119	**13%**	15/119	**13%**	14/119	**12%**	12/119	**10%**
invalid	17/119	**14%**	4/119	**3%**	10/119	**8%**	9/119	**8%**
missing	71/119	**60%**	80/119	**67%**	86/119	**72%**	79/119	**66%**

Tableau 77 : Comparaison longitudinale des taux de réponses totales des liaisons variables issus du test de jugement

Le taux des réponses du type *error* (correspondant au choix de la seule variante orthographique) (réponse 1bis) (cf. Paragraphe 1.6) varie faiblement au cours de l'étude longitudinale. Ce résultat indique que les apprenants non seulement choisissent la variante orthographique (ex. [ləgʀɑ̃dɑ̃klo]) mais qu'ils la considèrent comme la seule réponse correcte. Autrement dit, ils attribuent à de la liaison variable un caractère obligatoire en choisissant une seule variante à savoir celle orthographique. Cet aspect est encore plus intéressant si nous

considérons que ce type de réponse a représenté la principale sélection pour les liaisons variables les plus fréquentes telles que la liaison après l'adjectif épithète « grand » (24% - 24%) et la conjonction de subordination « quand » (35% -18%).

Les résultats sont plus variables dans le cas des réponses *invalid* pour lesquelles deux tendances ont été observées : d'une part les apprenants choisissent la variante correspondant à la CL cible à côté de la variante orthographique ([ləgʀɑ̃tɑ̃klo] et [ləgʀɑ̃dɑ̃klo]) (réponse 2bis) (cf. Paragraphe 1.6) ; de l'autre, les apprenants choisissent la variante orthographique ([ləgʀɑ̃dɑ̃klo] à côté de la variante avec non-réalisation [ləgʀɑ̃ɑ̃klo]) (réponse 3) (cf. Paragraphe 1.6).

Si dans les deux cas l'apprenant sélectionne une variante orthographique, ce n'est que dans le deuxième (réponse 3) (cf. Paragraphe 1.6) qu'il traite la liaison en tant que variable. Les données ne permettent pas d'établir laquelle de ces deux tendances est majoritaire. Il semblerait que dans ces cas l'apprenant hésite quant au caractère variable des contextes de liaison tout en sélectionnant une variante orthographique.

Quant aux réponses de type *missing*, elles sont sans aucun doute les plus représentées (60% - 66%). Autrement dit, les apprenants semblent sélectionner le plus souvent la seule variante correspondant à la CL cible ([ləgʀɑ̃tɑ̃klo]) (réponse 5) (cf. Paragraphe 1.6), en particulier dans les contextes de liaison les plus fréquents tels que la liaison après l'adjectif épithète « grand » (86% - 100%), la conjonction de subordination « quand » (71% - 100%) et entre le verbe « avoir » et l'adverbe « aussi » (« ils ont aussi ») (71% - 100%).

Cette tendance est en revanche moins marquée pour les contextes de liaison variable moins fréquente (NOM-ADJ, VER-ADJ et VER-PRP) (cf. Tableau 77) pour lesquels le choix de la seule variante avec non-réalisation est privilégié (réponse 4) (ex. [ɛ.ɛ̃pʀənabl] dans la séquence « la vue est imprenable sur le hall d'entrée»).

1.7. Bilan de l'étude longitudinale

En résumé, l'analyse des résultats généraux de cette étude longitudinale nous a permis de mettre en évidence les points suivants :

- de manière générale, les productions des apprenants conformes à la CL cible ne semblent pas évoluer de manière significative tout au long de l'étude longitudinale ;

Pour ce qui est du type de réalisation:

- la production de la liaison après les articles définis et indéfinis et après les pronoms personnels sujets n'est pas sensible à des effets de fréquence de cooccurrence et de

tâche. Ces observations rejoignent les résultats de l'étude menée par Tennant (sous presse) et Falbo (sous presse) qui montrent que la liaison après déterminants et après pronoms personnels sujets est maîtrisée dans toutes les tâches aussi bien par les apprenants les plus avancés (Tennant sous presse ; Falbo et al. sous presse corpus Trieste) que par les moins avancés (Tennant sous presse) et débutants (Falbo et al. sous presse corpus Milan) ;

- la liaison avant et après les pronoms clitiques fait état de faibles taux de réalisation ;
- le taux de non-réalisation des liaisons variables chez les apprenants n'est pas significativement différent de celui des natifs ;
- la plupart des liaisons sont produites avec enchaînement, mais il existe néanmoins des exemples de liaisons non-enchaînées ;
- les apprenants italophones produisent un taux très faible de liaisons non-enchaînées et presque uniquement dans des types de liaison variable ;
- les exemples de CL précédées par une pause-hésitation, contrairement aux formes de non enchaînement, sont produites non seulement dans les contextes de liaison variable mais également catégorique ;

Pour ce qui est de la nature de la CL réalisée :

- lorsque les apprenants italophones réalisent la liaison, ils produisent le plus souvent des CL conformes à la consonne cible ;
- la plupart des réalisations non conformes à la CL relèvent des productions orthographiques (cf. Partie 3.Chapitre 2.1.4) ;
- les apprenants produisent moins de prononciations orthographiques des liaisons catégoriques lors des dernières séances d'enregistrement, mais ils continuent à les réaliser dans des contextes de liaison variable ;
- de même, dans le test de jugement, le nombre de réponses contenant une variante orthographique dans les contextes catégoriques (*invalid*) se réduit progressivement au cours de l'étude longitudinale. Cette tendance semble également attestée indirectement dans les contextes variables (par une sélection progressivement majoritaire des réponses correspondant au choix de la seule variante avec CL cible) (*missing*) ;
- le nombre de substitutions semble augmenter au cours de l'étude longitudinale ;
- le taux de liaisons inattendues semble augmenter au cours des dernières séances de l'étude longitudinale, en particulier dans les contextes catégoriques.

Ces résultats partiaux nécessitent bien évidement d'analyses supplémentaires, ce qui nous amène à étudier les données issues des analyses transversales pour ensuite effectuer une comparaison entre les réalisations extraites des trois typologies de productions orales recueillies dans notre corpus (répétition, lecture et parole semi-spontanée).

2. Résultats de l'étude transversale

L'étude transversale, rappelons-le, comprend uniquement deux types de tâches : la dénomination d'image (cf. Partie 4.Chapitre 1.4.3.2.1) soumise au début de la troisième session d'enregistrement (S3) et la tâche de reformulation soumise au cours de la dernière séance d'enregistrement (S4).

2.1. Types de liaison potentielle

Le tableau ci-dessous (cf. Tableau 78) résume l'ensemble des types de liaison potentielle repartis par classe de liaison (catégorique et variable) observés dans les deux tâches de l'étude transversale. Ces dernières sont caractérisées par un nombre plus élevé de sites de liaison variable pour lesquels il nous a paru nécessaire de recueillir au cours de notre travail un échantillon plus important de données issues de plusieurs typologies de production orale. Parmi les types de liaison variable les plus attestés, nous pouvons citer la liaison entre l'adjectif épithète et le nom suivant (ADJ-NOM) et entre l'adverbe (monosyllabique et polysyllabique) et l'adjectif suivant (ADV-ADJ).

	Tâches		
Types de liaison	**Dénomination d'images**	**Reformulation**	**Total général**
Total CAT	**0**	**34**	**34**
DET-NOM	0	34	34
Total VAR	**169**	**252**	**421**
ADJ-NOM	169	17	186
ADV-ADJ	0	83	83
ADV-PPA	0	17	17
PRP-DET	0	68	68
PRP-NOM	0	16	16
VER-ADJ	0	51	51
Total général	**169**	**286**	**455**

Tableau 78 : Types de liaison potentielle catégorique et variable (étude transversale)

2.1.1. Dénomination d'image

2.1.1.1 Réalisation de la liaison

Le type représenté dans cette tâche est uniquement le contexte de liaison ADJ-NOM pour lequel nous avons enregistré un taux total de réalisations enchaînées de 60% et un faible pourcentage de pauses-hésitations. Les seuls adjectifs analysés sont « grand », « petit » et « gros » (ce dernier utilsé à la place de « grand »). Ce dernier ne présente que 4 occurrences (toutes correspondant à des non-réalisations), utilisées par les apprenants comme variantes de l'adjectif « grand » qui constitait la cible de cette tâche. De ces premiers résultats, la liaison semblerait le plus souvent réalisée après l'adjectif « grand » qu'après l'adjectif « petit » (cf. Tableau 79).

Type de liaison	Non-réalisation	%	Réalisation	%	Pause-hésitation	%	Total général
Grand	32	41%	44	56%	2	3%	78
Gros	8	100%	0	0%	0	0%	8
Petit	24	29%	57	69%	2	2%	83
Total général	**64**	**38%**	**101**	**60%**	**4**	**2%**	**169**

Tableau 79 : Résultats totaux - réalisations de la liaison (tâche de dénomination d'image)

2.1.1.2 Nature de la CL réalisée

Si la liaison semble être aussi bien fréquente après l'adjectif « grand » qu'après « petit », de nombreuses réalisations (63%) correspondent toutefois à des formes orthographiques (cf. Tableau 80). Celles-ci ont déjà été observées dans l'étude longitudinale et notamment dans la tâche de lecture des séquences. En effet, pendant la tâche de dénomination d'image, nous avons soumis les mêmes occurrences de liaison qui avaient déjà été présentées dans la tâche de lecture des séquences isolées de l'étude longitudinale. Cette démarche nous a permis de mener une analyse comparative des productions réalisées en lecture et en production semi-spontanée.

Types de liaison	Conforme	%	Orth	%	Total général
Grand	17	**37%**	29	**63%**	46
Petit	59	**100%**	0	**0%**	59
Total général	**76**	**72%**	**29**	**28%**	**105**

Tableau 80 : Taux de réalisation orthographique de la CL par type de liaison (tâche de dénomination d'image)

Les résultats de cette comparaison sont résumés dans le tableau ci-dessous (cf. Tableau 81). Cette étude comparative montre que les réalisations orthographiques de l'adjectif « grand» sont attestées aussi bien en lecture qu'en parole semi-spontanée.

	Lecture séquences isolées de liaison (étude longitudinale)					Dénomination d'images (étude transversale)				
	Conf	**%**	**Orth**	**%**	**Total**	**Conf**	**%**	**Orth**	**%**	**Total**
Grand	74	**56%**	57	**44%**	131	17	**37%**	29	**63%**	46
Petit	122	**100%**	0	**0%**	122	59	**100%**	0	**0%**	59
Total général	**196**	**77%**	**57**	**23%**	**253**	**76**	**72%**	**29**	**28%**	**105**

Tableau 81 : Comparaison entre les taux de réalisation orthographique de la CL issus de la lecture des séqueces isolées de liaison (étude longitudinale) et de la tâche de dénomination d'image (étude transversale)

2.1.2. Tâche de reformulation

Rappelons que dans cette tâche les sujets étaient séparés en deux groupes selon le type d'input écrit ou oral à travers lequel les séquences sources étaient présentées.

2.1.2.1 Réalisation de la liaison

Dans ce type de parole semi-spontanée, nous avons enregistré d'une part une presque totalité (97%) de réalisations dans des contextes catégoriques et de l'autre une forte idiosyncrasie interne de certains types de liaison variable comparable à celle déjà observée chez les natifs (cf. Partie 1.Chapitre 4.4).

En effet, si nous analysons les taux de réalisation au sein du type ADV-ADJ (cf. Tableau 82), nous pouvons remarquer que, comme dans les cas des données issues du corpus des natifs (cf. Partie 2.Chapitre 2.2.2.2.5), la liaison est davantage réalisée après les adverbes monosyllabiques qu'après les polysyllabiques. De plus, les apprenants produisent plus de liaisons après les adverbes monosyllabiques « très » et « plus » qu'après les adverbes « moins » et « trop ». Nous pouvons également souligner un taux plus important de réalisations après la forme « est » du verbe être qu'après la forme verbale « suis » au sein du type VER-DET.

Types	Non-réalisation	%	Réalisation	%	Total général
Total CAT	**1**	**3%**	**33**	**97%**	**34**
DET-NOM	**1**	**3%**	**33**	**97%**	**34**
Cet	1	**6%**	16	**94%**	17
Des	0	**0%**	17	**100%**	17
Total VAR	**125**	**50%**	**127**	**50%**	**252**
ADJ-NOM	**3**	**18%**	**14**	**82%**	**17**
Ancien	3	**18%**	14	**82%**	17
ADV-ADJ	**43**	**52%**	**40**	**48%**	**83**
Moins	13	**81%**	3	**19%**	16
Plus	3	**18%**	14	**82%**	17
Toujours	12	**71%**	5	**29%**	17
Très	3	**18%**	14	**82%**	17

Trop	12	**75%**	4	**25%**	16
ADV-PPA	**3**	**18%**	**14**	**82%**	**17**
Bien	3	**18%**	14	**82%**	17
PRP-DET	**46**	**68%**	**22**	**32%**	**68**
Devant	24	**71%**	10	**29%**	34
Sous	22	**65%**	12	**35%**	34
PRP-NOM	**6**	**38%**	**10**	**63%**	**16**
Sans	6	**38%**	10	**63%**	16
VER-ADJ	**24**	**47%**	**27**	**53%**	**51**
C'est	3	**18%**	14	**82%**	17
Est	14	**82%**	3	**18%**	17
Suis	7	**41%**	10	**59%**	17
Total général	**126**	**44%**	**160**	**56%**	**286**

Tableau 82 : Résultats totaux - taux de réalisation par type de liaison (tâche reformulation)

En ce qui concerne l'influence de la variable input sur la réalisation de liaison, nous n'avons pas enregistré de différences importantes au niveau des réalisations entre les deux groupes de sujets. Les apprenants exposés à un input écrit (cf. Tableau 84) ne réalisent pas plus de liaison que les apprenants n'ayant été soumis qu'à un input oral (cf. Tableau 83). Il convient cependant de souligner que le nombre exigu de locuteurs ne nous permet pas d'émettre des généralisations à ce sujet.

Groupe oral					
Type liaison	**Non-réalisation**	**%**	**Réalisation**	**%**	**Total**
Total CAT	**0**	**0%**	**18**	**100%**	**18**
DET-NOM	**0**	**0%**	**18**	**100%**	**18**
Cet	0	**0%**	9	**100%**	9
Des	0	**0%**	9	**100%**	9
Total VAR	**79**	**59%**	**55**	**41%**	**134**
ADJ-NOM	**2**	**22%**	**7**	**78%**	**9**
Ancien	2	**22%**	7	**78%**	9
ADV-ADJ	**29**	**66%**	**15**	**34%**	**44**
Moins	9	**100%**	0	**0%**	9
Plus	3	**33%**	6	**67%**	9
Toujours	7	**78%**	2	**22%**	9
Très	3	**33%**	6	**67%**	9
Trop	7	**88%**	1	**13%**	8
ADV-PPA	**2**	**22%**	**7**	**78%**	**9**
Bien	2	**22%**	7	**78%**	9
PRP-DET	**25**	**69%**	**11**	**31%**	**36**
Devant	14	**78%**	4	**22%**	18
Sous	11	**61%**	7	**39%**	18
PRP-NOM	**5**	**56%**	**4**	**44%**	**9**
Sans	5	**56%**	4	**44%**	9

VER-ADJ	**16**	**59%**	**11**	**41%**	**27**
C'est	3	**33%**	6	**67%**	9
Est	8	**89%**	1	**11%**	9
Suis	5	**56%**	4	**44%**	9
Total général	**79**	**52%**	**73**	**48%**	**152**

Tableau 83 : Groupe oral - taux de réalisation par type de liaison (tâche reformulation)

Groupe écrit					
Type liaison	**Non-réalisation**	**%**	**Réalisation**	**%**	**Total**
Total CAT	**1**	**6%**	**15**	**94%**	**16**
DET-NOM	**1**	**6%**	**15**	**94%**	**16**
Cet	1	13%	7	88%	8
Des	0	0%	8	100%	8
Total VAR	**46**	**39%**	**72**	**61%**	**118**
ADJ-NOM	**1**	**13%**	**7**	**88%**	**8**
Ancien	1	13%	7	88%	8
ADV-ADJ	**14**	**36%**	**25**	**64%**	**39**
Moins	4	57%	3	43%	7
Plus	0	0%	8	100%	8
Toujours	5	63%	3	38%	8
Très	0	0%	8	100%	8
Trop	5	63%	3	38%	8
ADV-PPA	**1**	**13%**	**7**	**88%**	**8**
Bien	1	13%	7	88%	8
PRP-DET	**21**	**66%**	**11**	**34%**	**32**
Devant	10	63%	6	38%	16
Sous	11	69%	5	31%	16
PRP-NOM	**1**	**14%**	**6**	**86%**	**7**
Sans	1	14%	6	86%	7
VER-ADJ	**8**	**33%**	**16**	**67%**	**24**
C'est	0	0%	8	100%	8
Est	6	75%	2	25%	8
Suis	2	25%	6	75%	8
Total général	**47**	**35%**	**87**	**65%**	**134**

Tableau 84 : Groupe écrit - taux de réalisation par type de liaison (tâche reformulation)

2.1.2.2 Nature de la CL réalisée

Quant à la nature de la CL des liaisons réalisées, la plupart des productions sont conformes à la CL cible (94%) (cf. Tableau 85). Néanmoins nous avons également relevé dans cette tâche la présence de formes orthographiques (4%) et de substitutions (3%).

Type liaison	Conforme	%	Orth	%	Substitution	%
Total CAT	**30**	**91%**	**1**	**3%**	**2**	**6%**
DET-NOM	**30**	**91%**	**1**	**3%**	**2**	**6%**
Cet	14	**88%**	0	**0%**	0	**0%**
Des	16	**94%**	1	**6%**	2	**12%**
Total VAR	**120**	**94%**	**5**	**4%**	**1**	**1%**
ADJ-NOM	**13**	**93%**	**0**	**0%**	**1**	**7%**
Ancien	13	**93%**	0	**0%**	1	**7%**
ADV-ADJ	**39**	**98%**	**0**	**0%**	**0**	**0%**
Moins	3	**100%**	0	**0%**	0	**0%**
Plus	14	**100%**	0	**0%**	0	**0%**
toujours	5	**100%**	0	**0%**	0	**0%**
Très	14	**100%**	0	**0%**	1	**7%**
Trop	3	**75%**	0	**0%**	0	**0%**
ADV-PPA	**14**	**100%**	**0**	**0%**	**0**	**0%**
Bien	14	**100%**	0	**0%**	0	**0%**
PRP-DET	**20**	**91%**	**2**	**9%**	**0**	**0%**
Devant	10	**100%**	0	**0%**	0	**0%**
Sous	10	**83%**	2	**17%**	0	**0%**
PRP-NOM	**8**	**80%**	**2**	**20%**	**0**	**0%**
Sans	8	**80%**	2	**20%**	0	**0%**
VER-ADJ	**26**	**96%**	**1**	**4%**	**0**	**0%**
C'est	14	**100%**	0	**0%**	0	**0%**
Est	3	**100%**	0	**0%**	0	**0%**
Suis	9	**90%**	1	**10%**	0	**0%**
Total général	**150**	**94%**	**6**	**4%**	**4**	**3%**

Tableau 85 : Résultats totaux - nature de la CL (tâche reformulation)

Nous aurions en outre pu nous attendre à observer un plus grand nombre de formes orthographiques chez le groupe écrit. Celles-ci sont au contraire reparties entre les deux groupes. Les apprenants exposés à un input écrit (cf. Tableau 87) ne réalisent pas plus de formes orthographiques que les apprenants ayant été soumis à un input uniquement oral (cf. Tableau 86). Bien que nous soyons conscients que notre échantillon est trop limité pour pouvoir tirer des conclusions, ces résultats nous confirment néanmoins la présence de formes orthographiques de liaison dans des productions de parole semi-spontanée.

Groupe oral						
Type liaison	**Conforme**	**%**	**Orth**	**%**	**Pause-hésitation**	**%**
Total CAT	**16**	**89%**	**1**	**6%**	**1**	**6%**
DET-NOM	**16**	**89%**	**1**	**6%**	**1**	**6%**
Cet	8	**89%**	0	**0%**	1	**11%**
Des	8	**89%**	1	**11%**	0	**0%**
Total VAR	**50**	**91%**	**3**	**5%**	**2**	**4%**
ADJ-NOM	**6**	**86%**	**0**	**0%**	**1**	**14%**
Ancien	6	**86%**	0	**0%**	1	**14%**
ADV-ADJ	**14**	**93%**	**0**	**0%**	**1**	**7%**
Moins	0	**0%**	0	**0%**	0	**0%**
Plus	6	**100%**	0	**0%**	0	**0%**
toujours	2	**100%**	0	**0%**	0	**0%**
Très	6	**100%**	0	**0%**	0	**0%**
Trop	0	**0%**	0	**0%**	1	**100%**
ADV-PPA	**7**	**100%**	**0**	**0%**	**0**	**0%**
Bien	7	**100%**	0	**0%**	0	**0%**
PRP-DET	**9**	**82%**	**2**	**18%**	**0**	**0%**
Devant	4	**100%**	0	**0%**	0	**0%**
Sous	5	**71%**	2	**29%**	0	**0%**
PRP-NOM	**3**	**75%**	**1**	**25%**	**0**	**0%**
Sans	3	**75%**	1	**25%**	0	**0%**
VER-ADJ	**11**	**100%**	**0**	**0%**	**0**	**0%**
C'est	6	**100%**	0	**0%**	0	**0%**
Est	1	**100%**	0	**0%**	0	**0%**
Suis	4	**100%**	0	**0%**	0	**0%**
Total général	**66**	**90%**	**4**	**5%**	**3**	**4%**

Tableau 86 : Groupe oral - nature de la CL (tâche reformulation)

Groupe écrit						
Type liaison	**Conforme**	**%**	**Orth**	**%**	**Pause-hésitation**	**%**
Total CAT	**14**	**93%**	**0**	**0%**	**1**	**7%**
DET-NOM	14	**93%**	0	**0%**	1	**7%**
Cet	6	**86%**	0	**0%**	1	**14%**
Des	8	**100%**	0	**0%**	0	**0%**
Total VAR	**70**	**97%**	**2**	**3%**	**0**	**0%**
ADJ-NOM	**7**	**100%**	**0**	**0%**	**0**	**0%**
Ancien	7	**100%**	0	**0%**	0	**0%**
ADV-ADJ	**25**	**100%**	**0**	**0%**	**0**	**0%**
Moins	3	**100%**	0	**0%**	0	**0%**
Plus	8	**100%**	0	**0%**	0	**0%**
Toujours	3	**100%**	0	**0%**	0	**0%**
Très	8	**100%**	0	**0%**	0	**0%**
Trop	3	**100%**	0	**0%**	0	**0%**
ADV-PPA	**7**	**100%**	**0**	**0%**	**0**	**0%**

Bien	7	**100%**	0	**0%**	0	**0%**
PRP-DET	**11**	**100%**	**0**	**0%**	**0**	**0%**
Devant	6	**100%**	0	**0%**	0	**0%**
Sous	5	**100%**	0	**0%**	0	**0%**
PRP-NOM	**5**	**83%**	**1**	**17%**	**0**	**0%**
Sans	5	**83%**	1	**17%**	0	**0%**
VER-ADJ	**15**	**94%**	**1**	**6%**	**0**	**0%**
C'est	8	**100%**	0	**0%**	0	**0%**
Est	2	**100%**	0	**0%**	0	**0%**
Suis	5	**83%**	1	**17%**	0	**0%**
Total général	**84**	**97%**	**2**	**2%**	**1**	**1%**

Tableau 87 : Groupe écrit - nature de la CL (tâche reformulation)

2.2. Bilan de l'étude transversale

En résumé, les résultats de l'étude transversale montrent que :

- les réalisations orthographiques sont attestées aussi bien en lecture qu'en parole semi-spontanée, dans la tâche de dénomination d'images comme de reformulation ;
- les résultats de la tâche de reformulation indiquent que les apprenants exposés à un input écrit ne produisent pas plus de réalisations que les apprenants ayant été soumis au seul input oral ;
- les apprenants exposés à un input écrit ne réalisent pas plus de formes orthographiques que les apprenants ayant été soumis à un input uniquement oral.

3. Résultats par type d'oral

Dans les paragraphes précédents, nous avons analysé séparément les résultats de l'étude longitudinale et transversale. Dans les pages qui suivent, nous nous attacherons à étudier les réalisations ainsi que la nature des CL réalisées par les apprenants sur la base du type de production orale dans lequel celles-ci ont été observées. Nous avons donc réparti les tâches de notre protocole selon trois typologies d'oral :

- Répétition : la tâche de répétition (cf. Partie 4.Chapitre 1.4.3.1.1) ;
- Lecture : (a) lecture des séquences isolées de liaison (cf. Partie 4.Chapitre 1.4.3.1.2) ; (b) lecture de texte PFC (cf. Partie 4.Chapitre 1.4.3.1.3) ; (c) lecture de texte journalistique (cf. Partie 4.Chapitre 1.4.3.1.5) ;

- Parole semi-spontanée : (a) tâche de dénomination d'image (cf. Partie 4.Chapitre 1.4.3.2.1) ; (b) tâche de reformulation (cf. Partie 4.Chapitre 1.4.3.2.3); (c) entretiens guidés (cf. Partie 4.Chapitre 1.4.3.1.7).

L'objectif de cette analyse est donc d'étudier l'effet de l'input écrit sur la production des liaisons chez les apprenants et notamment de vérifier l'hypothèse selon laquelle les apprenants produisent plus de liaisons orthographiques dans les tâches de lecture que dans les tâches de répétition et l'oral semi-spontané.

3.1. Répétition

3.1.1. Type de liaison potentielle

Le tableau ci-dessous (cf. Tableau 88) résume les types de liaisons potentielles (catégoriques et variables) que nous avons enregistrés dans la tâche de répétition de l'étude longitudinale.

Répétition	
Type de liaison	**Total général**
Total CAT	**955**
DET-NOM	**478**
Des	477
Un	1
EXF-EXF	**69**
Plus	69
PRO-AUX	**139**
Ils	139
PRO-VER	**133**
Ils	1
Vous	132
VER-PRO	**136**
Allons	67
Donnez	69
Total VAR	**1011**
ADJ-NOM	**369**
Grand	184
Petit	185
AUX-PPA	**138**
Ont	138
CON-PRO	**76**
Quand	76
NOM-ADJ	**136**
Efforts	69
Essais	67
PRP-NAM	**155**
En	155

VER-ADV	**137**
Sommes	68
Sort	8
Total général	**2047**

Tableau 88 : Types de liaison potentielle catégorique et variable (tâche de répétition)

3.1.2. Réalisation

Les résultats de la tâche de répétition (cf. Tableau 89) montrent que la liaison est toujours produite dans les types de liaison catégorique et qu'elle n'est presque jamais réalisée dans des contextes erratiques sauf pour deux occurrences de la séquence « au fur et // à mesure » (cf. Paragraphe 1.4).

En ce qui concerne les types de liaison variable, bien que l'input soumis aux apprenants présente toujours une CL réalisée, nous avons cependant enregistré des non-réalisations dans les productions des apprenants. C'est le cas par exemple du contexte ADJ-NOM pour lequel nous observons quelques liaisons non-réalisées, ou encore des types NOM-ADJ et CON-PRO où un nombre important de non-réalisations a été enregistré.

Pour ce qui est des liaisons sans enchaînement, quelques occurrences ont été relevées uniquement dans les contextes variables, mais le taux très faible suggère qu'elles correspondent à de simples hésitations.

Il convient de souligner qu'un nombre important d'input a donné lieu à des non-compréhensions, en particulier pendant la première séance de l'étude longitudinale.

Type de liaison	Non-réalisation	Enchaînement	Non-enchaînement	Enchaîn. incertain	Pause-hésitation	NC	Total général
Total CAT	**0**	**938**	**0**	**0**	**0**	**17**	**955**
DET-NOM	**0**	**475**	**0**	**0**	**0**	**3**	**478**
Des	0	475	0	0	0	2	477
Un	0	0	0	0	0	1	1
EXF-EXF	**0**	**66**	**0**	**0**	**0**	**3**	**69**
Plus	0	66	0	0	0	3	69
PRO-AUX	**0**	**133**	**0**	**0**	**0**	**6**	**139**
Ils	0	133	0	0	0	6	139
PRO-VER	**0**	**130**	**0**	**0**	**0**	**3**	**133**
Ils	0	0	0	0	0	1	1
Vous	0	130	0	0	0	2	132
VER-PRO	**0**	**134**	**0**	**0**	**0**	**2**	**136**
Allons	0	67	0	0	0	0	67
Donnez	0	67	0	0	0	2	69
Total ERR	**54**	**2**	**0**	**0**	**0**	**25**	**81**
EXF-EXF	**54**	**2**	**0**	**0**	**0**	**25**	**81**
Et	54	2	0	0	0	25	81
Total VAR	**39**	**829**	**8**	**1**	**6**	**128**	**1011**
ADJ-NOM	**12**	**299**	**3**	**0**	**2**	**53**	**369**

Grand	4	156	2	0	2	20	184
Petit	8	143	1	0	0	33	185
ADV-PRO	**0**	**55**	**0**	**0**	**0**	**21**	**76**
Quand	0	55	0	0	0	21	76
AUX-PPA	**4**	**129**	**0**	**0**	**0**	**5**	**138**
Ont	4	129	0	0	0	5	138
NOM-ADJ	**22**	**99**	**5**	**1**	**3**	**6**	**136**
Efforts	6	51	5	1	0	6	69
Essais	16	48	0	0	3	0	67
PRP-NAM	**1**	**114**	**0**	**0**	**1**	**39**	**155**
En	1	114	0	0	1	39	155
VER-ADV	**0**	**133**	**0**	**0**	**0**	**4**	**137**
Sommes	0	68	0	0	0	0	68
Sort	0	65	0	0	0	4	69
Total général	**93**	**1769**	**8**	**1**	**6**	**170**	**2047**

Tableau 89 : Résultats totaux - Réalisations par type de liaison (tâche de répétition)

3.1.3. Nature de la CL réalisée

Si la répétition assure un taux élevé de réalisations (bien que des non-réalisations soient attestées dans les contextes de liaison variable), la nature de la CL réalisée est quant à elle sujette à variations. En effet, nous avons relevé au cours de cette tâche la présence d'un taux significatif de réalisations orthographiques à la fois dans les contextes catégoriques (13%) et variables (10%) (cf. Tableau 90).

Parmi les contextes catégoriques, la plupart des formes orthographiques ont été produites après le déterminant « des » et le pronom personnel « ils ». En revanche, parmi les types de liaison variable, le type ADJ-NOM (après l'adjectif « grand ») et le type NOM-ADJ (dans les séquences « efforts inutiles » et « essais intéressants ») présentent le plus grand nombre de réalisations orthographiques.

	Conforme	**Orth**	**CL présente dans le M1**	**Substitution**	**NC**	**Total général**
Total CAT	**816**	**121**	**0**	**1**	**17**	**955**
DET-NOM	**431**	**44**	**0**	**0**	**3**	**478**
Des	431	44	0	0	2	477
Un	0	0	0	0	1	1
EXF-EXF	**61**	**5**	**0**	**0**	**3**	**69**
Plus	61	5	0	0	3	69
PRO-AUX	**78**	**55**	**0**	**0**	**6**	**139**
Ils	78	55	0	0	6	139
PRO-VER	**122**	**8**	**0**	**0**	**3**	**133**
Ils	0	0	0	0	1	1
Vous	122	8	0	0	2	132

VER-PRO	**124**	**9**	**0**	**1**	**2**	**136**
Allons	60	7	0	0	0	67
Donnez	64	2	0	1	2	69
Total ERR	**2**	**0**	**0**	**0**	**25**	**27**
EXF-EXF	**2**	**0**	**0**	**0**	**25**	**27**
Et	2	0	0	0	25	27
Total VAR	**744**	**99**	**1**	**0**	**128**	**972**
ADJ-NOM	**282**	**21**	**1**	**0**	**53**	**357**
Grand	138	21	1	0	20	180
Petit	144	0	0	0	33	177
AUX-PPA	**129**	**0**	**0**	**0**	**5**	**134**
Ont	129	0	0	0	5	134
CON-PRO	**53**	**2**	**0**	**0**	**21**	**76**
Quand	53	2	0	0	21	76
NOM-ADJ	**59**	**49**	**0**	**0**	**6**	**114**
Efforts	20	37	0	0	6	63
Essais	39	12	0	0	0	51
PRP-NAM	**115**	**0**	**0**	**0**	**39**	**154**
En	115	0	0	0	39	154
VER-ADV	**106**	**27**	**0**	**0**	**4**	**137**
Sommes	41	27	0	0	0	68
Sort	65	0	0	0	4	69
Total général	**1562**	**220**	**1**	**1**	**170**	**1954**

Tableau 90 : Résultats totaux - nature de la CL par type de liaison (tâche de répétition)

3.2. Lecture oralisée

3.2.1. Type de liaison potentielle

En ce qui concerne les tâches de lecture oralisée, nous avons regroupé les lectures des séquences isolées de liaison et les lectures des textes. Malgré cette hétérogénéité, l'ensemble de ces tâches comprend un nombre relativement comparable de sites de liaison potentielle.

Ceux-ci sont résumés dans le tableau ci-dessous (cf. Tableau 91).

	Lecture			
Type de liaison	**Lecture séquences**	**Lecture texte PFC**	**Lecture texte journalistique**	**Total général**
Total CAT	**894**	**915**	**2165**	**3974**
DET-ADJ	0	67	71	138
DET-NOM	409	368	675	1452
EXF-EXF	68	140	70	278
NUM-NOM	0	0	416	416
PDE-NOM	0	2	211	213
PRO-AUX	144	208	216	568
PRO-PRO	0	61	154	215
PRO-VER	137	69	218	424
VER-PRO	136	0	134	270

Total VAR	**1023**	**1241**	**2048**	**4312**
ADJ-NOM	340	137	348	825
ADV-ADJ	0	70	68	138
ADV-DET	0	0	67	67
ADV-INF	0	1	0	1
ADV-PRO	0	0	68	68
ADV-PRP	0	136	64	200
AUX-ADV	1	0	79	80
AUX-PPA	140	69	70	279
AUX-PRP	0	139	0	139
CON-PRO	66	0	75	141
INF-DET	0	67	0	67
INF-PRP	0	68	0	68
NOM-ADJ	136	205	206	547
NOM-AUX	0	68	0	68
NOM-PRP	0	68	0	68
NOM-VER	0	0	2	2
PRO-PRP	0	0	69	69
PRP-DET	0	76	1	77
PRP-NAM	136	0	0	136
PRP-NOM	0	1	70	71
VER-ADJ	1	0	208	209
VER-ADV	203	0	197	400
VER-DET	0	68	139	207
VER-INF	0	0	56	56
VER-PRP	0	68	261	329
Total général	**1917**	**2156**	**4213**	**8286**

Tableau 91 : Types de liaison potentielle catégorique et variable (tâches de lecture)

3.2.2. Réalisation

Les résultats de l'analyse des réalisations (cf. Tableau 92) montrent que même en lecture, il persiste 16% de non-réalisations en contexte catégorique et 61% en contexte variable. La plupart des non-réalisations en contexte catégorique ont été enregistrées encore une fois après l'adjectif possessif « leurs », après l'indéfini « quelques » ainsi qu'après les expressions figées « jeux olympiques » et « de temps en temps ». Cela tend à prouver que même en présence d'un input écrit, l'apprenant n'est pas en mesure de saisir le degré de cohésion qui unit ces séquences. La liaison après les numéraux « deux » et « cents » semblent également donner lieu à des non-réalisations alors que la liaison est complétement maîtrisée après « dix ». De plus, nous avons relevé un écart de réalisations entre « des » ayant valeur d'article indéfini (après lequel la liaison est presque toujours produite), et la même forme ayant la fonction de préposition. Quant à la liaison après pronom personnel sujet, nous pouvons mettre en évidence les mêmes tendances qu'il nous a déjà été donné d'observer dans l'analyse de l'étude longitudinale, à savoir que la

liaison est plus souvent produite devant un auxiliaire que devant d'autres verbes ou clitiques. Nous observons donc une forte idiosyncrasie interne au type de liaison catégorique mais également variable. En effet, si la liaison n'est presque jamais produite dans le type NOM-ADJ, les apprenants produisent plus de liaisons après les noms « enfants » et « conseils » qu'après d'autres noms tels que « cocons », « intérêts » ou « visites ». En revanche, la liaison dans le contexte ADJ-NOM est réalisée de façon homogène à hauteur de 58% après les adjectifs « petit », « grand » et « premier ». De nombreux cas (78%) de réalisations des liaisons erratiques devant « h » aspiré ont été produits dans la séquence « plus // haut » (cf. Paragraphe 1.4).

Si les liaisons non-enchaînées semblent apparaître presque exclusivement dans des contextes de liaison variable, les pauses-hésitations en revanche sont reparties entre des contextes catégoriques et variables.

	Lecture					
Type de liaison	**Non-réalisation**	**Enchaînement**	**Non-enchaînement**	**Enchain. incertain**	**Pause-hésitation**	**Total**
Total CAT	**636**	**3288**	**3**	**0**	**47**	**3974**
DET-ADJ	**9**	**129**	**0**	**0**	**0**	**138**
Un	9	129	0	0	0	138
DET-NOM	**114**	**1327**	**1**	**0**	**10**	**1452**
Ces	6	138	1	0	2	147
Des	5	476	0	0	1	482
Les	16	477	0	0	4	497
Leurs	32	10	0	0	3	45
Quelques	32	43	0	0	0	75
Son	23	183	0	0	0	206
EXF-EXF	**97**	**177**	**0**	**0**	**4**	**278**
En	5	64	0	0	0	69
Jeux	54	13	0	0	4	71
Temps	38	100	0	0	0	138
NUM-NOM	**176**	**234**	**1**	**0**	**5**	**416**
Cents	134	67	1	0	5	207
Deux	42	96	0	0	0	138
Dix	0	68	0	0	0	68
Trente-huit	0	3	0	0	0	3
PDE-NOM	**20**	**186**	**1**	**0**	**6**	**213**
Aux	0	2	0	0	0	2
Des	20	184	1	0	6	211
PRO-AUX	**68**	**492**	**0**	**0**	**8**	**568**
En	15	30	0	0	4	49
Ils	4	149	0	0	1	154
Les	0	1	0	0	0	1
Nous	5	134	0	0	2	141
On	1	84	0	0	1	86
Tout	42	31	0	0	0	73
Vous	1	63	0	0	0	64

PRO-PRO	**48**	**161**	**0**	**0**	**6**	**215**
En	1	0	0	0	1	2
Ils	30	34	0	0	5	69
Nous	8	76	0	0	0	84
On	9	51	0	0	0	60
PRO-VER	**54**	**366**	**0**	**0**	**4**	**424**
En	19	48	0	0	2	69
Ils	27	109	0	0	2	138
Nous	1	145	0	0	0	146
On	7	61	0	0	0	68
Vous	0	3	0	0	0	3
VER-PRO	**50**	**216**	**0**	**0**	**4**	**270**
Allez	4	64	0	0	0	68
Dit	4	61	0	0	0	65
Prenez	17	49	0	0	2	68
Sont	25	42	0	0	2	69
Total ERR	**462**	**86**	**0**	**0**	**0**	**548**
ADV-ADJ	**15**	**52**	**0**	**0**	**0**	**67**
Plus	15	52	0	0	0	67
ADV-DET	**65**	**3**	**0**	**0**	**0**	**68**
Vraiment	65	3	0	0	0	68
ADV-PRP	**67**	**1**	**0**	**0**	**0**	**68**
Comment	67	1	0	0	0	68
CON-VER	**0**	**9**	**0**	**0**	**0**	**9**
Et	0	9	0	0	0	9
EXF-EXF	**58**	**9**	**0**	**0**	**0**	**67**
Et	58	9	0	0	0	67
NAM-PRP	**65**	**3**	**0**	**0**	**0**	**68**
Berlin	65	3	0	0	0	68
NOM-AUX	**63**	**4**	**0**	**0**	**0**	**67**
Opposition	63	4	0	0	0	67
NOM-PRO	**63**	**2**	**0**	**0**	**0**	**65**
Coin	63	2	0	0	0	65
NOM-PRP	**66**	**2**	**0**	**0**	**0**	**68**
Région	66	2	0	0	2	68
PRP-VER	**0**	**1**	**0**	**0**	**0**	**1**
Et	0	1	0	0	0	1
Total VAR	**2643**	**1608**	**16**	**1**	**44**	**4312**
ADJ-NOM	**327**	**482**	**7**	**0**	**9**	**825**
Grand	196	247	1	0	6	450
Petit	101	198	6	0	3	308
Premier	30	37	0	0	0	67
ADV-ADJ	**22**	**115**	**0**	**0**	**1**	**138**
Plus	13	54	0	0	1	68
Très	9	61	0	0	0	70
ADV-DET	**49**	**17**	**1**	**0**	**0**	**67**
Pas	49	17	1	0	0	67
ADV-INF	**0**	**1**	**0**	**0**	**0**	**1**
Plus	0	1	0	0	0	1

ADV-PRO	**56**	**12**	**0**	**0**	**0**	**68**
Mais	56	12	0	0	0	68
ADV-PRP	**190**	**10**	**0**	**0**	**0**	**200**
Pas	2	0	0	0	0	2
Plus	120	10	0	0	0	130
Toujours	68	0	0	0	0	68
AUX-ADV	**45**	**35**	**0**	**0**	**0**	**80**
Ont	37	32	0	0	0	69
Sont	8	3	0	0	0	11
AUX-PPA	**152**	**121**	**1**	**0**	**5**	**279**
Ont	150	121	1	0	5	277
Sont	2	0	0	0	0	2
AUX-PRP	**118**	**21**	**0**	**0**	**0**	**139**
Est	118	21	0	0	0	139
CON-PRO	**39**	**102**	**0**	**0**	**0**	**141**
Quand	39	102	0	0	0	141
INF-DET	**65**	**2**	**0**	**0**	**0**	**67**
Provoquer	65	2	0	0	0	67
INF-PRP	**66**	**2**	**0**	**0**	**0**	**68**
Trouver	66	2	0	0	0	68
NOM-ADJ	**472**	**57**	**2**	**0**	**16**	**547**
Circuits	63	5	0	0	0	68
Cocons	63	7	1	0	1	72
Conseils	50	14	0	0	4	68
Emeutes	64	0	0	0	0	64
Enfants	36	23	1	0	8	68
Intérêts	61	6	0	0	3	70
Pâtes	68	0	0	0	0	68
Visites	67	2	0	0	0	69
NOM-AUX	**68**	**0**	**0**	**0**	**0**	**68**
Fanatiques	68	0	0	0	0	68
NOM-PRP	**68**	**0**	**0**	**0**	**0**	**68**
Amies	1	0	0	0	0	1
Chemises	67	0	0	0	0	67
NOM-VER	**0**	**1**	**0**	**0**	**1**	**2**
Résidents	0	1	0	0	1	2
PRO-PRP	**49**	**20**	**0**	**0**	**0**	**69**
Premier	49	19	0	0	0	68
Tout	0	1	0	0	0	1
PRP-DET	**13**	**57**	**3**	**0**	**4**	**77**
Dans	13	57	3	0	4	77
PRP-NAM	**3**	**133**	**0**	**0**	**0**	**136**
En	3	133	0	0	0	136
PRP-NOM	**6**	**64**	**0**	**0**	**1**	**71**
Aux	0	0	0	0	1	1
Sans	6	64	0	0	0	70
VER-ADJ	**145**	**61**	**1**	**0**	**2**	**209**
Est	95	44	0	0	1	140
Sont	50	17	1	0	0	68

Vient	0	0	0	0	1	1
VER-ADV	**214**	**183**	**1**	**1**	**1**	**400**
Descend	7	60	1	0	0	68
Est	2	0	0	0	0	2
Habitent	68	0	0	0	0	68
Sont	53	75	0	1	0	129
Vient	19	47	0	0	1	67
Vivent	65	1	0	0	0	66
VER-DET	**175**	**30**	**0**	**0**	**2**	**207**
Est	42	27	0	0	2	71
Faut	65	3	0	0	0	68
Préparent	68	0	0	0	0	68
VER-INF	**54**	**1**	**0**	**0**	**1**	**56**
Voulons	54	1	0	0	1	56
VER-PRP	**247**	**81**	**0**	**0**	**1**	**329**
Est	77	49	0	0	1	127
Fleurissent	62	5	0	0	0	67
Sent	59	8	0	0	0	67
Sont	49	19	0	0	0	68
Total général	**3741**	**4982**	**19**	**1**	**91**	**8834**

Tableau 92 : Résultats totaux - Réalisations par type de liaison (tâches de lecture)

3.2.3. Nature de la CL réalisée

Comme nous l'avons vu dans le paragraphe précédent, les réalisations orthographiques de la liaison sont produites en répétition à hauteur de 11% (tous contextes confondus) (cf. Tableau 90). Étant donné la présence d'un support écrit dans les tâches de lecture, nous pouvons nous attendre à enregistrer un plus grand nombre de ces réalisations orthographiques dans ce type de tâche (hypothèse 3) (cf. Partie 4.Chapitre 3.1.3). Cependant, les résultats de l'ensemble des lectures de notre protocole montrent que les apprenants produisent un pourcentage de forme orthographique (12%) comparable à celui déjà observé en répétition (11%). Pour être plus précis, ce taux correspond à 8% en contexte de liaison catégorique et à 20% en contexte variable. Parmi les contextes catégoriques, le taux le plus important de formes orthographiques a été produit après l'adjectif possessif « ces » et l'adjectif indéfini « quelques » ou encore après le pronom personnel « ils » (non seulement devant auxiliaire mais aussi devant n'importe quel autre verbe).

Dans le cas des contextes variables, la presque totalité des formes orthographiques correspond à la liaison après l'adjectif « grand » et la conjonction « quand ».

Comme nous l'avons observé dans les résultats de l'étude longitudinale, la plupart des réalisations comprenant une prononciation d'une consonne présente dans le M1 sont attestées dans l'expression figée « de temps en temps » et après le numéral « cents ». Ces formes sont

également produites dans le type de liaison variable NOM-ADJ après certains noms pluriels ayant une consonne /t/ qui précède la CL /z/ (ex. « enfants »). Il est de plus intéressant d'observer que la quasi-totalité des substitutions de la CL sont réalisées dans les tâches de lecture.

Ces résultats sont d'autant plus inattendus si nous considérons que les apprenants sont exposés au cours de ces tâches à un input écrit qui devrait assurer une plus grande conformité de la production de la CL par rapport à la cible.

Type de liaison	Conforme	Orth	CL présente dans le M1	Substitution	Total
Total CAT	**2951**	**256**	**109**	**22**	**3338**
DET-ADJ	**128**	**0**	**0**	**1**	**129**
Un	128	0	0	1	129
DET-NOM	**1211**	**125**	**0**	**2**	**1338**
Ces	124	17	0	0	141
Des	419	58	0	0	477
Les	443	38	0	0	481
Leurs	11	2	0	0	13
Quelques	33	10	0	0	43
Son	181	0	0	2	183
EXF-EXF	**131**	**11**	**38**	**1**	**181**
En	64	0	0	0	64
Jeux	16	0	0	1	17
Temps	51	11	38	0	100
NUM-NOM	**151**	**15**	**71**	**3**	**240**
Cents	2	0	71	0	73
Deux	86	10	0	0	96
Dix	63	5	0	0	68
Trente-huit	0	0	0	3	3
PDE-NOM	**169**	**23**	**0**	**1**	**193**
Aux	2	0	0	0	2
Des	167	23	0	1	191
PRO-AUX	**447**	**47**	**0**	**6**	**500**
En	34	0	0	0	34
Ils	108	42	0	0	150
Les	1	0	0	0	1
Nous	133	1	0	2	136
On	85	0	0	0	85
Tout	28	0	0	3	31
Vous	58	4	0	1	63
PRO-PRO	**157**	**10**	**0**	**0**	**167**
En	1	0	0	0	1
Ils	32	7	0	0	39
Nous	73	3	0	0	76
On	51	0	0	0	51
PRO-VER	**339**	**25**	**0**	**6**	**370**

En	44	0	0	6	50
Ils	92	19	0	0	111
Nous	139	6	0	0	145
On	61	0	0	0	61
Vous	3	0	0	0	3
VER-PRO	**218**	**0**	**0**	**2**	**220**
Allez	64	0	0	0	64
Dit	61	0	0	0	61
Prenez	50	0	0	1	51
Sont	43	0	0	1	44
Total ERR	**83**	**3**	**0**	**0**	**86**
ADV-ADJ	**49**	**3**	**0**	**0**	**52**
Plus	49	3	0	0	52
ADV-DET	**3**	**0**	**0**	**0**	**3**
Vraiment	3	0	0	0	3
ADV-PRP	**1**	**0**	**0**	**0**	**1**
Comment	1	0	0	0	1
CON-VER	**9**	**0**	**0**	**0**	**9**
Et	9	0	0	0	9
EXF-EXF	**9**	**0**	**0**	**0**	**9**
Et	9	0	0	0	9
NAM-PRP	**3**	**0**	**0**	**0**	**3**
Berlin	3	0	0	0	3
NOM-AUX	**4**	**0**	**0**	**0**	**4**
Opposition	4	0	0	0	4
NOM-PRO	**2**	**0**	**0**	**0**	**2**
Coin	2	0	0	0	2
NOM-PRP	**2**	**0**	**0**	**0**	**2**
Région	2	2	2	2	2
PRP-VER	**1**	**0**	**0**	**0**	**1**
Et	1	0	0	0	1
Total VAR	**1294**	**327**	**26**	**22**	**1669**
ADJ-NOM	**345**	**137**	**0**	**16**	**498**
Grand	104	137	0	13	254
Petit	204	0	0	3	207
Premier	37	0	0	0	37
ADV-ADJ	**104**	**12**	**0**	**0**	**116**
Plus	52	3	0	0	55
Très	52	9	0	0	61
ADV-DET	**18**	**0**	**0**	**0**	**18**
Pas	18	0	0	0	18
ADV-INF	**1**	**0**	**0**	**0**	**1**
Plus	1	0	0	0	1
ADV-PRO	**12**	**0**	**0**	**0**	**12**
Mais	12	0	0	0	12
ADV-PRP	**10**	**0**	**0**	**0**	**10**
Plus	10	0	0	0	10
AUX-ADV	**34**	**0**	**0**	**1**	**35**
Ont	31	0	0	1	32

Sont	3	0	0	0	3
AUX-PPA	**125**	**0**	**0**	**2**	**127**
Ont	125	0	0	2	127
AUX-PRP	**21**	**0**	**0**	**0**	**21**
Est	21	0	0	0	21
CON-PRO	**51**	**51**	**0**	**0**	**102**
Quand	51	51	0	0	102
INF-DET	**2**	**0**	**0**	**0**	**2**
Provoquer	2	0	0	0	2
INF-PRP	**2**	**0**	**0**	**0**	**2**
Trouver	2	0	0	0	2
NOM-ADJ	**29**	**20**	**24**	**0**	**75**
Circuits	1	0	3	0	5
Cocons	6	3	0	0	9
Conseils	14	4	0	0	18
Enfants	5	9	18	0	32
Intérêts	2	3	3	0	9
Visites	1	1	0	0	2
NOM-VER	**2**	**0**	**0**	**0**	**2**
Résidents	2	0	0	0	2
PRO-PRP	**20**	**0**	**0**	**0**	**20**
Premier	19	0	0	0	19
Tout	1	0	0	0	1
PRP-DET	**39**	**25**	**0**	**0**	**64**
Dans	39	25	0	0	64
PRP-NAM	**133**	**0**	**0**	**0**	**133**
En	133	0	0	0	133
PRP-NOM	**35**	**28**	**0**	**2**	**65**
Aux	1	0	0	0	1
Sans	34	28	0	2	64
VER-ADJ	**62**	**1**	**0**	**1**	**64**
Est	43	1	0	1	45
Sont	18	0	0	0	18
Vient	1	0	0	0	1
VER-ADV	**135**	**51**	**0**	**0**	**186**
Descend	11	50	0	0	61
Sont	76	0	0	0	76
Vient	46	1	0	0	47
Vivent	1	0	0	0	1
Vient	1	0	0	0	1
VER-DET	**32**	**0**	**0**	**0**	**32**
Est	29	0	0	0	29
Faut	3	0	0	0	3
VER-INF	**2**	**0**	**0**	**0**	**2**
Voulons	2	0	0	0	2
VER-PRP	**80**	**2**	**0**	**0**	**82**
Est	48	2	0	0	50
Fleurissent	5	0	0	0	5
Sent	8	0	0	0	8

Sont	19	0	0	0	19
Total général	**4328**	**586**	**135**	**44**	**5093**

Tableau 93 : Résultats totaux - Nature de la CL par type de liaison (tâches de lecture)

3.3. Parole semi-spontanée

3.3.1. Type de liaison potentielle

Comme nous l'avons déjà souligné à plusieurs reprises, le nombre de sites potentiels de liaison en contexte de parole spontanée est bien plus faible que celui enregistré en lecture.

De plus, ces réalisations sont souvent réparties sur plusieurs types de liaison différents. Le tableau ci-dessous (cf. Tableau 94) résume l'ensemble des productions en parole semi-spontanée.

Type de liaison	Parole semi-spontanée
Total CAT	**285**
DET-ADJ	**6**
Des	1
Les	3
Un	2
DET-NOM	**176**
Cet	25
Des	52
Les	45
Mes	27
Mon	2
Plusieurs	1
Ses	1
Trois	1
Un	22
DET-PRO	**3**
Les	3
EXF-EXF	**18**
En	7
Plus	7
Tout	2
Vingt	2
NUM-NOM	**29**
Cents	1
Deux	11
Quatre-vingts	1
Trente-six	1
Trois	4
Un	1
Vingt	10
PDE-ADJ	**1**
Des	1

PDE-NOM	**2**
Aux	2
PRO-AUX	**34**
Ils	3
Les	3
Nous	24
On	4
PRO-VER	**16**
Elles	1
Ils	4
Les	1
Nous	5
On	5
Total VAR	**658**
ADJ-NOM	**190**
Ancien	17
Autres	1
Bon	1
Grand	78
Gros	8
Mêmes	1
Petit	83
Petits	1
ADV-ADJ	**98**
Assez	1
Moins	16
Plus	19
Toujours	17
Très	27
Trop	16
Vraiment	2
ADV-ADV	**21**
Mais	10
Pas	8
Plus	1
Puis	2
ADV-DET	**2**
Pas	1
Seulement	1
ADV-INF	**3**
Bien	1
Pas	2
ADV-NOM	**4**
Moins	1
Pas	1
Plus	1
Seulement	1
ADV-PPA	**23**

Bien	19
Jamais	1
Pas	1
Toujours	2
ADV-PRP	**3**
Mais	1
Pas	2
AUX-INF	**1**
Vais	1
AUX-PPA	**20**
Est	1
Ont	1
Sommes	4
Sont	4
Suis	10
AUX-PRP	**12**
Est	2
Sommes	1
Suis	9
INF-ADV	**2**
Commencer	1
Fréquenter	1
NOM-ADJ	**7**
Cultures	1
Langues	5
Relations	1
NOM-ADV	**1**
Jours	1
PPA-ADV	**1**
Fait	1
PRO-INF	**1**
Nous	1
PRO-PRP	**1**
Tout	1
PRP-DET	**92**
Chez	3
Dans	21
Devant	34
Sous	34
PRP-NAM	**8**
En	8
PRP-NOM	**21**
Dans	1
En	4
Sans	16
PRP-PRO	**1**
Chez	1
VER-ADJ	**62**
C'est	17

Est	19
Etait	5
Suis	21
VER-ADV	**15**
Avait	1
Devons	1
Dois	3
Était	1
Parlent	1
Pouvais	1
Sais	1
Sont	1
Suis	1
Voudrais	4
VER-DET	**26**
Avait	1
Est	19
Étais	1
Était	3
Sont	1
Suis	1
VER-INF	**27**
Devais	1
Devrais	4
Dois	4
Fait	1
Faut	2
Peux	3
Préférais	1
Serais	2
Voudrais	7
Voulais	2
VER-NOM	**2**
Suis	2
VER-PDE	**2**
Suis	2
VER-PRP	**12**
Est	2
Étais	1
Sommes	2
Suis	6
Vais	1
Total général	**943**

Tableau 94 : Types de liaison potentielle catégorique et variable (tâches de parole semi-spontanée)

3.3.2. Réalisation

En parole spontanée, les apprenants maintiennent un taux de non-réalisations assez faible en contexte de liaison catégorique, de l'ordre de 6% contre 16% en lecture et 0% en répétition (cf. Tableau 95)[305]. Il convient toutefois de souligner qu'en lecture, l'apprenant est exposé à des séquences plus rares qu'il n'utilise pas forcement en parole spontanée puisqu'elles ne font pas partie de son interlangue. C'est le cas par exemple du type DET-NOM, qui en lecture comprend des liaisons après « quelques » et « leurs » donnant le plus souvent lieu à des non-réalisations. Néanmoins, si on soustrait ces dernières du total des non-réalisations du type DET-NOM, les résultats des non-réalisations de la lecture sont tout à fait comparables avec ceux de l'oral semi-spontané et correspondent à un pourcentage de 6% des non-réalisations (50 / 1388 en lecture et 11 / 162 en parole semi-spontanée). Cela revient à dire qu'il existe peu de différences entre le taux de réalisations de la liaison après déterminants dans la lecture et la parole spontanée.

Il en va de même pour la réalisation de la liaison après les pronoms personnels sujets. Le taux des non-réalisations reste stable après les pronoms personnels sujets « ils », « nous » et « vous » entre les trois types de productions soumises aux apprenants (répétition, lecture et parole semi-spontanée).

En revanche, les résultats totaux en contexte de liaison variable montrent un écart d'environ 10% entre les non-réalisations en lecture (61%) et en parole semi-spontanée. Ces données s'expliquent par le fait que, comme l'a souligné De Moras (2011), les apprenants de FLE en lecture sont exposés à des sites de liaison et des séquences syntaxiques qu'ils auraient difficilement produits de manière spontanée.

Quant aux réalisations non-enchaînées, nous n'avons enregistré aucune forme de liaison non-enchaînée en parole semi-spontanée.

La faible présence de ces dernières réalisations par rapport à celles observées en lecture semble confirmer une fois de plus que les non-enchaînements relèvent des productions disfluentes influencées par l'input écrit.

	Parole semi-spontanée			
Type de liaison	**Non-réalisation**	**Enchaînement**	**Pause-hésitation**	**Total**
Total CAT	**18**	**261**	**6**	**285**
DET-ADJ	**0**	**6**	**0**	**6**
Des	0	1	0	1
Les	0	3	0	3
Un	0	2	0	2

[305] Nous n'avons enregistré que deux réalisations de liaison erratique (enchaînées et conformes à la cible) dans les tâches de parole semi-spontanée à savoir dans la séquence « plus // haut ».

DET-NOM	**11**	**162**	**3**	**176**
Cet	2	23	0	25
Des	2	49	1	52
Les	5	38	2	45
Mes	0	27	0	27
Mon	1	1	0	2
Plusieurs	1	0	0	1
Ses	0	1	0	1
Un	0	22	0	22
DET-PRO	**0**	**3**	**0**	**3**
Les	0	3	0	3
EXF-EXF	**0**	**18**	**0**	**18**
En	0	7	0	7
Plus	0	7	0	7
Tout	0	2	0	2
Vingt	0	2	0	2
NUM-NOM	**2**	**26**	**1**	**29**
Cents	0	1	0	1
Deux	1	10	0	11
Quatre-vingts	0	0	1	1
Trente-six	0	1	0	1
Trois	0	4	0	4
Un	0	1	0	1
Vingt	1	9	0	10
PDE-ADJ	**0**	**1**	**0**	**1**
Des	0	1	0	1
PDE-NOM	**0**	**2**	**0**	**2**
Aux	0	2	0	2
PRO-AUX	**3**	**30**	**1**	**34**
Ils	1	2	0	3
Les	2	1	0	3
Nous	0	23	1	24
On	0	4	0	4
PRO-VER	**2**	**13**	**1**	**16**
Elles	0	1	0	1
Ils	1	3	0	4
Les	1	0	0	1
Nous	0	4	1	5
On	0	5	0	5
Total VAR	**335**	**317**	**6**	**658**
ADJ-NOM	**70**	**116**	**4**	**190**
Ancien	3	14	0	17
Autres	1	0	0	1
Bon	0	1	0	1
Grand	32	44	2	78
Gros	8	0	0	8
Mêmes	1	0	0	1
Petit	24	57	2	83
Petits	1	0	0	1

ADV-ADJ	**47**	**51**	**0**	**98**
Assez	1	0	0	1
Moins	13	3	0	16
Plus	4	15	0	19
Toujours	12	5	0	17
Très	3	24	0	27
Trop	12	4	0	16
Vraiment	2	0	0	2
ADV-ADV	**15**	**6**	**0**	**21**
Mais	7	3	0	10
Pas	6	2	0	8
Plus	0	1	0	1
Puis	2	0	0	2
ADV-DET	**1**	**1**	**0**	**2**
Pas	1	0	0	1
Seulement	0	1	0	1
ADV-INF	**2**	**1**	**0**	**3**
Bien	0	1	0	1
Pas	2	0	0	2
ADV-NOM	**3**	**1**	**0**	**4**
Moins	1	0	0	1
Pas	1	0	0	1
Plus	0	1	0	1
Seulement	1	0	0	1
ADV-PPA	**7**	**16**	**0**	**23**
Bien	5	14	0	19
Jamais	0	1	0	1
Pas	0	1	0	1
Toujours	2	0	0	2
ADV-PRP	**2**	**1**	**0**	**3**
Mais	1	0	0	1
Pas	1	1	0	2
AUX-INF	**1**	**0**	**0**	**1**
Vais	1	0	0	1
AUX-PPA	**12**	**7**	**1**	**20**
Est	1	0	0	1
Ont	1	0	0	1
Sommes	1	2	1	4
Sont	2	2	0	4
Suis	7	3	0	10
AUX-PRP	**5**	**7**	**0**	**12**
Est	2	0	0	2
Sommes	0	1	0	1
Suis	3	6	0	9
INF-ADV	**2**	**0**	**0**	**2**
Commencer	1	0	0	1
Fréquenter	1	0	0	1
NOM-ADJ	**6**	**1**	**0**	**7**
Cultures	1	0	0	1

Langues	4	1	0	5
Relations	1	0	0	1
NOM-ADV	**1**	**0**	**0**	**1**
Jours	1	0	0	1
PPA-ADV	**1**	**0**	**0**	**1**
Fait	1	0	0	1
PRO-INF	**0**	**1**	**0**	**1**
Nous	0	1	0	1
PRO-PRP	**1**	**0**	**0**	**1**
Tout	1	0	0	1
PRP-DET	**51**	**41**	**0**	**92**
Chez	1	2	0	3
Dans	4	17	0	21
Devant	24	10	0	34
Sous	22	12	0	34
PRP-NAM	**0**	**8**	**0**	**8**
En	0	8	0	8
PRP-NOM	**6**	**14**	**1**	**21**
Dans	0	0	1	1
En	0	4	0	4
Sans	6	10	0	16
PRP-PRO	**0**	**1**	**0**	**1**
Chez	0	1	0	1
VER-ADJ	**33**	**29**	**0**	**62**
C'est	3	14	0	17
Est	15	4	0	19
Etait	5	0	0	5
Suis	10	11	0	21
VER-ADV	**13**	**2**	**0**	**15**
Avait	1	0	0	1
Devons	0	1	0	1
Dois	2	1	0	3
Etait	1	0	0	1
Parlent	1	0	0	1
Pouvais	1	0	0	1
Sais	1	0	0	1
Sont	1	0	0	1
Suis	1	0	0	1
Voudrais	4	0	0	4
VER-DET	**14**	**12**	**0**	**26**
Avait	1	0	0	1
Est	7	12	0	19
Etais	1	0	0	1
Etait	3	0	0	3
Sont	1	0	0	1
Suis	1	0	0	1
VER-INF	**27**	**0**	**0**	**27**
Devais	1	0	0	1
Devrais	4	0	0	4

Dois	4	0	0	4
Fait	1	0	0	1
Faut	2	0	0	2
Peux	3	0	0	3
Préférais	1	0	0	1
Serais	2	0	0	2
Voudrais	7	0	0	7
Voulais	2	0	0	2
VER-NOM	**2**	**0**	**0**	**2**
Suis	2	0	0	2
VER-PDE	**2**	**0**	**0**	**2**
Suis	2	0	0	2
VER-PRP	**11**	**1**	**0**	**12**
Est	2	0	0	2
Etais	1	0	0	1
Sommes	2	0	0	2
Suis	5	1	0	6
Vais	1	0	0	1
Total général	**353**	**578**	**12**	**943**

Tableau 95 : Résultats totaux - Réalisations par type de liaison (tâches de parole semi-spontanée)

3.3.3. Nature de la CL réalisée

Les données concernant la nature de la CL en production de parole semi-spontanée montrent la présence de formes orthographiques de la liaison. De plus, celles-ci sont réalisées à 4% dans les contextes catégoriques et à 13% dans les contextes variables. Ces résultats se rapprochent de façon significative des pourcentages de réalisation des productions orthographiques observées en lecture (8% en contexte de liaison catégorique et 20% en contexte variable).

De plus, si nous comparons le taux global de réalisations orthographiques tous contextes confondus, les productions orthographiques sont autant attestées en parole semi-spontanée (9,3%) qu'en lecture (12%) et en répétition (11%) (cf. Tableau 96).

	Parole semi-spontanée				
Type de liaison	**Conforme**	**Orth**	**CL présente dans le M1**	**Substitution**	**Total**
Total CAT	**251**	**12**	**2**	**2**	**267**
DET-ADJ	**6**	**0**	**0**	**0**	**6**
Des	1	0	0	0	1
Les	3	0	0	0	3
Un	2	0	0	0	2
DET-NOM	**152**	**11**	**0**	**2**	**165**

Cet	21	0	0	2	23
Des	46	4	0	0	50
Les	36	4	0	0	40
Mes	24	3	0	0	27
Mon	1	0	0	0	1
Ses	1	0	0	0	1
Un	22	0	0	0	22
DET-PRO	**3**	**0**	**0**	**0**	**3**
Les	3	0	0	0	3
EXF-EXF	**18**	**0**	**0**	**0**	**18**
En	7	0	0	0	7
Plus	7	0	0	0	7
Tout	2	0	0	0	2
Vingt	2	0	0	0	2
NUM-NOM	**25**	**0**	**2**	**0**	**27**
Cents	0	0	1	0	1
Deux	10	0	0	0	10
Quatre-vingts	0	0	1	0	1
Trente-six	1	0	0	0	1
Trois	4	0	0	0	4
Un	1	0	0	0	1
Vingt	9	0	0	0	9
PDE-ADJ	**1**	**0**	**0**	**0**	**1**
Des	1	0	0	0	1
PDE-NOM	**2**	**0**	**0**	**0**	**2**
Aux	2	0	0	0	2
PRO-AUX	**30**	**1**	**0**	**0**	**31**
Ils	2	0	0	0	2
Les	1	0	0	0	1
Nous	23	1	0	0	24
On	4	0	0	0	4
PRO-VER	**14**	**0**	**0**	**0**	**14**
Elles	1	0	0	0	1
Ils	3	0	0	0	3
Nous	5	0	0	0	5
On	5	0	0	0	5
Total VAR	**278**	**43**	**0**	**2**	**323**
ADJ-NOM	**90**	**29**	**0**	**1**	**120**
Ancien	13	0	0	1	14
Bon	1	0	0	0	1
Grand	17	29	0	0	46
Petit	59	0	0	0	59
ADV-ADJ	**49**	**1**	**0**	**1**	**51**
Moins	3	0	0	0	3
Plus	15	0	0	0	15
Toujours	5	0	0	0	5
Très	23	1	0	0	24
Trop	3	0	0	1	4

ADV-ADV	**5**	**1**	**0**	**0**	**6**
Mais	3	0	0	0	3
Pas	1	1	0	0	2
Plus	1	0	0	0	1
ADV-DET	**1**	**0**	**0**	**0**	**1**
Seulement	1	0	0	0	1
ADV-INF	**1**	**0**	**0**	**0**	**1**
Bien	1	0	0	0	1
ADV-NOM	**1**	**0**	**0**	**0**	**1**
Plus	1	0	0	0	1
ADV-PPA	**16**	**0**	**0**	**0**	**16**
Bien	14	0	0	0	14
Jamais	1	0	0	0	1
Pas	1	0	0	0	1
ADV-PRP	**1**	**0**	**0**	**0**	**1**
Pas	1	0	0	0	1
AUX-PPA	**7**	**1**	**0**	**0**	**8**
Sommes	2	1	0	0	3
Sont	2	0	0	0	2
Suis	3	0	0	0	3
AUX-PRP	**7**	**0**	**0**	**0**	**7**
Sommes	1	0	0	0	1
Suis	6	0	0	0	6
NOM-ADJ	**1**	**0**	**0**	**0**	**1**
Langues	1	0	0	0	1
PRO-INF	**1**	**0**	**0**	**0**	**1**
Nous	1	0	0	0	1
PRP-DET	**34**	**7**	**0**	**0**	**41**
Chez	2	0	0	0	2
Dans	12	5	0	0	17
Devant	10	0	0	0	10
Sous	10	2	0	0	12
PRP-NAM	**8**	**0**	**0**	**0**	**8**
En	8	0	0	0	8
PRP-NOM	**12**	**3**	**0**	**0**	**15**
Dans	0	1	0	0	1
En	4	0	0	0	4
Sans	8	2	0	0	10
PRP-PRO	**1**	**0**	**0**	**0**	**1**
Chez	1	0	0	0	1
VER-ADJ	**28**	**1**	**0**	**0**	**29**
C'est	14	0	0	0	14
Est	4	0	0	0	4
Suis	10	1	0	0	11
VER-ADV	**2**	**0**	**0**	**0**	**2**
Devons	1	0	0	0	1
Dois	1	0	0	0	1
VER-DET	**12**	**0**	**0**	**0**	**12**

Est	12	0	0	0	12
VER-PRP	**1**	**0**	**0**	**0**	**1**
Suis	1	0	0	0	1
Total général	**529**	**55**	**2**	**4**	**590**

Tableau 96 : Résultats totaux - Nature de la CL par type de liaison (tâches de parole semi-spontanée)

3.4. Bilan des résultats par tâche

Les tableaux ci-dessous (cf. Tableau 97 ; Tableau 98) permettent une comparaison de la production des liaisons enchaînées réalisées dans les contextes catégoriques et variables dans les trois différentes tâches de production (répétition, lecture et parole semi-spontanée). De manière générale, cette analyse inter-tâche montre l'absence de différences significatives dans la production des liaisons catégoriques (cf. Tableau 97).

Tâche	Enchaînement liaisons catégoriques	Total général	%
Répétition	938	955	98%
Lecture	3288	3974	83%
Parole semi-spontanée	261	285	91,5%

Tableau 97 : Comparaison des taux d'enchaînement des liaisons catégoriques entre les trois tâches de production orale

De plus, si nous avons enregistré des écarts de production de réalisation des liaisons variables entre la tâche de répétition et les autres tâches (cf. Tableau 98), aucune différence significative n'a été observée cette fois encore entre la lecture et la parole semi-spontanée. Ces résultats rejoignent ceux déjà recueillis auprès d'autres sujets débutants et intermédiaires (Galazzi et al. 2013) qui montrent qu'à ce niveau l'apprenant n'est pas encore capable de maîtriser la variation de liaison entre ces deux registres.

Tâche	Enchaînement liaisons variables	Total général	%
Répétition	829	1011	82%
Lecture	1608	4312	37%
Parole semi-spontanée	317	658	48,1%

Tableau 98 : Comparaison des taux d'enchaînement des liaisons variables entre les trois tâches de production orale

De plus, les données issues de l'analyse comparative des réalisations de la liaison par type de production orale (répétition, lecture et parole semi-spontanée) montrent que :

- les résultats de la tâche de répétition indiquent la présence de non-réalisations dans les contextes de liaison variable ;
- les liaisons non-enchaînées apparaissent presque exclusivement dans des contextes de liaison variable alors que les pauses-hésitations sont reparties entre des contextes catégoriques et variables ;
- contrairement aux tâches de répétition et de lecture oralisée, aucune forme de liaison non-enchaînée n'a été enregistrée en parole semi-spontanée (cf. Tableau 99). La faible présence de ces dernières réalisations par rapport à celles observées en lecture semble confirmer une fois de plus que les non-enchaînements relèvent des productions disfluentes influencées par l'input écrit ;

Tâche	Non-enchaînement	Total général	%
Répétition	8	2047	0,4 %
Lecture	19	8834	0,2%
Parole semi-spontanée	0	945	0%

Tableau 99 : Comparaison des taux de non-enchaînement des liaisons totales entre les trois tâches de production orale

- des formes orthographiques de la CL sont attestées dans la tâche de répétition à la fois dans les contextes catégoriques et variables ;
- les résultats issus de la comparaison inter-tâches montrent que les apprenants produisent en parole semi-spontanée un pourcentage de formes orthographiques (9,5%) comparable à celui déjà observé en lecture (12%) et répétition (11%) (cf. Tableau 100) ;

Tâche	Orth	Total général	%
Répétition	224	1954	11,5%
Lecture	586	5093	11,5%
Parole semi-spontanée	55	592	9,3%

Tableau 100 : Comparaison des taux de réalisation des CL orthographiques totales entre les trois tâches de production orale

- contrairement aux contextes variables, il n'existe pas de différences inter-tâches entre le taux de réalisations de la liaison après déterminants (DET-ADJ ; DET-NOM) (cf. Tableau 101) et après les pronoms personnels sujets (PRO-VER ; PRO-AUX) (cf. Tableau 102) ;

DET-ADJ DET-NOM	Enchaînement	Total général	%
Répétition	475	478	99,4%
Lecture	1456	1590	91,6%
Parole semi-spontanée	168	182	92,3%

Tableau 101 : Comparaison des taux d'enchaînement des types DET-ADJ et DET-NOM entre les trois tâches de production orale

PRO-VER PRO-AUX	Enchaînement	Total général	%
Répétition	263	272	96,7%
Lecture	492	568	86,6%
Parole semi-spontanée	43	50	86%

Tableau 102 : Comparaison des taux d'enchaînement des types PRO-VER et PRO-AUX entre les trois tâches de production orale

Chapitre 3. Discussion et aspects d'un itinéraire d'acquisition de la liaison en français langue étrangère

1. Discussion

À la lumière des résultats que nous avons présentés dans les pages précédentes, dans ce chapitre nous nous attacherons tout d'abord à vérifier les hypothèses que nous avons formulées au début de cette étude (cf. Partie 4.Chapitre 1.4.1). Au terme de cette discussion nous souhaiterons avancer des propositions d'itinéraires d'acquisition de la liaison en FLE (cf. Paragraphe 3) en nous appuyant sur les résultats de notre corpus d'apprenants italophones ainsi que sur les données issues de la comparaison avec les autres études IPFC menées auprès de différentes populations d'apprenants (hispanophones, anglophones, japonophones etc.) (cf. Partie 4.Chapitre 1.2.2).

1.1. Hypothèse 1

Les résultats de l'étude longitudinale montrent que les apprenants ne semblent pas maîtriser complétement la production de la liaison catégorique même dans les dernières séances d'enregistrement (S3, S4). Nous pourrions supposer que les apprenants de niveau intermédiaire ne sont pas en mesure d'opérer des généralisations phonologiques quant au fonctionnement de la liaison catégorique puisqu'ils produisent ce type de liaison dans un pourcentage de 90%. Ces données sembleraient ne pas valider l'hypothèse (hypothèse 1) que nous avions formulée dans les chapitres précédents (cf. Partie 4.Chapitre 1.4.1.1.1). Il nous paraît toutefois nécessaire de prendre également en compte certains aspects que nous nous attacherons à decrire dans les pages qui suivent.

1.1.1. Non-réalisation des liaisons catégoriques

Tout d'abord, comme nous l'avons évoqué dans le chapitre précédent (cf. Partie 4.Chapitre 2.1.2.1.2), la production de la liaison après les articles définis et indéfinis et après les pronoms personnels sujets n'est pas sensible à des effets de cooccurrence du M1 et du M2[306]. En d'autres termes, l'apprenant ne produit pas plus de liaisons dans des séquences M1-M2 ayant des indices de cooccurrence plus importants (ex. « des arbres », « nous avons ») que dans

[306] Dans d'autres contextes catégoriques tels que la liaison au sein des expressions figées en revanche, la fréquence de la construction pourrait jouer un rôle dans la production des liaisons chez des apprenants FLE (cf. Partie 4.Chapitre 2.1.2.1.2).

d'autres ne possédant que de faibles indices de cooccurrence (ex. « des âtres », « nous amarrons ») que ce soit en répétition ou en lecture.

En effet, les résultats de l'analyse comparative entre les différents types de productions réalisées par des apprenants (répétition, lecture et parole semi-spontanée) (cf. Partie 4.Chapitre 2.3.4) ont montré que contrairement aux contextes variables, il n'existe pas de différences significatives entre le taux de réalisations de la liaison après déterminants et après pronoms personnels sujets en lecture et en parole semi-spontanée. Ces observations rejoignent les résultats de l'étude menée par Tennant (sous presse) et Falbo (sous presse) qui montrent que la liaison après déterminants et après pronoms personnels sujets est maîtrisée dans toutes les tâches aussi bien par les apprenants les plus avancés (Tennant sous presse ; Falbo et al. sous presse corpus Trieste) que par les moins avancés (Tennant sous presse) et même par les débutants (Falbo et al. sous presse corpus Milan).

Il convient en outre de souligner que la persistance de non-réalisation en contexte catégorique n'est pas uniquement attestée en acquisition L2 mais également en L1 même chez des enfants de 6-7 ans (Ruvoletto 2014 ; Ruvoletto et Wauquier 2015)[307]. Cependant, chez l'apprenant adulte de FLE, contrairement à ce qui se passe chez l'enfant, les marques (à savoir les CL) qui caractérisent les fonctions morphosyntaxiques des liaisons catégoriques n'émergent pas de façon graduelle d'une simple exposition à l'environnement linguistique, mais sont apprises de façon le plus souvent explicite[308] (cf. Partie 3.Chapitre 1.3.2.3). En effet, si l'enfant traite dans un premier stade les déterminants et les noms comme une seule unité (cf. Partie 1.Chapitre 5.3.5), les apprenants adultes L2 acquièrent le plus souvent ces éléments séparément, également sur la base de leurs connaissances en L1.

D'après Arnon et Ramscar (2012), cela amènerait l'apprenant à se concentrer sur les éléments qu'il perçoit comme plus saillants d'un point de vue sémantique et morphosyntaxique, à savoir les noms.

> « Whereas children may have no choice other than to gradually learn to discriminate the various elements of languages, adult L2 learners will have prior experience of another language, and this should enable them to focus on what they perceive to be the ‘‘meaning-carrying’’ units (such as the noun-labels) of a new language». (Arnon et Ramscar 2012, p. 294)

Dans cette perspective, la présence de non-réalisations de liaisons catégoriques attestées même dans les dernières phases de la période d'observation pourrait être expliquée, selon l'hypothèse formulée par Wauquier (2009), comme le résultat d'une difficulté des apprenants

[307] Cette étude montre que des enfants de 6-7 ans produisent encore un taux important d'omission (13,2%) (ex. « les ours » prononcé [leuʁs] au lieu de [lezuʁs]) dans une tâche de dénomination d'image guidée.

[308] Au moins en milieu institutionnel (cf. Partie 3.Chapitre 1.3.2.3).

d'« attacher les mots ensemble », en l'occurrence les déterminants et les noms suivants (Wauquier 2009, p. 125). Or cette analyse implique que les apprenants traitent la liaison « au cas par cas ». Comme nous l'avons déjà observé, cette hypothèse s'accorde mal avec les formes de liaisons non attendues et les substitutions (cf. Partie 4.Chapitre 2.1.5) que nous avons enregistrées dans notre étude longitudinale et transversale (également attestées par des études précédentes cf. Partie 3.Chapitre 2.2).

En effet, l'hypothèse formulée par Wauquier (2009) repose essentiellement sur l'observation d'un taux élevé de liaisons non-enchaînées sans valeur stylistique observées majoritairement chez les apprenants anglophones de niveau avancé[309] (cf. Partie 3.Chapitre 2.4). Comme nous l'avons vu dans les paragraphes précédents ces formes de liaison non-enchaînées sont très rares dans notre corpus et le plus souvent produites en lecture.

Bien que nous reconnaissions une difficulté des apprenants de français L2 à « attacher » les mots au sein du groupe accentuel, nous ne le considérons pas comme un argument en faveur d'un traitement exclusivement lexical de l'acquisition de la liaison en L2 (cf. Partie 3.Chapitre 2.3).

1.1.1.1 La liaison comme marque redondante dans l'acquisition du français langue étrangère (FLE)?

Dans le paragraphe précédent, nous avons souligné que les apprenants de FLE, contrairement aux natifs, traitent le déterminant et le nom comme des unités distinctes en focalisant leur attention sur la catégorie principale à savoir le nom. À partir de cette observation, une recherche psycholinguistique récente menée par Arnon et Ramscar (2012) s'interroge sur la gradualité de l'acquisition du genre dans une langue étrangère, aspect donnant souvent lieu à des erreurs même à des niveaux d'apprentissage avancés. Ces auteurs ont mené une étude de l'acquisition d'une langue artificielle auprès de deux groupes de sujets adultes. Les résultats de cette recherche ont montré que le groupe ayant été exposé dès le début de l'acquisition à des énoncés contenant à la fois le déterminant et le nom (et ensuite au seul nom) apprenaient mieux le genre que ceux ayant appris dans un premier temps les noms de façon isolée (et ensuite l'énoncé entier). Ces données semblent suggérer que l'acquisition d'unités restreintes et séparées du reste de l'énoncé, avant l'exposition à des unités plus larges, risque de bloquer l'apprentissage des relations existantes entre ces unités en les rendant redondantes.

[309] « […] la prégnance de liaisons non enchaînées sans valeur stylistique particulière en contexte obligatoire invite à considérer que ces apprenants recherchent dans l'input oral les mots segmentés qu'ils ont appris par l'écrit ». (Wauquier 2009, p. 122)

> « […] prior learning can block (Kamin, 1969) the learning of later information by rendering it redundant. Thus learning noun-labels before article - noun sequences might affect the learning of articles if a learner's knowledge about the relationship between a noun and its meaning marginalizes the informative contribution of an article. For example, if an English speaker learning the French word for cat has already learned to fully expect a label (''chat'') given a semantic context (chat/cat), then there will be little room for a definite article (''le'') to add new information to this relationship: what the learner already knows will interfere with learning the article ». (Arnon et Ramscar 2012, p. 294)

Or, bien que la liaison soit apprise le plus souvent à partir des suites M1-M2 (ex. « les amis ») insérées dans un énoncé ou présentées de façon isolée, la plupart des syllabus pour l'enseignement du FLE font précéder le plus souvent l'enseignement de la liaison par l'acquisition de la morphologie des différentes catégories morphosyntaxiques (entre lesquelles la liaison est réalisée).

Ce fait semble confirmé par les recherches menées sur les itinéraires acquisitionnels du français qui montrent qu'aux stades initiaux de l'acquisition l'apprenant se sert d'une structuration en grande partie nominale mais aussi des moyens grammaticaux (des articles définis et indéfinis ainsi que des pronoms sujets) qui cependant sont « très souvent accentués et non élidés (Granfeldt et Schlyter, 2004) » (Bartning et Schlyter 2004, p. 294). Ces données semblent étayer l'hypothèse d'un traitement lexical des éléments dans lequel, en dépit de la présence d'unités fonctionnelles, les échanges d'information au sein du groupe syntaxique ne sont pas encore activés (Pienemann 1998) (cf. Partie 3.Chapitre 1.4.3.1.5). À ce stade donc, l'apprenant est capable d'attribuer aux lexèmes des catégories syntaxiques spécifiques en utilisant les morphèmes grammaticaux tels que les désinences du genre et du nombre. Les informations morphosyntaxiques sont donc encodées au niveau des lexèmes (« procédure de catégorie » *category procedure*) sans pour autant que des échanges entre les différentes unités ne soient activées (« procédure intra-syntagmatique » *phrasal procedure*) (Pienemann 1998). À la suite des résultats des travaux menés par Arnon et Ramscar (2012), nous pouvons avancer l'hypothèse que l'acquisition préalable des marques morphosyntaxiques au niveau du lexème et donc des catégories principales est en mesure de ralentir l'acquisition de la liaison et de la rendre redondante. Autrement dit, le passage d'un marquage au niveau catégoriel (« procédure de catégorie » *category procedure*) à des accords entre constituants d'un syntagme (« procédure intra-syntagmatique » *phrasal procedure*), ou encore entre syntagmes (« procédure inter-syntagmatique » « *S-procedure* »), pourrait être ralenti par le fait que les informations morphosyntaxiques sont déjà encodées sous formes de désinences au niveau du lexème, ce qui rend la production de la liaison redondante d'un point de vue fonctionnel. Cet effet pourrait être

amplifié en contexte hétéroglotte et institutionnel par une faible exposition à l'input oral ainsi que par l'absence d'un réel besoin communicatif[310] (cf. Partie 3.Chapitre 1.3.2.2).

Cependant, comme nous le verrons dans le paragraphe suivant, plusieurs éléments semblent suggérer qu'au cours du processus d'acquisition l'apprenant est en mesure de développer des représentations phonologiques de la liaison sous forme de constructions.

1.1.2. Vers des représentations phonologiques de la liaison

Comme nous l'avons souligné dans le paragraphe précédent, l'acquisition préalable des marques morphosyntaxiques au niveau du lexème pourrait ralentir les échanges entres constituants d'un syntagme (« procédure intra-syntagmatique » *phrasal procedure*) (Pienemann 1998) et ainsi favoriser un traitement lexical de la liaison dans les premières étapes de l'apprentissage.

La nature des erreurs attestées par la présente étude ainsi que par les travaux précédents semble nous suggérer que la CL n'est pas encodée dans la représentation phonologique du M1 chez les apprenants (ex. « des » /de/) (cf. hypothèse abstractionniste Partie 1.Chapitre 6.2.1.1). En effet si la représentation phonologique du M1 contenait la CL (ex. /dez/), nous aurions dû observer une présence massive de liaisons non-attendues devant des mots à initiale consonantique (ex. « des garçons » [dezgaʀsɔ̃]) (Mastromonaco 1999) (cf. Partie 3.Chapitre 2.1.4) comme résultat d'une généralisation morpho-phonologique de la liaison.

En revanche, comme nous l'avons montré dans le chapitre précédent (cf. Partie 4.Chapitre 2.1.5), ces formes sont très rares alors que les non-réalisations dans des contextes catégoriques et des prononciations orthographiques de la CL représentent les erreurs les plus attestées par toutes les études (cf. Partie 3.Chapitre 2.1.4). Ces données semblent apporter des arguments en faveur d'une représentation phonologique des mots de liaison qui ne contient pas la CL (ex. /de/).

Comme nous le verrons plus en détail dans les prochains paragraphes (cf. Paragraphe 1.3.1), nous faisons l'hypothèse que dans les premières phases de l'acquisition, la production de la liaison catégorique serait majoritairement influencée par la représentation orthographique (ex. < des >), ce qui justifierait la présence de réalisations orthographiques (ex. « des arbres » [desaʀbʀ]) à côté des non-réalisations des liaisons catégoriques (ex. « des arbres » [deaʀbʀ]).

[310] Le processus inverse est en revanche observé en acquisition L1, dans lequel la liaison recouvre dès le début de l'acquisition un rôle fonctionnel important puisque les enfants mémorisent des CL encodées au début du M2. C'est seulement au cours des étapes successives qu'ils parviennent à détacher la CL et à la ré-analyser.

Ces deux types d'erreurs se réduisent au cours de l'acquisition comme le démontrent les résultats de notre étude longitudinale (cf. Partie 4.Chapitre 2.1.1). En revanche des formes de liaisons non attendues et de substitutions émergent (même si en faible mesure) en particulier à partir de la deuxième séance de la période d'observation. Les premières, comme nous l'avons déjà évoqué, apparaissent le plus souvent en contexte de liaison catégorique et notamment dans les contextes DET-NOM, NUM-NOM, PRO-AUX et PRO-VER. Bien que le taux de production de ces formes soit très faible, ces réalisations de liaison non attendue nous indiquent que les apprenants opèrent des généralisations de schémas de liaison tels que « DET - z - NOM » (ex. « les [z] halls » au lieu de « le hall »), NUM - z - X (ex. « cinq [z] ans », « quatrième [z] aux ») ou encore « PRO - z - VER au pluriel » (ex. « qui [z] ont » ; « y [z] entrer »).

Quant à la nature de la réalisation de la CL, comme nous l'avons souligné, des formes de substitution des CL, souvent attestées en acquisition L1 (cf. Partie 1.Chapitre 5.3.5), apparaissent graduellement tout le long des quatre sessions d'enregistrement dans notre corpus. Parmi ces formes (cf. Partie 4.Chapitre 2.1.3.4), la plupart correspondent à des remplacements des CL /n/ et /t/ par une CL /z/ (71%). Celles-ci sont encore plus intéressantes si nous considérons qu'elles sont produites presque exclusivement en lecture oralisée, lorsque la présence d'un support écrit devrait assurer une plus grande conformité des réalisations à la CL cible. Certaines parmi ces productions semblent suivre les mêmes schémas de généralisation observés dans les cas des liaisons non attendues (cf. Partie 4.Chapitre 2.1.5), à savoir « NUM - z – X » (ex. « trente-huit [z] ans ») ou encore « PRO - z -VER au pluriel » (ex. « nous en [z] installions »). D'autres peuvent être expliquées comme des formes d'erreur à distance (cf. Partie 1.Chapitre 4.2.4.3) (ex. « ils ont [z] aussi », « ils ont [z] acheté »), d'autres encore restent difficiles à interpréter (ex. « grand [z] enclos », « petit [z] enfant » ou « il est [z] impossible »). Ces dernières pourraient être analysées comme le résultat d'une surgénéralisation de la CL /z/ perçue sans doute par l'apprenant comme plus fréquente et fonctionnelle (cf. Partie 1.Chapitre 2.1) et donc produite dans la plupart des substitutions.

En résumé, malgré la persistance de non-réalisations dans les contextes catégoriques au cours des sessions finales de la période d'observation (tendance également relevée en acquisition L1) (cf. Ruvoletto 2014 ; Ruvoletto et Wauquier 2015), les formes de liaison non attendues ainsi que des cas de substitutions de la CL indiquent que les apprenants lors des dernières phases de la période d'observation commencent à mettre en place des généralisations d'ordre phonologique, tout au moins dans les types de liaison catégorique DET-NOM et PRO-VER. Ces généralisations dépassent les simples stratégies d'acquisition lexicale ou de traitement « au cas par cas » des contextes de liaison (Wauquier 2009). Une acquisition de ce

type présupposerait en effet que dans les contextes DET-NOM et PRO-VER les apprenants produisent un taux plus important de liaison dans les cas des séquences M1-M2 ayant des indices de cooccurrence plus importants et en présence d'un support écrit. En revanche, nous n'avons observé aucun effet significatif de la fréquence de cooccurrence ou de tâche dans les cas de réalisations des contextes DET-NOM et PRO-VER.

Arrivés à ce stade, sommes-nous en droit de nous interroger sur la nature des facteurs capables de déclencher ces généralisations ? Un élément de réponse à cette question pourrait nous venir des théories de probabilité transitionnelle (Coté 2013). Dans cette perspective, nous pouvons supposer que les apprenants développent des généralisations en s'appuyant sur l'observation des fortes probabilités transitionnelles qu'un déterminant soit suivi par un nom qui aboutit à la mise en place de représentations sous forme de constructions telles que « les - z - NOM » etc.

Malgré le développement de généralisations décrites ci-dessus, des non-réalisations continuent d'être observées au cours des dernières séances de notre étude longitudinale ainsi que dans les productions d'apprenants avancés attestées par les travaux précédents (cf. Partie 3.Chapitre 2.2). La persistance de ces non-réalisations pourrait à notre avis être interprétée comme un indice du faible rôle fonctionnel et informatif que la liaison recouvre pendant les premières phases d'acquisition en FLE (cf. Paragraphe 1.1.1.1).

Il convient en outre de souligner que tous les contextes catégoriques ne font pas l'objet de généralisations, comme le prouve la présence de non-réalisations à l'intérieur des expressions figées, ou encore le taux important de non-réalisations après les pronoms clitiques. Si les premières peuvent être facilement expliquées comme le résultat d'un manque d'exposition à l'input oral en milieu hétéroglotte, les deuxièmes demandent une étude supplémentaire.

1.1.3. L'acquisition de la liaison après les pronoms clitiques

Bien que les clitiques possèdent un fort degré de cohésion syntaxique et prosodique avec les constituants verbaux dont ils dépendent, les taux de réalisations associés à ces productions sont assez faibles. Les études menées sur les itinéraires acquisitionnels du FLE (Bartning et Schlyter 2004) pourraient apporter des éléments de réponse aux résultats des réalisations de liaison les pronoms clitiques qui semblent nier l'influence de la cohésion syntaxique dans l'acquisition de la liaison en FLE.

Comme nous l'avons évoqué dans la troisième partie de ce travail (cf. Partie 3.Chapitre 1.4.3.1.4), les travaux de Bartning et Schlyter (2004) ont permis de mettre en évidence qu'au

sein de l'itinéraire acquisitionnel des pronoms clitiques apparaissent assez tardivement dans l'interlangue de l'apprenant de FLE. Après une première phase de non-réalisation des pronoms objet ou de remplacement par des syntagmes nominaux, l'apprenant commence à placer ces formes après le verbe. C'est seulement à des stades plus avancés qu'a lieu l'antéposition (en particulier devant auxiliaire) et que l'apprenant acquiert donc le contexte morphosyntaxique adéquat pour la réalisation de liaison entre pronom clitique et verbe suivant.

Le taux élevé de non-réalisations associé aux pronoms clitiques peut donc s'expliquer par l'acquisition tardive de ces éléments pour lesquels aucun schéma de réalisation de la liaison n'est accessible chez l'apprenant de niveau intermédiaire. Cette hypothèse semble être étayée par les formes de substitution observées après les pronoms clitiques « nous en [z] installions » résultant soit de l'application par analogie du schéma « PRO - z - VERpluriel » soit d'erreurs à distance (cf. Partie 1.Chapitre 4.2.4).

À la suite des travaux menés sur les itinéraires d'acquisition en FLE, en nous appuyant sur les résultats de notre corpus ainsi que sur les données issues des autres points d'enquête IPFC (cf. Partie 4.Chapitre 1.2.2), nous présenterons dans les pages qui suivent (cf. Paragraphe 3) un possible itinéraire d'acquisition de la liaison en FLE.

1.2. Hypothèse 2

Nous rappelons que dans notre deuxième hypothèse (hypothèse 2) (cf. Partie 4.Chapitre 1.4.1.2.1), à la suite des recherches précédentes (Racine sous presse ; Tennant sous presse ; Andreassen et Lyche sous presse), nous avons suggéré de considérer les non-enchaînements comme des formes qui relèveraient de l'influence de l'input orthographique dans la lecture et non pas de l'influence du système syllabique de la L1 des apprenants (Thomas 2002; Wauquier 2009; Tennant sous presse) (cf. Partie 3.Chapitre 1.3.1.1). C'est pour cette raison que nous nous attendions à observer la présence d'un écart important entre le taux des liaisons non-enchaînées réalisées dans des tâches de lecture et des tâches de parole semi-spontanée

Les résultats de notre étude longitudinale et transversale semblent étayer cette hypothèse. En effet, nous pouvons observer que, contrairement aux tâches de répétition et de lecture oralisée, aucune forme de liaisons non-enchaînées n'a été enregistrée en parole semi-spontanée. Ces données rejoignent les résultats des études précédentes qui montrent une plus grande présence de liaisons non-enchaînées en lecture qu'en parole semi-spontanée chez des apprenants ayant des L1 typologiquement différentes telles que l'anglais (Tennant sous presse), l'espagnol (Racine sous presse) et le norvégien (Andressen et Lyche sous presse).

L'absence de formes de liaisons non-enchaînées en parole semi-spontanée semble aller dans la direction des conclusions déjà formulées en L1 (Durand et al. 2011) ainsi que par certaines recherches en L2 (Racine sous presse ; Tennant sous presse ; Andreassen et Lyche sous presse), à savoir que les non-enchaînements constitueraient un phénomène marginal dans la production de la liaison et relèveraient de l'influence de l'input orthographique dans la lecture et non pas de l'influence du système syllabique de la L1 des apprenants.

En effet, le faible nombre de liaisons non-enchaînées attesté chez les apprenants italophones (et presque uniquement dans des types de liaison variable) ne permet pas de valider l'hypothèse formulée par Thomas (2002), Wauquier (2009) et Tennant (sous presse)[311], mais il nous indique simplement que les apprenants italophones ne rencontrent pas de difficulté à resyllaber la CL. Il convient en effet de rappeler que d'autres recherches (Racine sous presse ; Falbo sous presse) relèvent la présence des productions de liaisons non-enchaînées également chez des apprenants ayant pour L1 des langues romanes (espagnol et italien) caractérisées par un système syllabique ouvert. À la lumière de ces observations, l'hypothèse de l'influence de la L1 sur l'acquisition de la liaison (Thomas 2002 ; Wauquier 2009 ; Tennant sous presse) ne peut pas être vérifiée.

1.3. Hypothèse 3

L'analyse de l'influence de l'orthographe sur les réalisations de la liaison est au centre de notre troisième hypothèse (hypothèse 3) (cf. Partie 4.Chapitre 1.4.1.3.1). Pour être plus précis, nous nous sommes proposé de vérifier le niveau d'interaction entre orthographe et phonologie dans la production de la liaison chez les apprenants de FLE.

Les résultats de notre étude longitudinale ont montré que les réalisations orthographiques des CL sont attestées aussi bien en répétition qu'en lecture ou en parole semi-spontanée. Ce fait est confirmé par les résultats de la comparaison menée entre les différentes tâches de production orale (cf. Partie 4.Chapitre 2.3). En effet, contrairement à nos prévisions initiales, les apprenants ne réalisent pas plus de formes orthographiques dans les tâches de lecture oralisée que dans les tâches de répétition ou de parole semi-spontanée.

De plus, les résultats de la tâche de reformulation montrent que les apprenants exposés à un input écrit ne produisent pas plus de formes orthographiques que les apprenants ayant été soumis à un input uniquement oral. L'absence d'écart entre les formes orthographiques réalisées par les deux groupes d'apprenants semble indiquer que l'exposition à un input écrit ne

[311] À savoir que les apprenants ayant une L1 caractérisée par un système syllabique ouvert produisent moins de liaisons non-enchaînées que d'autres apprenants.

produit pas un effet linguistique immédiat (Chevrot 2015) sur les productions des réalisations orthographiques des CL.

De plus, les résultats du test de jugement montrent que des réalisations orthographiques sont également sélectionnées par les apprenants en tant que seules réponses jugées comme correctes (à savoir les réponses *error,* « erreur », réponse 1 et 1bis) (cf. Partie 4.Chapitre 2.1.6) ou en combinaison avec d'autres variantes (variantes avec non-réalisation ou variante correspondant à la CL cible). En effet, comme nous l'avons vu dans le chapitre précédent, nous avons enregistré des réponses où les apprenants sélectionnaient à la fois des variantes de liaison conforme à la CL cible et des variantes orthographiques non seulement en contexte variable (à savoir les réponses *invalid* « nul », réponse 2bis) (cf. Partie 4.Chapitre 2.1.6) mais également catégorique (à savoir les réponses *invalid plus*, « nul plus », réponse 2) (cf. Partie 4.Chapitre 2.1.6). Ces résultats sont encore plus intéressants si nous considérons que les apprenants sont conscients de la nature obligatoire des contextes catégoriques, comme le montre l'absence presque totale de sélection de la variante avec non-réalisation (à savoir les réponses *error plus* « erreur plus », réponse 6) (cf. Partie 4.Chapitre 2.1.6). De plus, ces typologies d'erreurs comportant le choix d'une variante orthographique semblent se réduire progressivement au cours des dernières séances de la période d'observation en faveur d'une augmentation des sélections de la seule variante conforme à la CL cible. Comment alors justifier la présence de choix qui combinent les variantes des CL conformes à la cible et les réalisations orthographiques (à savoir les réponses *invalid plus*, «nul plus » réponse 2) (cf. Partie 4.Chapitre 2.1.6) dans des contextes que les apprenants traitent comme catégoriques? Il serait possible d'expliquer ces résultats en reconnaissant que la forme orthographique des mots conditionne de façon profonde les représentations phonologiques de la liaison, tout au moins au cours des premières phases de l'étude longitudinale.

L'influence de l'orthographe semblerait donc avoir un effet durable (Chevrot 2015) au niveau des connaissances épilinguistiques (Gombert 1990) et des représentations phonologiques, ce qui expliquerait la présence de productions orthographiques dans les trois types de production orale ainsi que les choix opérés dans le test de jugement.

1.3.1. Interactions entre représentations phonologiques et orthographiques

Comme nous l'avons déjà évoqué (cf. Paragraphe 1.1), nous rejetons l'hypothèse selon laquelle les apprenants ne sont pas en mesure de développer des représentations phonologiques de la CL. La présence de productions de liaisons orthographiques, particulièrement observées

au cours des étapes initiales de la période d'observation, laisse cependant penser qu'à ce stade, la représentation phonologique du mot de liaison (qui ne contient pas la CL ex. /de/ cf. Paragraphe 1.1.2) est encore profondément influencée par la représentation orthographique. Afin de rendre compte de cette influence, nous adaptons ci-dessous le schéma proposé par le modèle interactif bimodale (Grainger et Ferrand 1996) (cf. Figure 34).

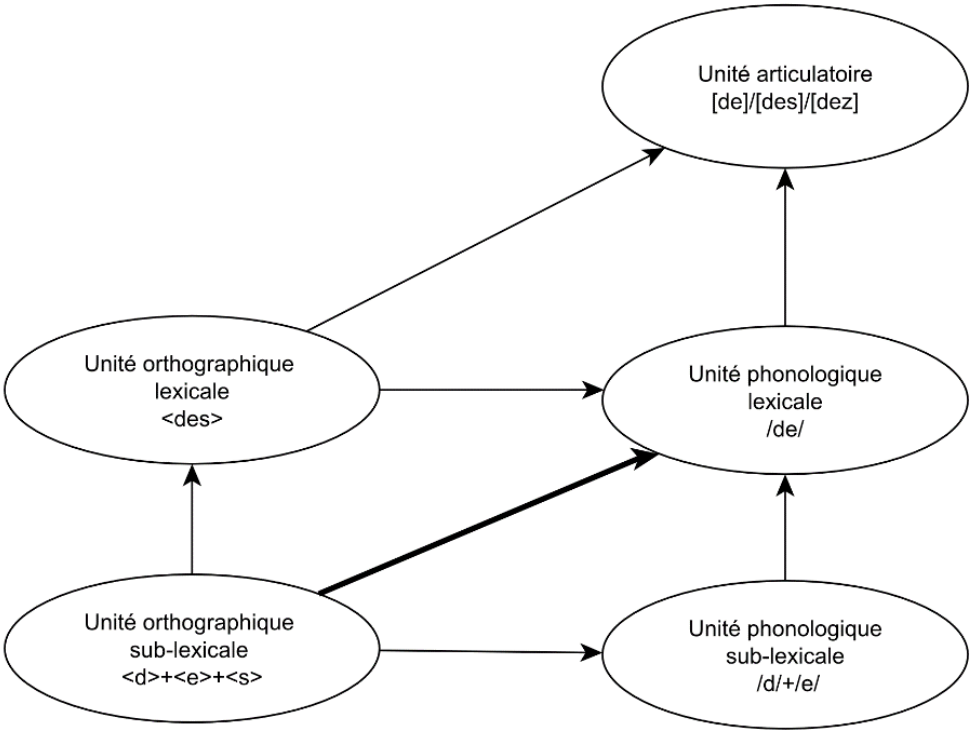


Figure 34 : Interaction entre représentations phonologiques et orthographiques des mots de liaison (Grainger et Ferrand 1996) dans la production des réalisations orthographiques de la liaison chez les apprenants FLE

Les interactions entre représentations orthographiques et phonologiques permettent de rendre compte de la présence de réalisations orthographiques dans des tâches de production orale et de perception ainsi que du choix des variantes orthographiques dans le test de jugement. Cela s'expliquerait encore une fois par le fait qu'au début de leur acquisition, les apprenants FLE débutants ou faux-débutants (A2-B1) ont encodé les informations morphosyntaxiques au niveau des lexèmes (« procédure de catégorie » *category procedure*).

C'est seulement à l'étape suivante que des échanges entre les différentes unités commencent à être mises en place entres constituants d'un syntagme (« procédure intra-syntagmatique » *phrasal procedure*) ou entre syntagmes différents (« procédure inter-syntagmatique » *S-procedure*).

Au début de cette phase, nous pouvons supposer que les apprenants commencent à produire la liaison en s'appuyant sur les seules représentations phonologiques (ex. /de/)[312] et orthographiques (ex. < des >) des unités dont ils disposent. Autrement dit, lors des premières phases du développement des échanges syntaxiques (« procédure de catégorie » *category procedure*), la liaison peut être envisagée comme le résultat de l'influence des représentations orthographiques sur les représentations phonologiques qui, dans le cas de non correspondance entre consonnes graphiques et CL, donnent lieu à des réalisations orthographiques de la CL (ex. « des enfants » [desɑ̃fɑ̃] au lieu de [dezɑ̃fɑ̃]) tant en perception qu'en production.

À ce stade donc la CL est encore traitée comme une consonne finale du M1, présente dans la représentation orthographique du lexème dont l'activation permettrait de produire la liaison de façon relativement mécanique. La liaison, comme nous l'avons observé dans les pages précédentes (cf. Paragraphe 1.1.1.1), est dans un premier temps redondante et pauvre d'un point de vue fonctionnel.

Cette analyse que nous venons de décrire correspond au traitement lexical décrit par Wauquier, où l'apprenant essaie d' « attacher » les mots les uns aux autres. Si la production de la liaison en acquisition L2 s'arrêtait à ce processus mécanique, nous devrions continuer à observer uniquement des formes orthographiques tout au long de la période d'observation.

En revanche, nous avons montré comment ces réalisations diminuent progressivement au cours des dernières séances de la période d'observation, en particulier dans les contextes catégoriques. Nous avons suggéré dans le paragraphe 1.1.2 que cette réduction pourrait être la conséquence d'une mise en place de généralisations phonologiques de la part de l'apprenant (cf. Figure 35).

[312] Il convient de souligner que l'apprenant est conscient de la forme phonologique des mots en contexte de non liaison ex /de/ puisqu'il est capable de produire la forme correcte de l'item devant pause ou devant mot commençant par consonne.

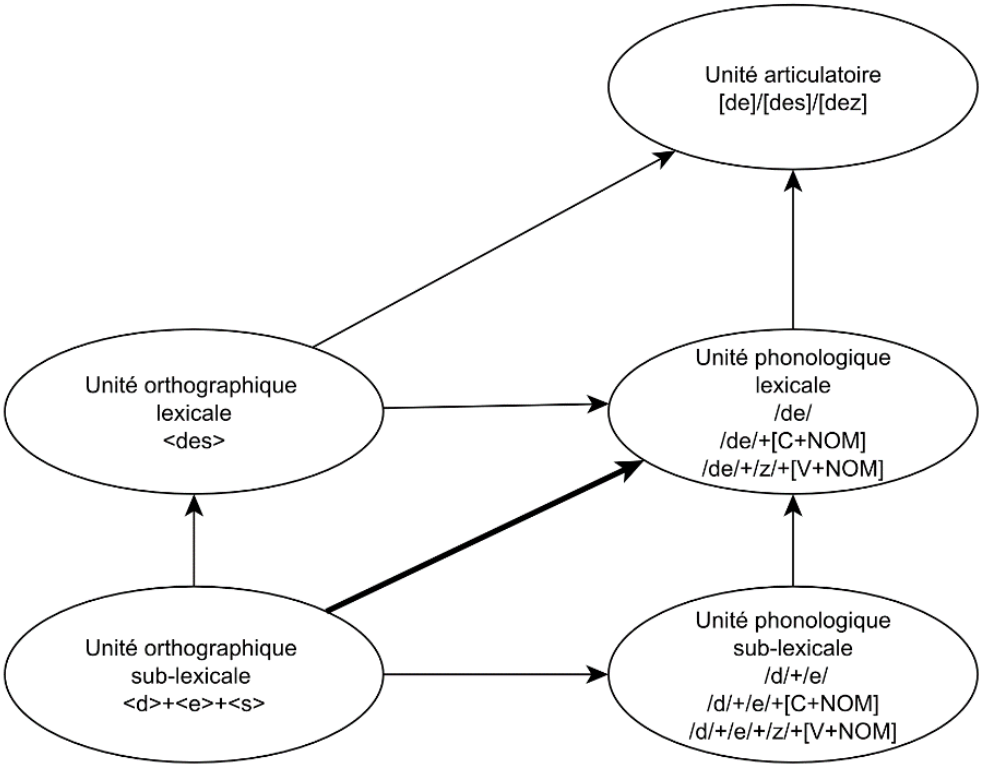


Figure 35 : Développement des généralisations phonologiques de la CL chez les apprenants de FLE. Persistance interactions entre représentations phonologiques et orthographiques des mots de liaison (Grainger et Ferrand 1996)

Les fortes probabilités transitionnelles (Côté 2013) des déterminants d'être suivis par un nom amèneraient l'apprenant à mettre en place des généralisations d'ordre phonologique sous formes de constructions et à s'éloigner progressivement des représentations orthographiques des M1 apprises lors des premières phases de l'acquisition (*category procedure*) (cf. Partie 3.Chapitre 1.4.3.1.5.).

1.4. Hypothèse 4

Quant à l'hypothèse formulée à propos de l'acquisition de la variabilité sociolinguistique, les résultats de l'étude longitudinale et transversale montrent que les apprenants de niveau intermédiaire ne maîtrisent pas la variabilité sociolinguistique de la liaison, ce qui rejoint les résultats recueillis auprès d'autres sujets débutants et intermédiaires (Galazzi et al. 2013 ; Falbo et al. sous presse). L'étude comparative des réalisations des liaisons variables issues des lectures et de la parole semi-spontanée indique l'absence de différences significatives inter-tâches. Autrement dit comme le soulignent Galazzi et al. (2013) lorsque les apprenants « passent d'une

situation communicative à une autre, qui demanderait un registre plus soutenu, ils ne se conforment pas à cette exigence ».

De plus, contrairement à nos prévisions initiales, les résultats de la présente étude montrent que les connaissances épilinguistiques de la variabilité sociolinguistique ne sont pas plus développées que la variabilité attestée en production. En effet, les apprenants font également état de faibles connaissances épilinguistiques (Gombert 1990) de cette variation.

L'apprenant, au lieu de sélectionner les deux variantes correctes (avec et sans réalisation) dans le contexte de liaison variable, il choisit comme seule forme valable celle avec réalisation de la liaison (indépendamment de la nature de la CL réalisée) (cf. réponse *error* « erreur », réponse 1bis ; réponse *missing* « manqué », réponse 5) (cf. Partie 4.Chapitre 2.1.6).

Les résultats de l'étude des productions des liaisons variables font également état de taux de réalisation plus élevés chez les apprenants que chez les natifs (cf. Chapitre 2.1.2.1.3).

Ces données rejoignent les études précédentes (Mougeon, Nadasdi et Rehner 2002 ; Gadet et Tyne 2007) (cf. Partie 3.Chapitre 3.1.3) qui ont montré que les apprenants en milieu institutionnel sur-généralisent les variantes qu'ils jugent être les plus standards (dans notre cas, la variante avec réalisation quelle que soit la nature de la CL). L'apprenant en contexte hétéroglotte et guidé est en effet exposé les plus souvent à des variantes issues de styles relativement formels tels que les dialogues extraits des films ou des lectures de textes journalistiques, etc. Cette hypothèse est étayée par les résultats de l'étude menée par Racine (sous presse) et par Detey, Kawaguchi et Kondo (sous presse) qui montrent que de manière générale seuls les apprenants ayant effectué un séjour (AS) dans un pays francophone produisent un nombre inférieur de liaisons par rapport aux natifs et ce, uniquement dans la tâche de conversation. Racine (sous presse) souligne en effet que, comme dans le cas d'autres aspects sociolinguistiques (ex. l'adverbe de négation « ne »), « le séjour permet aux apprenants d′acquérir des formes moins formelles qu′ils ont tendance à sur-généraliser en conversation, alors que le contexte ″classe de FLE″ favorise les formes formelles »[313].

2. Entre stades post-initiaux et stades intermédiaires

L'acquisition de la liaison semble soumise à des contraintes d'ordre acquisitionnel comme nous l'avons vu lors de l'analyse des productions de la liaison associée au clitiques chez

[313] Ces observations permettraient d'expliquer également les cas où les apprenants sélectionnent des réponses correspondant au choix des variantes conformes à la CL à côté des variantes orthographiques (à savoir les réponses invalid, « nul », réponse 2bis) (cf. Partie 4.Chapitre 2.1.6).

les apprenants italophones de niveau A2-B1. L'étude des productions de liaisons réalisées auprès de ces sujets semblerait se rapprocher de certains éléments issus des stades post-initiaux et intermédiaires identifiés par Bartning et Schlyter (2004) dans l'acquisition du FLE. De manière approximative, nous pouvons donc comparer les deux premières séances de notre étude longitudinale (S1, S2 correspondant à la fin de la première année de licence et au tout début de la deuxième année) aux stades post-initiaux décrits par Bartning et Schlyter (2004).

À ce stade, nous avons observé que les échanges intra-syntagmatiques sont déjà activés, bien que la production de la liaison chez l'apprenant italophone soit encore profondément influencée par une représentation orthographique du mot fixée lors des premières phases de l'acquisition («procédure catégorique » *category procedure*).

En revanche, les deux dernières séances de la période d'observation (S3, S4 correspondant à la deuxième année de licence) pourraient être comparées aux stades intermédiaires, où l'apprenant possède une interlangue relativement systématique souvent caractérisée par des « sur-extensions et des régularisations qui font qu'elle n'est pas toujours en accord avec les normes de la langue cible » (Bartning et Schlyter 2004, p. 295). À ce stade, comme nous l'avons déjà observé, la position du pronom personnel objet n'est pas encore maîtrisée et le pronom lui-même est souvent placé de façon incorrecte. La forme verbale « Nous V-ons » est marquée la plupart du temps. C'est justement après ce pronom sujet que nous avons enregistré l'un des taux les plus importants de réalisation de la liaison en contexte catégorique. En revanche, Bartning et Schlyter (2004) soulignent que l'opposition entre les troisièmes personnes du pluriel « ont » et « sont », bien qu'elle soit plus constante (environ la moitié des cas), n'est pas encore entièrement maîtrisée. En effet, nos résultats font état d'un nombre plus important de non-réalisations après le pronom « ils » qu'après les autres pronoms personnels sujets tels que « nous » et « vous ».

En reprenant la caractérisation des stades établie par Pienemann (1998) et celle définie par Bartning et Schlyter (2004., p. 293‑296), nous nous attacherons dans les pages qui suivent à esquisser le développement d'un itinéraire acquisitionnel de la liaison catégorique chez des apprenants de FLE.

3. Itinéraire d'acquisition de la liaison catégorique en français langue étrangère (FLE)

Le tableau ci-dessous (cf. Tableau 103) illustre l'itinéraire d'acquisition de la liaison catégorique en milieu hétéroglotte, basé à la fois sur les résultats de notre étude longitudinale

menée auprès d'apprenants italophones et les données issues de la comparaison avec les autres recherches réalisées dans le cadre du projet IPFC (cf. Partie 4.Chapitre 1.2.2). Cet itinéraire se situe dans le cadre de la description des stades d'acquisition établie par Pienemann (1998) et par Bartning et Schlyter (2004., p. 293‑296).

Stade acquisitionnel (Bartning et Schlyter 2004)	**Procédure de processabilité (Pienemann 1998)**	**Domaine structural**	**Echange d'information/ Marquage d'accord**	**Itinéraire acquisitionnel des liaisons catégoriques**
Stades avancés	5 Proc. de phrase subordonnée (*Subordinate clause procedure*)	Phrase principale et phrase subordonnée	Accord entre la phrase principale et la phrase subordonnée	Maîtrise presque totale des liaisons catégoriques (« zone de fragilités »)
Stade intermédiaire	4 Proc. Inter-syntagmatique (*S-procedure*)	Phrase simple	Accord entre syntagmes	- Réduction des non-réalisations et des réalisations orthographiques - Persistance des non-réalisations (expressions figées et après pronoms clitiques) - Augmentation des liaisons non attendues et substitutions des CL
Stade post-initial	3 Proc. Intra-syntagmatique (*Phrasal procedures*)	Syntagme	Accord entre les constituants d'un syntagme	Taux élevé de non-réalisations et des réalisations orthographiques
Stade initial	2 Proc. de catégorie (*Category procedure*)	Morphèmes lexicaux	Attribution du genre et du nombre	Faible taux de réalisations
	1 mot/lemme (*Lemma acces*)	Lexème	-	-

Tableau 103 : Itinéraire d'acquisition de la liaison catégorique en FLE

3.1. Stade initial - procédure par lemme et par catégorie

Dans un stade initial de l'acquisition, l'interlangue de l'apprenant est encore basée sur une structuration majoritairement nominale mais elle fait également appel à l'article défini et indéfini ainsi qu'à des pronoms personnels sujets. À ce stade donc, les informations morphosyntaxiques sont encodées au niveau des lexèmes (« procédure de catégorie » *category*

procedure) sans pour autant que des échanges entre les différents unités soient activées (« procédure intra-syntagmatique » *phrasal procedure*).

Sur la base de ces descriptions, nous supposons qu'à ce stade la liaison catégorique après des déterminants et des pronoms personnels est encore peu développée.

3.2. Stade post-initial - procédure intra-syntagmatique

Au stade suivant, le stade post-initial, des échanges entre les différents composants d'un même syntagme commencent à se mettre en place (« procédure intra-syntagmatique » *phrasal procedure*).

L'apprenant commence à acquérir l'accord sujet-verbe, en particulier après le pronom « nous ». À ce moment de l'apprentissage toutefois les pronoms objets sont encore généralement postposés ce qui exclut la possibilité pour l'apprenant de réaliser la liaison entre les pronoms clitiques et les verbes, malgré la nature catégorique de cette liaison. De plus, Bartning et Schlyter (2004) soulignent qu'à ce stade l'apprenant fait souvent appel à « c'est » utilisé comme « constructeur d'énoncés et de structures passe-partout »[314].

Toutefois, la réalisation de la liaison pourrait encore être ralentie par le fait que les informations morphosyntaxiques sont déjà encodées sous formes de désinences au niveau du lexème, ce qui rend la production de la liaison redondante d'un point de vue fonctionnel et explique la présence de non-réalisations.

De plus, au début de cette phase, nous pouvons supposer que les apprenants commencent à produire la liaison en s'appuyant sur des représentations phonologiques (ex. /de/)[315] mais également orthographiques (ex. < des >) des unités dont ils disposent. Autrement dit, l'influence des représentations orthographiques apprises au stade d'acquisition précédent (« procédure de catégorie » *category procedure*) sur les représentations phonologiques donnerait lieu à des réalisations orthographiques de la CL (ex. « des enfants » [desɑ̃fɑ̃]) tant en perception qu'en production (lecture et parole semi-spontanée). Nous avons formulé l'hypothèse selon laquelle

[314]« Pour la construction des énoncés, aussi pour les énoncés complexes (la subordination), l'apprenant a recours, presque dès le début, à la prédication en *c'est*. L'apprenant s'en sert tôt et fréquemment. Plus tard dans le développement, la structuration en *c'est* est liée à la subordination par le biais des clivées, des pseudo-clivées etc. Tous ces phénomènes se reflètent de façon parallèle dans les deux corpus ». (Bartning et Schlyter 2004, p. 297). L'émergence précoce de cette forme en tant qu'introducteur pourrait expliquer le taux élevé (82%) de réalisations de la liaison enregistré dans la tâche de reformulation (cf. Partie 4.Chapitre 2.2.1.2) après « c'est » par rapport au faible pourcentage (18%) observé après la seule forme verbale « est ».

[315] Il convient de souligner que l'apprenant est conscient de la forme phonologique des mots en contexte de non liaison ex. /de/ puisqu'il est capable de produire la forme correcte de l'item devant pause ou devant mot commençant par consonne.

la CL est encore traitée à ce stade comme une consonne finale du M1 présente dans la représentation orthographique du lexème dont l'activation permettrait de produire la liaison.

3.3. Stade intermédiaire - procédure inter-syntagmatique

Comme nous l'avons déjà souligné, l'interlangue de l'apprenant commence à être caractérisée au stade intermédiaire par l'émergence de sur-extensions et de régularisations (Bartning et Schlyter 2004, p. 295). Il est probable qu'à ce moment, sous l'effet des fortes probabilités transitionnelles (cf. Partie 2.Chapitre 2.2.3.1.2) des déterminants d'être suivis par un nom, l'apprenant commence à mettre en place des généralisations phonologiques sous forme de constructions telles que « les - z - NOM » et « des - z - NOM » et « nous - z - VER au pluriel ».

À ce stade en effet, on observe d'une part une réduction des formes de non-réalisations des liaisons catégoriques et de réalisations orthographiques et de l'autre, l'émergence de liaisons non attendues et de substitutions que nous avons recueillies dans notre corpus. Celles-ci sont produites par analogie avec les schémas déjà appris à savoir « les / des - z - NOM », « nous - z - VER au plurie l». En effet, les apprenants italophones réalisent des liaisons non attendues ou des substitutions uniquement à partir de ces constructions qu'ils appliquent aux contextes de liaison qu'ils ne maîtrisent pas et auxquels ils sont exposés pendant les tâches de lecture oralisée. C'est le cas par exemple des types de liaisons associées aux clitiques pour lesquels l'apprenant produit des substitutions de la CL telles que « nous en [z] installions », par analogie avec le schéma « nous - z - VER - ons ». Nous pouvons également supposer que les formes de substitutions comme « grand [z] enclos », « petit [z] enfant », qui impliquent une liaison antéposée à une catégorie principale (cf. Partie 1.Chapitre 2.2.3.2), sont calquées sur le schéma des liaisons associées aux articles définis et indéfinis « les / des - [z] - NOM » appris en premiers.

3.4. Stades avancés - procédure de phrase subordonnée

Les stades avancés sont caractérisés par une émergence de la morphologie flexionnelle qui devient fonctionnelle tout en continuant à présenter des zones « fragiles » de développement. Les pronoms clitiques apparaissent devant l'auxiliaire. Ce n'est qu'à partir de ce stade que l'apprenant est en mesure de réaliser la liaison dans ces contextes catégoriques.

Les formes « ont » et « sont » sont entièrement maîtrisées et nous pouvons supposer que les non-réalisations de liaison se réduisent davantage dans ces contextes.

En ce qui concerne les liaisons variables, l'accord du genre de l'adjectif semble toutefois poser toujours des problèmes, ce qui pourrait sans doute constituer un obstacle à la réalisation de la liaison dans ce contexte. L'acquisition des autres formes de liaison variable reste pour l'instant difficilement prévisible. Une comparaison avec les résultats d'autres enquêtes IPFC pourrait offrir des éléments importants pour aboutir à l'identification d'un itinéraire complet d'acquisition de la liaison.

4. Conclusion

Au terme de ce travail, plusieurs éléments nous paraissent émerger dont nous proposons un bilan ci-dessous.

En ce qui concerne les liaisons catégoriques, si les apprenants maîtrisent la production de la plupart des contextes catégoriques (sauf au sein des expressions figées et des pronoms clitiques), un certain nombre de non-réalisations de liaisons catégoriques persistent cependant, même au cours des dernières phases de la période d'observation. Afin de rendre compte de ces erreurs, nous avons dans le cadre de la théorie de la processabilité de Pienemann (1998) formulé l'hypothèse selon laquelle l'acquisition préalable des marques morphosyntaxiques au niveau des catégories principales pourrait ralentir le développement des généralisations en rendant la liaison redondante. Autrement dit, le fait que les informations morphosyntaxiques soient déjà encodées sous forme de désinences au niveau du lexème pourrait représenter un obstacle au passage d'un marquage au niveau catégoriel (« procédure de catégorie » *category procedure*) vers des accords entre constituants d'un syntagme (« procédure intra-syntagmatique » *phrasal procedure*) ou encore entre syntagmes (« procédure inter-syntagmatique » *S-procedure*). Cela rendrait la production de la liaison redondante d'un point de vue fonctionnel. Néanmoins, plusieurs aspects semblent suggérer que l'apprenant est en mesure de développer des représentations phonologiques de la liaison sous forme de constructions. En particulier, l'émergence progressive de formes de liaisons non attendues et de substitutions lors des dernières séances de la période d'observation montre que l'apprenant est capable de dépasser un simple traitement « au cas par cas » des contextes de liaison (Wauquier 2009). D'autres résultats semblent également indiquer que la liaison suit un itinéraire d'acquisition articulé en échelles implicationnelles. En effet, le taux élevé de non-réalisations associées aux pronoms clitiques pourrait s'expliquer par l'acquisition tardive de ces éléments (Bartning et Schlyter 2004) pour lesquels aucun schéma de réalisation de la liaison n'est accessible chez l'apprenant de niveau intermédiaire.

Quant à la production des liaisons variables, au niveau de la « macro-planification » (Racine 2014), les résultats de la présente étude rejoignent les résultats obtenus auprès d'autres sujets débutants et intermédiaires (Galazzi et al. 2013 ; Falbo et al. sous presse) qui montrent que les apprenants de niveau intermédiaire ne maîtrisent pas la variabilité sociolinguistique associée à la liaison. De plus, le recours à un test de jugement nous a permis de mettre en évidence la faible nature des connaissances épilinguistiques de la variation (Gombert 1990).

Quant aux difficultés rencontrées par les apprenants au niveau de la « micro-planification » (Racine 2014) dans la production, nos résultats ont relevé que les réalisations orthographiques sont présentes non seulement dans les lectures oralisées mais aussi dans des tâches de production orale semi-spontanée, de perception ainsi que dans le choix des variantes orthographiques pendant le test de jugement. Afin de rendre compte de ces données, nous avons postulé, dans le cadre du modèle interactif bimodal (Grainger et Ferrand 1996), que les représentations orthographiques et phonologiques de la liaison interagissent entre elles dans les productions des apprenants. Ces interactions seraient favorisées encore une fois par le fait qu'au début de leur acquisition les apprenants ont encodé les informations morphosyntaxiques au niveau des lexèmes (« procédure de catégorie » *category procedure*). Autrement dit, lors des premières phases du développement (« procédure de catégorie » *category procedure*), la liaison peut être envisagée comme le résultat de l'influence des représentations orthographiques sur les représentations phonologiques qui, dans le cas de non correspondance entre consonnes graphiques et CL, donnent lieu à des réalisations orthographiques de la CL tant en perception qu'en production.

Quant à la production de liaison non-enchaînées, l'absence de formes de liaisons non-enchaînées en parole semi-spontanée semble suggérer que les non-enchaînements représenteraient un phénomène marginal dans la production de la liaison relevant majoritairement de l'influence de l'input orthographique dans la lecture qui favorise un traitement lexical de la liaison et non pas d'une difficulté à resyllaber la CL.

Compte tenu de l'ensemble des difficultés rencontrées par les apprenants de FLE dans la production de la liaison (à la fois au niveau de la « micro-planification » et de la « macro-planification »), il nous a paru intéressant, dans une démarche didactique, de développer des outils didactiques pour l'enseignement de la liaison en FLE, tout en nous appuyant sur des résultats issus de l'analyse fréquentielle du corpus PFC (cf. Partie 2.Chapitre 2.2.2.1).

Partie 5. DES ATOUTS D'UN CORPUS DE NATIFS DANS L'ACQUISITION DE LA LIAISON EN FRANCAIS LANGUE ETRANGERE

Chapitre 1. Ressources pour l'acquisition de la liaison en français langue étrangère

Les résultats de l'étude menée auprès des apprenants italophones de FLE que nous avons présentés dans la quatrième partie de ce travail (cf. Partie 4), ainsi que les comparaisons réalisées avec les autres travaux issus du projet IPFC nous ont permis d'etablir des hypothèses sur l'acquisition de la liaison chez l'apprenant FLE concernant: (a) les représentations phonologiques des mots de liaison (cf. Partie 4.Chapitre 3.1.1.2) ; (b) le développement de généralisations phonologiques dans la production des liaisons catégoriques (cf. Partie 4.Chapitre 3.1.1.2) ; (c) les interactions entre phonologie et orthographe dans la production de la liaison (cf. Partie 4.Chapitre 3.1.3.1) ; (d) les compétences ainsi que les connaissances épilinguistiques de la variation de la liaison.

Dans les pages qui suivent, compte tenu des représentations phonologiques des mots de liaison de l'apprenant qui semblent émerger de notre étude, nous nous interrogeons sur le traitement que la CL a reçu dans l'enseignement du FLE.

Les données issues de l'analyse des réalisations orthographiques de la CL seront également l'occasion de développer une réflexion sur les approches d'étude des correspondances phonographiques dans l'acquisition de la liaison en FLE.

Pour conclure, étant données les faibles compétences et connaissances de la variation de la liaison observées chez l'apprenant FLE, nous présenterons des outils (notamment une classification et un concordancier multi-niveaux) issus de l'analyse fréquentielle menée sur le corpus PFC (cf. Partie 2.Chapitre 2.2.2.1). Le public auquel ces ressources s'adressent est varié : chercheurs, enseignants et apprenants de français langue seconde et étrangère. Nous espérons que celles-ci pourront servir de supports complémentaires pour l'acquisition de la liaison, dans le but de sensibiliser l'apprenant de FLE à la variation de la liaison (cf. Partie 3.Chapitre 3.1.3).

1. Quel traitement de la CL dans l'enseignement du FLE ?

L'étude de la nature des erreurs[316] de liaison attestées dans les productions des apprenants FLE semble suggérer que la CL n'est pas encodée dans la représentation phonologique du mot de liaison (cf. Partie 4.Chapitre 3.1.1.2). La production de la liaison semblerait, au moins dans

[316] Notamment la production des non-réalisations dans des contextes de liaisons catégoriques ainsi que le faible taux de liaisons non-attendues devant des mots à initiale consonantique (ex. « les garçons » [lezgaʀsɔ̃]) (cf. Partie 4.Chapitre 3.1.1.2).

les premières phases de l'acquisition, fortement influencée par la représentation orthographique (ex. < des >) des mots de liaison, ce qui justifierait la présence de réalisations orthographiques (ex. « des arbres » [desaʀbʀ]) à côté des non-réalisations des liaisons catégoriques (ex. « des arbres » [deaʀbʀ]). Dans notre étude nous avons observé que la mise en place de généralisations phonologiques dans la production de la liaison (attestées par des erreurs de substitutions et des liaisons non-attendues) s'accompagne d'une réduction progressive des réalisations orthographiques (cf. Partie 4.Chapitre 2.1.1). Nous avons suggeré que le retard dans le développement des généralisations pourrait être expliqué par l'acquisition préalable en FLE des informations morphosyntaxiques encodées sous formes de désinences au niveau du lexème, ce qui rendrait la production de la liaison redondante d'un point de vue fonctionnel (cf. Partie 4.Chapitre 3.1.1.1.1)

À la lumière de ces résultats nous nous interrogeons sur les approches d'enseignement de la liaison en FLE et notamment sur le traitement des CL proposé par les matériaux didactiques.

Dans la plupart des manuels, comme nous l'avons vu dans la troisième partie de ce travail (cf. Partie 3.Chapitre 3.2), la CL est presentée comme une consonne finale du mot de liaison (à savoir M1) qui est prononcée devant un mot suivant (M2) commençant par voyelle, les deux mots étant syntaxiquement (prosodiquement selon les auteurs cf. Léon 1996) liés. En dehors de ces conditions phonologiques et syntaxiques, la CL est considérée comme « muette » (Fouché 1959).

En guise d'exemple nous pouvons citer deux définitions : l'une issue du « Traité de prononciation française » de Fouché (1959) et l'autre bien plus récente tirée de « Les 500 exercices de phonétique : niveau A1/A2 » d'Abry et Chalaron (2009).

> « La liaison consiste […] à prononcer devant un mot commençant par voyelle un consonne finale, muette en dehors de cette condition ». (Fouché 1959, p. 434)

> « La consonne 'n' du mot *un* est prononcée seulement devant voyelle. Elle passe alors dans la syllabe suivante. C'est la liaison. Exemple : un aéroport [ɛ̃-na-e-ʀo-pɔʀ] ». (Abry et Chalaron 2009, 23)[317]

En dépit de la distance temporelle et méthodologique qui sépare ces deux manuels, il semblerait que le traitement retenu pour l'analyse de la CL est identique. La CL est encore analysée comme une consonne finale du M1 qui lorsqu'elle est réalisée (sous certaines

[317] Cette définition rappelle celle qui avait déjà été donnée par Charliac et Montron (1998) (cf. Partie 3.Chapitre 3.2.2.).

« En français oral, de nombreuses consonnes finales de mots sont muettes (voir p.11). Certaines consonnes finales de mot peuvent, dans certains cas, être prononcées avec la voyelle initiale du mot qui suit. C'est la liaison ». (Charliac et Motron 1998, p. 16)

conditions) « passe dans la syllabe suivante ». En revanche, dans « Les 500 exercices de phonétique : niveau B1/B2 », Abry et Chalaron (2011) décrivent la CL comme une consonne finale « latente ». Cette définition nous paraît toutefois incomplète puisque la notion de latence ne semble s'appliquer qu'au contexte où le M1 est « pris en isolement ». Aucune indication n'est donc donnée pour les cas où le mot suivant commence par consonne.

> « La liaison est le passage, dans le mot suivant, d'une consonne 'latente' (non prononcée dans un mot pris isolement). Exemple : sans [sɑ̃], sans elle [sɑ̃-zɛl], chez [ʃe] chez elle [ʃe-zɛl]. Elles sont obligatoires, facultatives ou interdites ». (Abry et Chalaron 2011, p. 98)

Lauret (2007) fait en revanche référence à une consonne finale « muette » (cf. ci-dessus Fouché 1959).

> « […] à différence de l'enchaînement, la consonne dite de liaison n'est prononcée que lorsqu'elle est suivie par une voyelle (*mon petit-ami*). En revanche, elle est muette devant une autre consonne (*mon peti(t) copain*) ou en finale absolue (*il est peti(t)*) ». (Lauret 2007, p. 60)

L'ensemble de ces analyses de la CL en tant que consonne finale du M1 et en même temps « latente » ou « muette » dans certaines conditions nous paraissent complexifier la représentation phonologique des mots que l'apprenant FLE a construite suite à l'étude du comportement des consonnes finales. En effet, l'apprenant FLE dès les premiers cours (au moins dans la plupart des approches) est confronté à l'apprentissage des règles qui régissent la prononciation des consonnes finales. Il est donc probable qu'à partir de cet enseignement l'apprenant commence à construire des représentations phonologiques qui excluent toutes les consonnes graphiques finales.

Comme nous l'avons suggéré dans le chapitre précédent, au début de l'acquisition l'apprenant serait donc amené à avoir recours à la forme orthographique des mots afin de rechercher dans l'écrit la CL à prononcer dans les contextes de réalisation de liaison, ce qui rendrait la réalisation de la liaison plus complexe. Compte tenu de ces observations et de la nature des erreurs de liaison attestées dans l'ensemble des études menées sur l'acquisition de la liaison en FLE, nous pouvons supposer que le traitement didactique de la CL comme consonne finale dans les manuels pourrait contribuer d'une part à ralentir la mise en place de généralisations phonologiques de liaison et d'autre part à favoriser la production de réalisations orthographiques des CL[318].

[318] La définition proposée par Clédat (1890) nous paraît intéressante et semble suggérer un traitement épenthétique de la CL.

« On remarque que les formes de l'article défini pluriel devant les voyelles ne diffèrent des formes de cet article devant consonnes que par l'adjonction d'un z de liaison ». (Clédat 1890, p. 29)

Celle-ci s'inscrit dans la « méthode phonétique » (Galazzi 2002a, p. 131) proposée par Clédat (1980) selon lequel les étrangers ainsi que les natifs « peuvent évidemment se contenter, pour écrire le français, d'une graphie

1.1. L'étude des correspondances phonographiques

Les raisons de la production de réalisations orthographiques en acquisition du FLE peuvent être également recherchées dans la place marginale accordée à l'enseignement des correspondances entre la graphie et la phonie dans les manuels d'enseignement du FLE.

En effet comme le soulignent Racine et Detey (sous presse), (cf. Partie 3.Chapitre 3.2), à l'exception de quelques ouvrages (Léon 1966 ; Léon 2007), la plupart des auteurs proposent des descriptions sommaires des correspondances entre la graphie et les réalisations phonétiques des CL. Celles-ci sont le plus souvent présentées à partir des graphèmes (ex. < d > > [t]). Léon (1966, p. 128) est le seul à proposer des exercices qui demandent de retrouver la graphie de mots à partir des transcriptions phonétiques des mots commençant par « h » aspiré (ex. [leero] « les héros »). Nous suggérons que cet exercice pourrait être élargi à d'autres contextes de liaison notamment dans les cas où la CL n'est pas identique à la dernière consonne écrite du M1, à savoir pour la plupart des CL les plus fréquentes (cf. Partie 1.Chapitre 1.4) :

- < s > : comme dans « les amis » [lezami] ;
- < x > : comme dans « deux autres » [døzotʀ] ;
- < d > : comme dans « quand il pleut » [kɑ̃tilplø].

Nous pourrions donc imaginer de demander à l'apprenant de reconstruire la graphie de ces séquences à partir des transcriptions phonétiques correspondantes.

Cette activité contrairement à une analyse de la CL comme une consonne finale du M1 propose une approche qui pourrait d'une part favoriser la mise en place de généralisations phonologiques, et d'autre part limiter la production de réalisations orthographiques de la CL (cf. Partie 3.Chapitre 2.1.4)[319]. D'un point de vue didactique cet exercice simple nous paraît tout à fait compatible avec les inconvénients résultant de la place limitée accordée à

phonétique, dégagée de toutes les subtilités de l'orthographe officielle, et qui […] sera beaucoup plus logique et beaucoup plus intelligible que les graphies gauches, et si souvent barbares, des Français peu lettrés, où l'orthographe officielle est écorchées sans être simplifiée ». (Clédat 1890, p. 6)

En effet, comme le souligne Galazzi (2002a, p. 131), Clédat (1890) voyait dans l'orthographe un obstacle pour la diffusion du français à l'étranger, que la méthode phonétique qu'il adopte dans sa grammaire avait précisement pour objectif de faire disparaître. Cependant, comme l'observe Galazzi (2002a) la transcription phonétique qu'il propose (ex. *un pieû-z ami* « un pieux ami ») « est purement française, c'est-à-dire que les lettres ont la valeur qu'elles ont le plus souvent en français » (Galazzi 2002a, p. 132).

[319] Quant à l'enseignement des liaisons catégoriques, comme l'avait proposé Léon (1966, p. 123), la présentation du rôle fonctionnel joué par ces liaisons (cf. Partie 1.Chapitre 2.1) (Léon 1966, p. 123) en tant que marques du pluriel pourrait permettre à l'apprenant de comprendre le caractère « obligatoire » de la réalisation au-délà de la mémorisation des règles établies par les classifications.

l'enseignement des deuxièmes langues étrangères dans les universités italiennes (malgré les réformes ministérielles) (Galazzi 2002b, p. 60-61)[320].

2. L'acquisition de la variation de la liaison au-délà des règles classificatoires

Dans la première partie de ce travail (cf. Partie 1.Chapitre 2), nous avons vu que la variation sociolinguistique chez des apprenants en milieu institutionnel est assez limitée, même à des niveaux plus avancés. Ces observations sont étayées par les résultats du test de jugement issu du corpus IPFC-italien (Rome) qui montre comment les apprenants italophones ont des connaissances épilinguistiques (Gombert 1990) modestes de la variation de la liaison, tout au moins dans les contextes de liaison les moins fréquentes. Il est possible d'expliquer ces résultats en invoquant le faible contact de l'apprenant en milieu hétéroglotte avec la langue cible. L'apprenant de FLE est en effet exposé le plus souvent à des variantes issues des styles relativement formels (les dialogues extraits des films ou des lectures de textes journalistiques) comme le soulignent Mougeon et al. (2002) « […] l'enseignement du français aux apprenants de FLE avancés, ne peut à lui seul assurer l'appropriation de la variation sociolinguistique de cette langue » (Mougeon, Nadasdi et Rehner 2002, p. 21). D'après ces auteurs, les enseignants et les auteurs de manuels proposent aux apprenants « un modèle linguistique fortement standardisé, qui ne permet guère à ces derniers d'acquérir un répertoire sociolinguistique proche de celui des locuteurs de FL1 » (Mougeon, Nadasdi et Rehner 2002, p. 21).

Gadet et Tyne (2007) font également référence à la faible attention consacrée à la variation sociolinguistique dans les manuels.

> « Les conceptions courantes du style mises en œuvre dans les manuels, de L1 comme de L2, ne proposent aucune représentation un tant soit peu globale du phénomène : elles regardent celui-ci comme localisé en des points linguistiques, dont la maîtrise est une étape ultime, un « en plus » d'une langue déjà maîtrisée ». (Gadet et Tyne 2007, p. 132)

En effet, comme nous l'avons observé (cf. Partie 3.Chapitre 3.2), la liaison dans les manuels destinés à l'enseignement du FLE est le plus souvent (sauf quelques rares exceptions Putska 2009 ; Detey et al. 2010) traitée uniquement sous forme de listes de règles.

La description de la liaison sous forme de classifications se révèle relativement peu économique surtout si on considère que certaines liaisons proposées sont rarement produites

[320] « […] dans le répertoire des quarante-deux classes disciplinaires où doivent nécessairement être ancrés les titres universitaires de premier niveau (après 3 ans), huit seulement prévoient la connaissance de 2 LV [deuxième *langue vivante*] ». (Galazzi 2002b, p. 60-61)

par les natifs et uniquement dans des registres très soutenus (ex. entre les adverbes polysyllabiques et les mots suivants « vraiment adorable »).

L'enseignement de la liaison variable en FLE pourrait donc commencer par une présentation des séquences de liaison le plus fréquentes, réalisées par les francophones indépendemment du registre de parole. En effet l'analyse fréquentielle de la liaison que nous avons menée sur les données de parole semi-spontanée (de registres soutenu et informel) issues du corpus PFC (cf. Partie 2.Chapitre 2.2.2.2.8) a permis de mettre en évidence que les seules 16 constructions les plus fréquentes (cf. Tableau 18) donnaient lieu à 65% des réalisations totales de la liaison variable (à savoir 2755 / 4296 occurrences).

Cela ne signifie pas que l'enseignement du FLE doit renoncer à apprendre la variation sociolinguistique qui caractérise la réalisation de la liaison, mais il nous paraît que l'identification des ces tendances puisse représenter un point de départ pour l'acquisition des usages du « français standard »[321] (cf. Partie 3.Chapitre 3)

Une analyse approfondie de la variation de la liaison est donc évidemment nécessaire aussi bien pendant les heures d'enseignement qu'en apprentissage autonome. C'est dans ce but que nous présentons dans les pages qui suivent des ressources complémentaires (notamment une classification et un concordancier multi-niveaux) pour l'enseignement et l'acquisition de la liaison, issues de l'analyse fréquentielle menée sur le corpus PFC (cf. Partie 2.Chapitre 2.2.2.1)[322]. Le public auquel ces ressources s'adressent est varié : chercheurs, enseignants et apprenants de français langue seconde et étrangère.

Ces deux ressources ne sont pas isolées l'une de l'autre dans la mesure où le concordancier permet d'approfondir l'acquisition (conçue ici comme une recherche) des tendances représentées de façon synthétique par la classification. Cela peut être le cas par exemple de l'étude des productions des liaisons après des adverbes monosyllabiques. Si la description fournie par la classification multi-niveaux permet d'observer une hétérogénéité au sein de cette classe, c'est seulement par l'observation et l'écoute des concordances que l'apprenant prend conscience de cet écart.

[321] Nous faisons référence à la valeur attribuée au terme « standard » par Porcher (cité par Galazzi 2015b).
« J'ai beaucoup utilisé l'expression 'français standard' comme tout le monde parce que une approximation me suffisait alors… Ce concept abstrait servait à y rapporter les parlures concrètes qui évidemment prolifèrent. C'est une idée méthodologique forte que Kant a bien théorisée quand il expliquait que pour comprendre un ou des changements il fallait impérativement sur le plan conceptuel définir une permanence. En ce premier sens donc on peut considérer que le concept de français standard est nécessaire pour comprendre les variations […] ». (Porcher cité par Galazzi 2015b)

[322] Ce corpus, comme nous l'avons vu dans la deuxième partie de ce travail (Partie 2.Chapitre 1.1), grâce à ses nombreux points d'enquête représente une ressource capable de rendre compte de la diversité du français parlé au sein de l'espace francophone.

2.1. Proposition d'une classification multi-niveaux de la liaison

Les résultats issus de l'analyse de la distribution de la fréquence de réalisation de la liaison menée sur le corpus PFC (cf. Partie 2.Chapitre 2.2.2.1) nous ont permis de parvenir à une description plus fine de la variation de la liaison au sein des types de liaison variable. Celle-ci a abouti à l'élaboration d'une classification multi-niveaux de la liaison réalisée à l'aide de graphiques *drill-down*[323]. Cette description est donc articulée sur deux niveaux. Le premier niveau (cf. Figure 36) comprend les indices de fréquence[324] de réalisation de liaison (sur l'axe des ordonnées) associés à l'ensemble des types de liaison catégorique et variable (sur l'axe des abscisses) (ex. ADJ-NOM) tels qu'ils sont décrits dans les classifications traditionnelles (Delattre 1966 ; Encrevé 1988). La représentation choisie est un histogramme empilé comportant deux séries, une représentant les liaisons réalisées (à savoir « LR » en bleu) et l'autre les liaisons non-réalisées (à savoir « LNR » en noir).

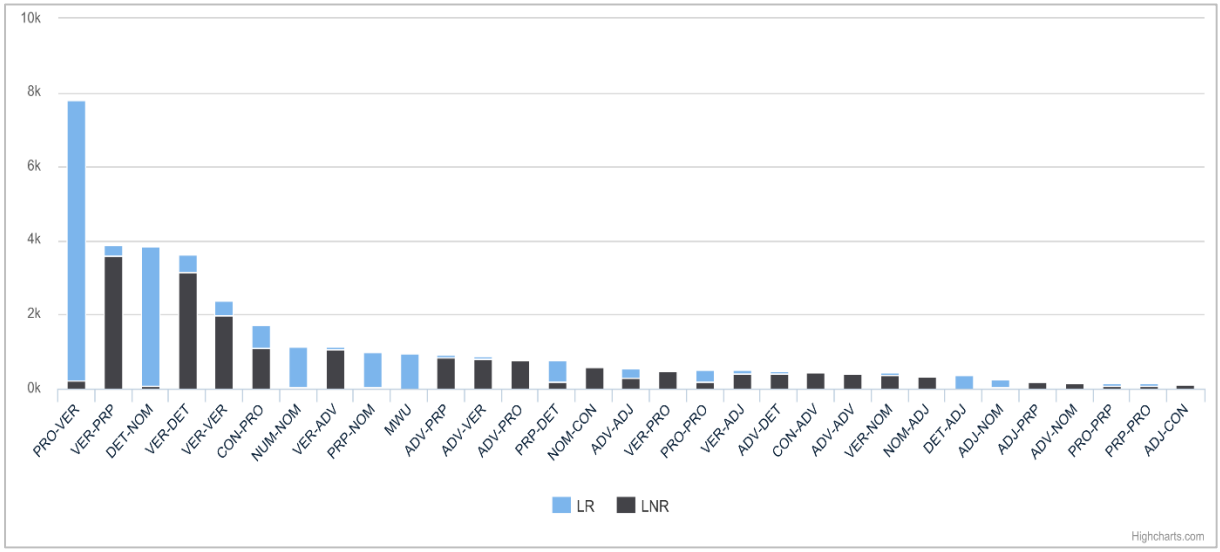


Figure 36: Premier niveau de la classification multi-niveaux

[323] Nous remercions Arnaud Dufranne, ingénieur informatique, pour avoir développé ces outils de répresentation, réalisés à l'aide de la bibliothèque graphique *Highcharts*.

[324] Le symbole « k » à côté des indices de fréquence représente l'unité de mésure à savoir mille.

En positionnant le curseur sur la série souhaitée il est possible de consulter les indices de fréquence de réalisation (LR) et de non-réalisation (LNR) de la liaison (cf. Figure 37).

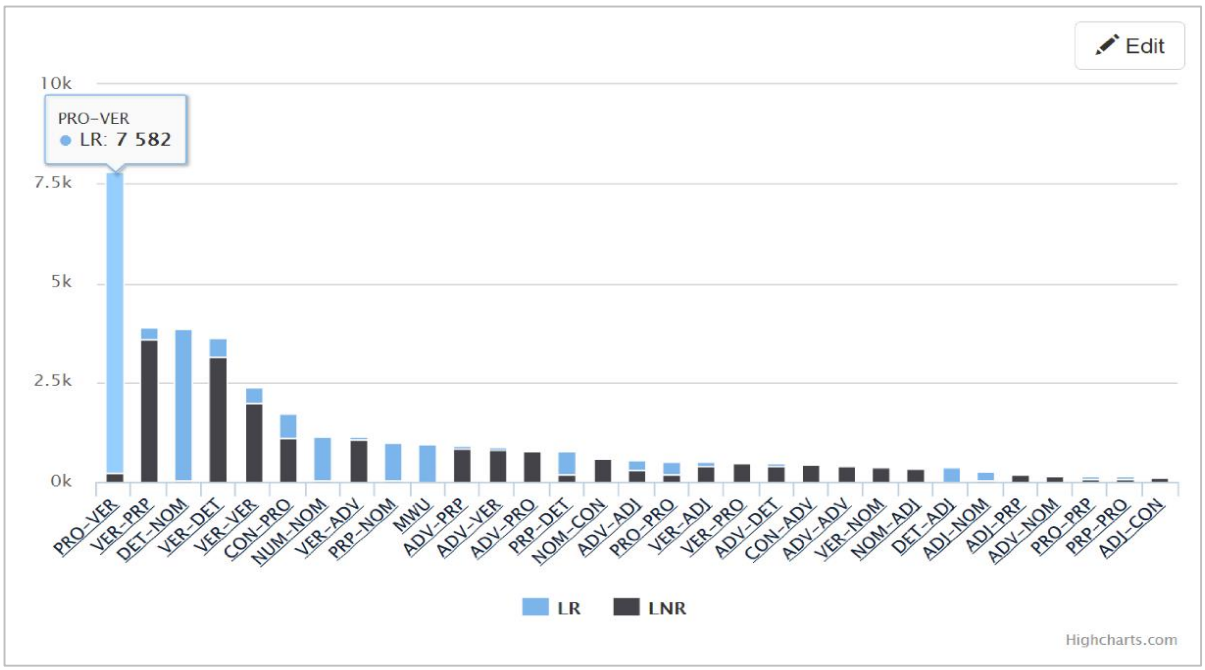


Figure 37 : Affichage icône des indices de fréquence de réalisation (niveau 1)

L'utilisateur peut également décider d'afficher un seul type de réalisation (LR ou LNR), en décochant le carré de la valeur qu'il ne souhaite pas représenter, situé à côté de la légende. Cette modalité d'affichage permet de mieux comparer les réalisations entre les types de liaison (cf. Figure 38).

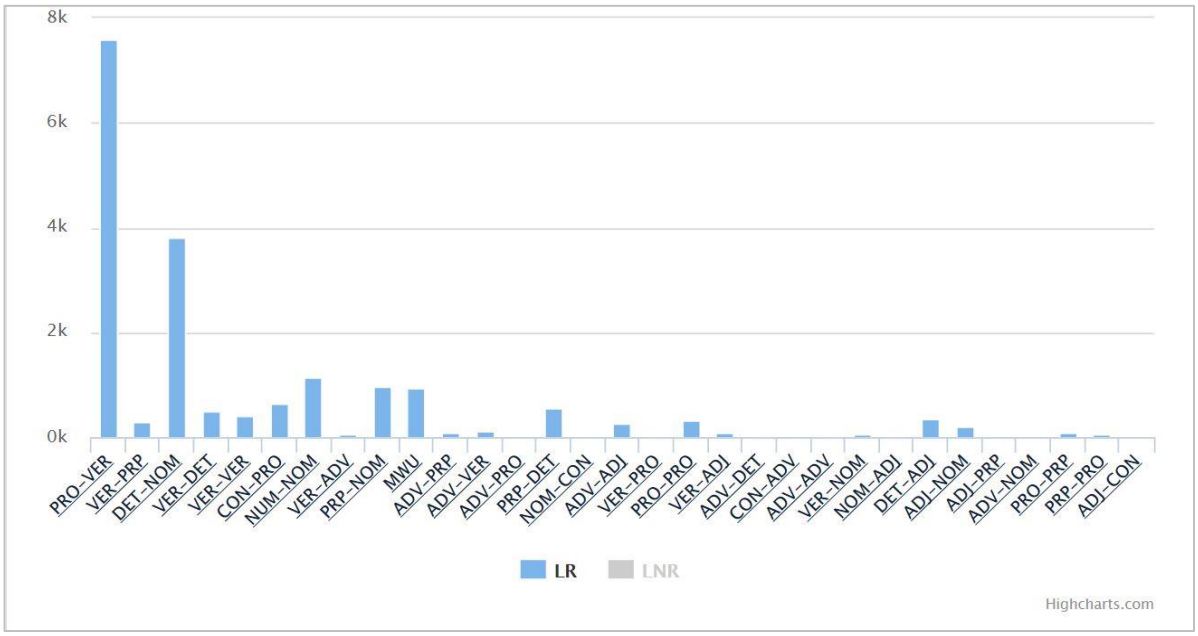


Figure 38 : Sélection de l'histogramme des liaisons réalisées (niveau 1)

De plus, en sélectionnant en un seul clic de souris un type particulier du premier niveau (ex. ADJ-NOM), l'utilisateur a accès au deuxième niveau de la classification (cf. Figure 39), correspondant aux constructions internes (ex. « petit - t - NOM ») correspondant au type de premier niveau choisi (ex. ADJ-NOM).

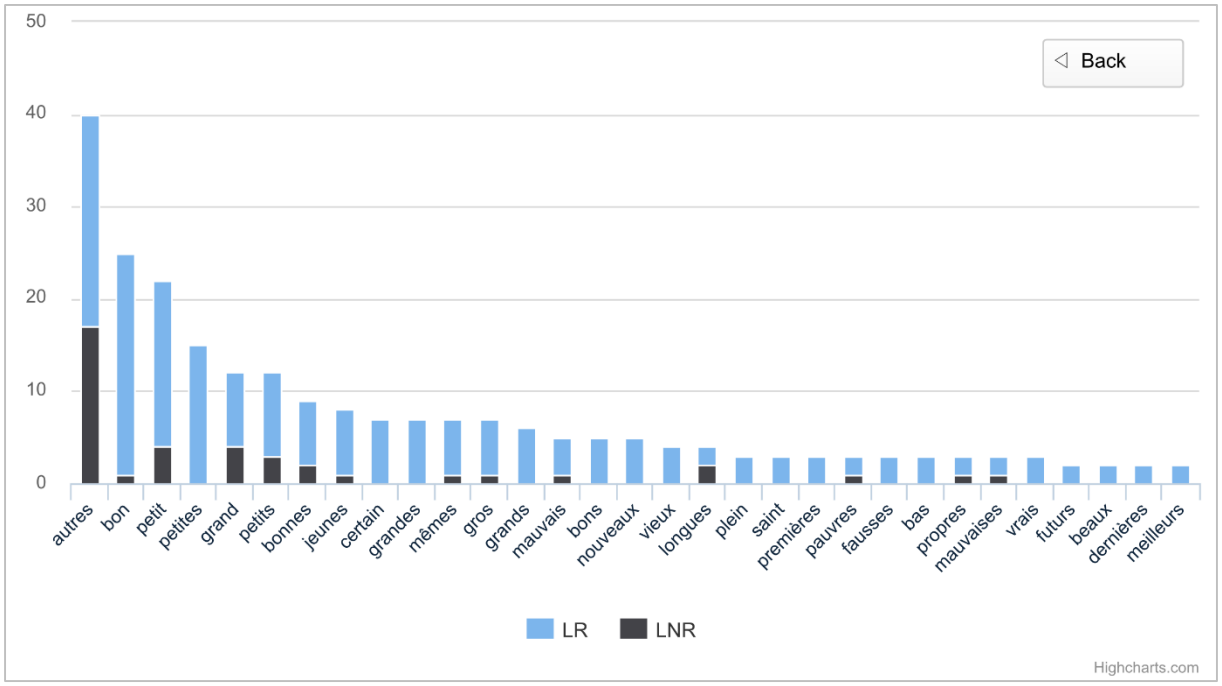


Figure 39 : Deuxième niveau de la classification multi-niveaux du type ADJ-NOM

À ce stade, il est possible de consulter les indices de fréquence des productions associés aux différents M1 (ex. différents ADJ « petit » ; « grand » etc.). Par exemple, en cliquant sur le type ADJ-NOM au premier niveau de la classification l'utilisateur a accès directement aux fréquences de réalisation de la liaison après les différents adjectifs. Cette simple représentation permet de mettre en évidence le nombre réduit d'adjectifs antéposés après lesquels la liaison est réalisée de façon presque systématique. On observe ainsi la même tendance que nous avons évoquée dans la deuxième partie de ce travail (cf. Partie 2.Chapitre 2.2.2.3), à savoir que les adjectifs les plus fréquents sont également ceux qui présentent le nombre le plus important de non-réalisations.

En revanche l'analyse au deuxième niveau du type ADV-ADJ (cf. Figure 40) permet de se rendre rapidement compte, d'une part, de l'écart de réalisation de la liaison qui caractérise les adverbes monosyllabiques et d'autre part de l'absence presque totale de réalisation après les adverbes polysyllabiques.

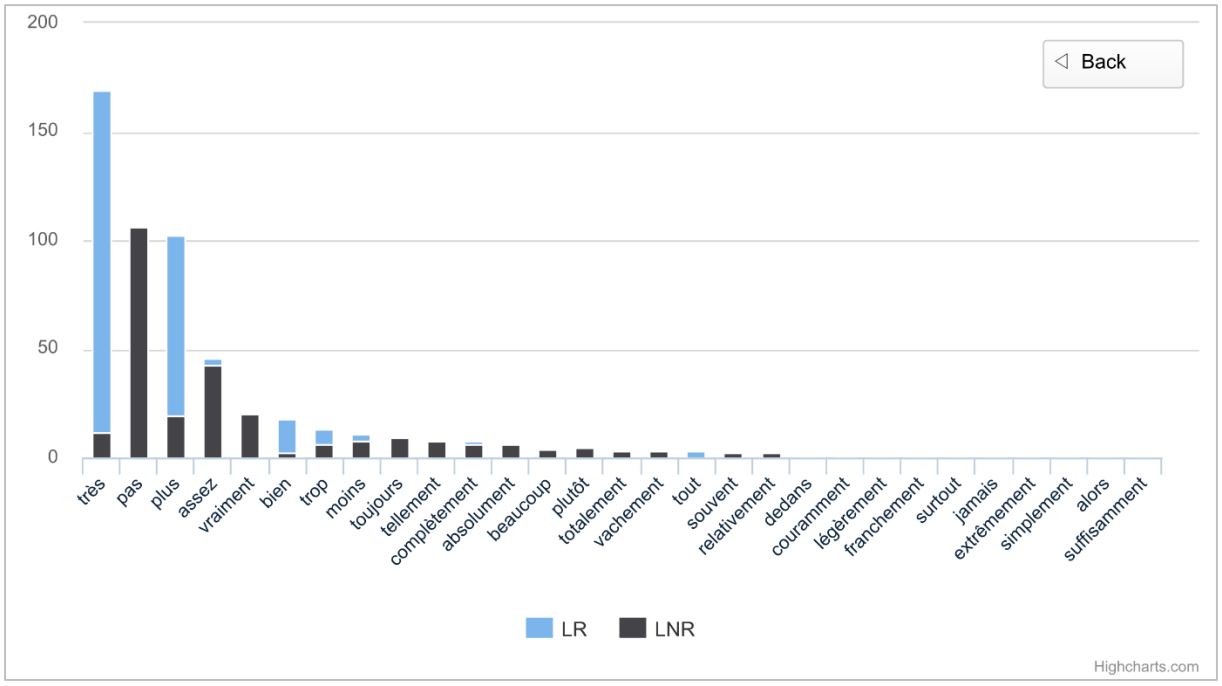


Figure 40 : Deuxième niveau de la classification multi-niveaux du type ADV-ADJ

Un bouton *back* (« retour en arrère ») permet de revenir au premier niveau de la classification pour effectuer une nouvelle recherche.

Nous avons décidé de ne pas attribuer aux données les étiquettes des classifications traditionnelles (ex. liaison catégorique, variable et erratique) mais tout simplement d'essayer de représenter la variabilité au sein des différents types de liaison. Comme nous l'avons déjà évoqué, cette démarche permettrait de mettre en évidence les principales idiosyncrasies internes aux contextes morphosyntaxiques de liaison. Celles-ci pourront être analysées plus en détails grâce au recours au concordancier multi-niveaux et multimédia que nous avons développé afin d'exploiter les données issus du projet PFC dans le cadre d'un apprentissage sur corpus.

2.2. Un concordancier multi-niveaux et multimédia pour l'étude de la liaison

Les concordanciers jouent depuis longtemps un rôle important dans l'analyse des corpus linguistiques, tout comme dans les domaines de la philologie, de la littérature, de la traduction et de l'enseignement des langues[325].

L'utilisation d'outils informatiques tels que les concordanciers, qui permettent d'extraire des collocations de mots et leurs contextes, est en effet fréquente dans la communauté des

[325] Le contenu des pages qui suivent est issu du travail commun que nous avons mené avec George Christodoulides (cf. Barreca et Christodoulides 2014).

chercheurs en sciences du langage. La taille des corpus actuellement disponibles est en augmentation constante, de même que les schèmes d'annotation, ce qui entraîne la création de nouveaux systèmes pour interroger ces données. Dans les pages qui suivent, après avoir décrit brièvement quelques-uns des travaux précédents, nous présenterons un prototype de concordancier multi-niveaux et multimédia (Barreca et Christodoulides 2014) que nous avons développé à partir des données du corpus PFC (Durand, Laks et Lyche 2009). L'outil permet non seulement d'élargir les modalités d'accès au corpus mais constitue également une ressource pour l'acquisition de la liaison.

Notre objectif consiste à nous rapprocher le plus possible des besoins des linguistes et des apprenants, qui souhaitent dépouiller de façon rapide et précise les contextes d'occurrences d'un lexème ou d'une séquence de lexèmes, et bien évidemment des contextes de liaison. Notre outil, contrairement à d'autres concordanciers, intègre l'écoute des extraits d'enregistrements au travail de recherche des concordances. De plus, cet outil permet de lancer des requêtes en croisant plusieurs couches (niveaux) d'annotation alignées temporellement avec le signal. Le moteur de recherche du concordancier construit de manière dynamique des requêtes SQL vers une base de données relationnelle, dans laquelle sont stockées les annotations. Dans le cas du corpus PFC, ces annotations associées à la transcription sont de nature morphosyntaxique (POS, lemme) et phonologique (codage des phénomènes de la liaison et du *schwa*).

2.2.1. Typologies

Le concordancier a toujours été un outil de grand intérêt pour l'analyse des contextes d'emploi des lexèmes dans des corpus oraux ou écrits. Ses applications théoriques et pratiques sont nombreuses, et intéressent plusieurs disciplines : philologie, littérature, syntaxe et sémantique, traduction et didactique des langues (cf. Barreca et Christodoulides 2015). Pourtant, comme le soulignent Pincemin et al. (2006), la plupart des concordanciers disponibles pour l'oral n'exploitent pas les informations linguistiques introduites par l'annotation des corpus (ex. catégories morphosyntaxiques, lemmatisation, etc.). Les résultats produits par ce genre de concordancier n'aboutissent qu'à de simples collections d'occurrences de chaînes de caractères (ex. *Lextutor* et *Lexiqum*) (Stevens 1991).

Plusieurs concordanciers existent déjà pour les corpus écrits : nous pouvons citer parmi ceux-ci *Condor* (développé par le Laboratoire lorrain de recherche en informatique - LORIA), qui permet d'effectuer des recherches simples ou de cooccurrences en utilisant le mot-forme ou le lemme, et d'indexer les formes présentes dans le corpus. La version publiquement disponible propose une collection des corpus écrits, pour la plupart littéraires. Le concordancier *TXM*

permet de produire des concordances KWIC (*keyword in context*) à partir des propriétés des mots (ex. lemme, catégorie POS) ou de lancer des requêtes plus complexes grâce à une syntaxe propre à son moteur de recherche.

L'intervention des informations de nature linguistique, et donc l'affinement des résultats des concordances, demande un traitement préalable du corpus qui est souvent très couteux. C'est pour cette raison que de nombreux concordanciers lemmatisés font appel à des dictionnaires. Le recours à ce type de ressources linguistiques permet notamment de mettre en relation les occurrences du texte avec les lemmes et les formes fléchies contenues dans les dictionnaires. Parmi les concordanciers qui utilisent ce type d'outils linguistiques, nous pouvons citer *Unitex* et *Stella*.

Toutefois, l'application des concordanciers conçus pour l'écrit à l'analyse des corpus oraux pose un problème de fond : ces outils ne permettent pas de tenir compte de l'alignement des annotations linguistiques avec le signal sonore. Les requêtes possibles sont ainsi limitées à ce qu'on peut représenter avec les conventions de transcription choisies. À notre connaissance, les concordanciers lemmatisés pour l'analyse de corpus de français parlé sont peu nombreux. Mis à part le concordancier de la base lexicale « Lexique 3 » (New 2006), basé sur les transcriptions des sous-titrages de films, les autres concordanciers disponibles pour les corpus oraux ne permettent d'obtenir que des relevés des occurrences et des données statistiques sur leurs collocations. Nous considérons qu'un concordancier multimédia spécifique pour l'analyse d'un corpus oral pourrait permettre aux utilisateurs de mieux prendre en compte l'articulation des différents niveaux d'annotation dans le temps et d'exploiter les résultats des requêtes tout en gardant un lien avec le signal sonore.

Comme nous l'avons souligné, ces concordanciers sont parfois utilisés dans l'enseignement du FLE, en particulier comme support pour l'enseignement de la grammaire (Rézeau 1997) ou pour l'analyse et la correction des productions des apprenants (Rézeau 2007). Il s'agit principalement de concordanciers basés sur des corpus écrits comme le forum de discussion, ou sur des extraits des journaux. Ces ressources, bien qu'elles soient précieuses dans la didactique du FLE, ne sont pas exploitables pour l'enseignement du français oral et notamment de la liaison.

2.2.2. Un concordancier pour une étude multi-niveaux et multimédia de la liaison

Contrairement aux concordanciers que nous venons d'évoquer, basés principalement sur des corpus écrits, nous proposons un concordancier multi-niveaux et multimédia s'appuyant

sur des données issues d'un corpus oral, exploitable pour l'enseignement de la liaison et des recherches linguistiques en français parlé. Ce concordancier constitue aujourd'hui une des fonctionnalités du logiciel Praaline (Christodoulides 2014), un outil en libre source pour l'analyse et la gestion de corpus de langue[326].

Pour ce faire, nous avons utilisé le corpus PFC annoté en morphosyntaxe (POS et lemmes) à l'aide de l'étiqueteur *DisMo* (Christodoulides, Avanzi et Goldmann 2014 ; Christodoulides, Barreca et Avanzi 2015) et qui présente des annotations des réalisations de la liaison. Nous avons essayé de créer un outil qui permette d'établir une articulation entre ces différents niveaux d'analyse (morphosyntaxique, lexicale et phonologique), pouvant ainsi offrir une observation linguistique plus précise de la variation de la liaison. De plus, ce concordancier permet à l'utilisateur d'accéder directement aux écoutes des intervalles des enregistrements contenant les concordances, ce qui élargit les modalités d'accès à l'analyse des sequences récherchées.

L'application d'un concondancier lemmatisé et multimédia du français parlé constituerait une ressource précieuse pour l'enseignement de la liaison, en favorisant un apprentissage basé sur des données authentiques et conçu comme une recherche (*data-driven learning*, apprentissage sur corpus) (Johns 1988), dans la lignée des travaux réalisés dans le cadre du sous-projet PFC-EF (PFC Enseignement du Français, Detey et al. 2010) (cf. Partie 3.Chapitre 3.2.3). Le recours à cette ressource fournirait aux apprenants une observation directe de la langue en favorisant une sensibilisation aux usages réels des natifs.

Dans les paragraphes suivants, nous illustrons plus en détails la structure du concordancier.

[326] http://www.corpusannotation.org

2.2.2.1 *Illustration*

Dans les pages qui suivent, nous allons donner une démonstration des requêtes qui peuvent être réalisées à partir du prototype de notre concordancier.

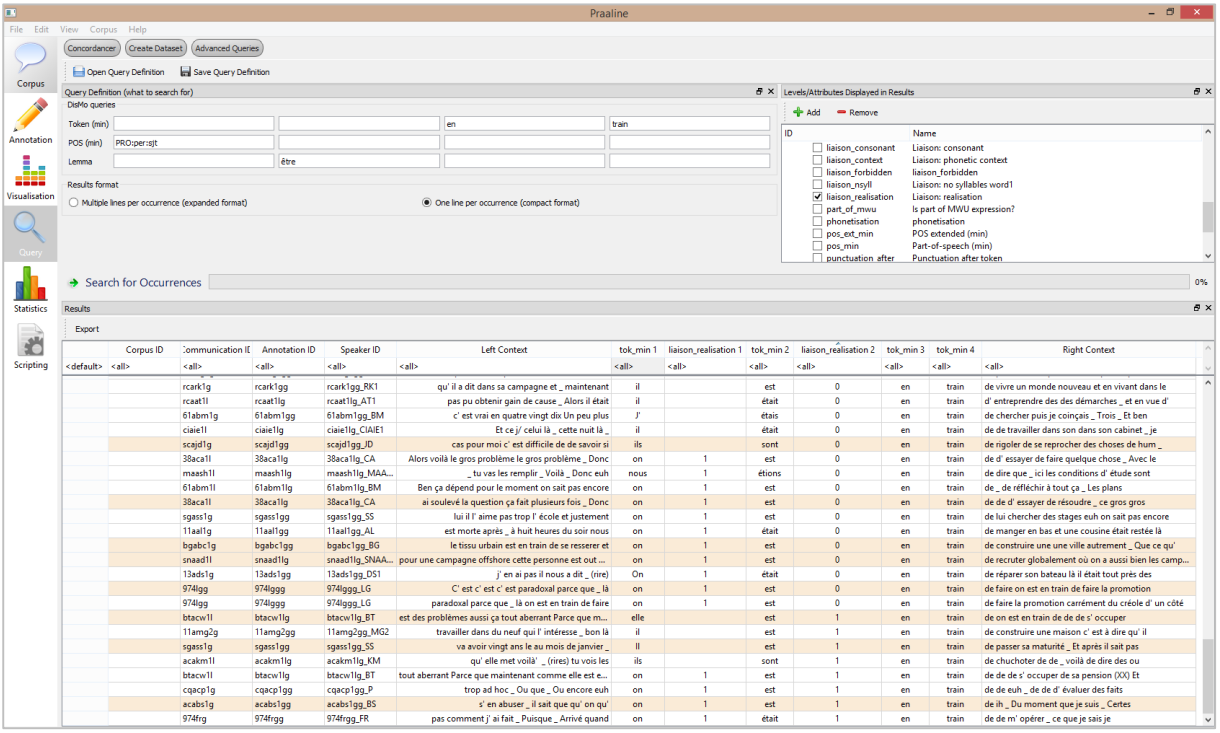


Figure 41 : Trois champs de l'interface du concordancier multi-niveaux et multimédia

L'interface du concordancier (cf. Figure 41) s'articule en trois sections : (a) un champ pour formuler les requêtes (*Query Definition*) sur la base de trois attributs de recherche ; (b) un champ pour la sélection des attributs du corpus à afficher dans les résultats (*Levels / Attributes Displayed in Results*) ; (c) un champ d'affichage des résultats (*Results*).

2.2.2.2 *Définition de la requête*

Dans le champ *Query definition* (cf. Figure 42) l'utilisateur formule la requête de la concordance qui peut comprendre jusqu'à quatre unités. Cette recherche peut être effectuée en combinant trois différents attributs: (a) l'unité lexicale minimale (à savoir le M1 et le M2 de liaison) (*Token (min)*) (cf. Partie 2.Chapitre 1.2); (b) le lemme du M1 et du M2 (*Lemma*); (c) la catégorie morphosyntaxique d'appartenance du M1 et du M2 (*POS(min)*) (cf. Partie 2.Chapitre 1.2). Ces attributs (*Token (min)*, *Lemma* et *POS(min)*) font référence aux couches de l'annotation (*tok:min*, *lemma* et *POS:min*) du corpus PFC réalisée par le biais de l'annotateur morphosyntaxique *DisMo* (Christodoulides, Avanzi et Goldmann 2014 ; Christodoulides, Barreca et Avanzi 2015), que nous avons présentée dans la deuxième partie de ce travail (cf. Partie 2.Chapitre 1.2). Il convient donc de rappeler que dans le cadre de l'annotation proposée

par l'annotateur *DisMo*, les termes *tok:min* (à savoir *Minimal Token*) et *POS:min* désignent respectivement le lexème minimal et l'étiquette morphosyntaxique minimale issus du premier niveau de *tokenisation* et d'annotation morphosyntaxique réalisée à l'aide de *DisMo* (cf. Partie 2.Chapitre 1.2). Cette première *tokenisation* et annotation morphosyntaxique minimale est suivie d'une deuxième annotation morphosyntaxique combinée avec la détection des unités poly-lexicales. C'est à ce deuxième niveau que les unités poly-lexicales (ex. « tout à fait ») sont détectées (et donc analysées comme *tok:mwu*) et ensuite annotées (comme *POS:mwu* « ADV »).

Étant donné que le concordancier s'appuie sur la ré-annotation morphosyntaxique du corpus PFC, il est évident que les requêtes ayant pour attribut *Token (min)* et *POS(min)* doivent être formulées en utilisant le même jeu d'étiquettes utilisé pour l'annotation du corpus PFC (cf. Partie 2.Chapitre 1.2.2.1)

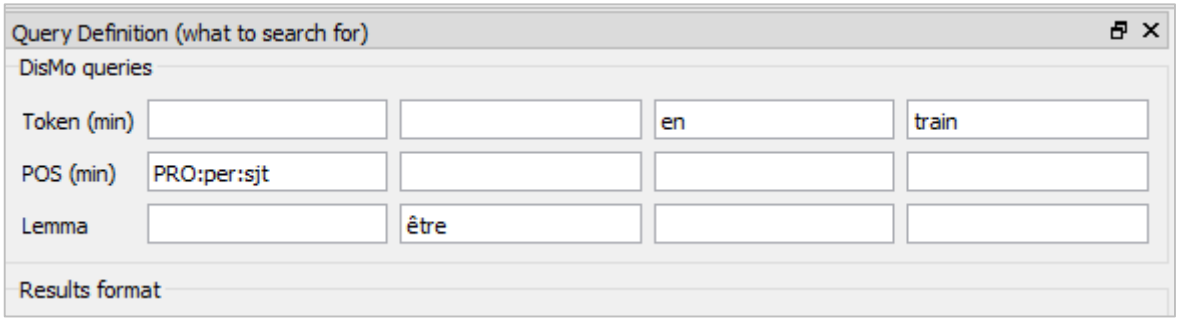


Figure 42 : Champ de requête des concordances

La figure ci-dessus (cf. Figure 42) montre la section *query definition* du concordancier qui répresente les trois attributs de requête que nous venons d'évoquer. Dans la requête ci-dessus nous avons combiné ces trois attributs afin de formuler une recherche qui contient les élements suivants : (a) attribut *POS (min)* : « PRO:per:sjt » à savoir le pronom personnel sujet; (b) attribut *Lemma* : le lemme du verbe « être » ; attribut *Token (min)* : l'unité lexicale « en » ; (c) attribut *Token (min)* ; l'unité lexicale « train ». La requête ainsi formulée permet d'accéder à des concordances contenant tout type de pronom personnel sujet suivi de n'importe quelle forme du verbe « être » se trouvant dans la locution verbale « être en train de ».

2.2.2.3 Sélection des attributs du corpus à afficher dans les résultats

Une fois définie la requête, l'utilisateur peut décider quels attributs des éléments recherchés (dans la section *query definition*) afficher dans les résultats.

La section (*Levels / Attributes Displayed in Results*) (en haut à droit de l'interface du logiciel) (cf. Figure 43) permet de sélectionner les attributs contenus dans le corpus PFC.

Ceux-ci sont articulés sur deux niveaux principaux correspondant aux deux niveaux de tokenisation du corpus (cf. Partie 2.Chapitre 1.2.2).

D'une part les attributs des unités lexicales minimales (*tok:min*) :

- les CL associées aux unités lexicales minimales recherchées (*liaison_consonant*)
- nombre de syllabes des unités lexicales minimales recherchées, si correspondant au M1 d'une séquence de liaison (*liaison_nsyll*)[327];
- les contextes de liaison erratique (*liaison_forbidden*) ;[328]
- les lemmes des unités lexicales minimales recherchées (*lemma_min*) ;
- les étiquettes morphosyntaxiques des unités lexicales minimales recherchées (*pos_min*) ;
- le genre et le nombre des unités lexicales minimales recherchées (*pos_ext_min*) ;
- l'éventuelle appartenance à une unité poly-lexicale au niveau de tokenisation supérieur (à savoir *tok:mwu*) (*part_of_mwu*).

D'autre part les attributs associés aux unités poly-lexicales (*tok:mwu*) :

- les lemmes des unités poly-lexicales recherchées (*lemma_mwu*) ;
- les étiquettes morphosyntaxiques des unités poly-lexicales recherchées (*pos_mwu*) ;
- le genre et le nombre des unités poly-lexicales recherchées (*pos_ext_mwu*).

Afin de sélectionner les attributs évoqués ci-dessus il faut tout d'abord choisir le niveau de tokenisation souhaité (*tok-min* ou *tok-mwu*) et ensuite cocher la case des attributs située à côté de leurs identifiants (*ID*) (ex. *liaison_realisation*) associé à leurs noms complets (ex. *Liaison : realisation*).

[327] Ces deux derniers attributs font référence aux conventions établies par le codage liaison PFC (cf. Partie 2.Chapitre 1.1.3.3).

[328] D'autres attributs sont également disponibles : (1) le « codage liaison » utilisé par le protocole du projet PFC (*liaison*) ; (2) le « codage schwa » utilisé par le protocole du projet PFC (*schwa*) ; (3) le score de confiance d'annotation de DisMo (*dismoconfidence*) ; (4) la présence d'éventuels commentaires des transcripteurs (*comment*) ; (5) la présence de ponctuation avant ou après le lexème (*punctation_before*; *punctuation after*) ; (6) les éventuelles disfluences associées aux unités minimales recherchées (*disfuency*).

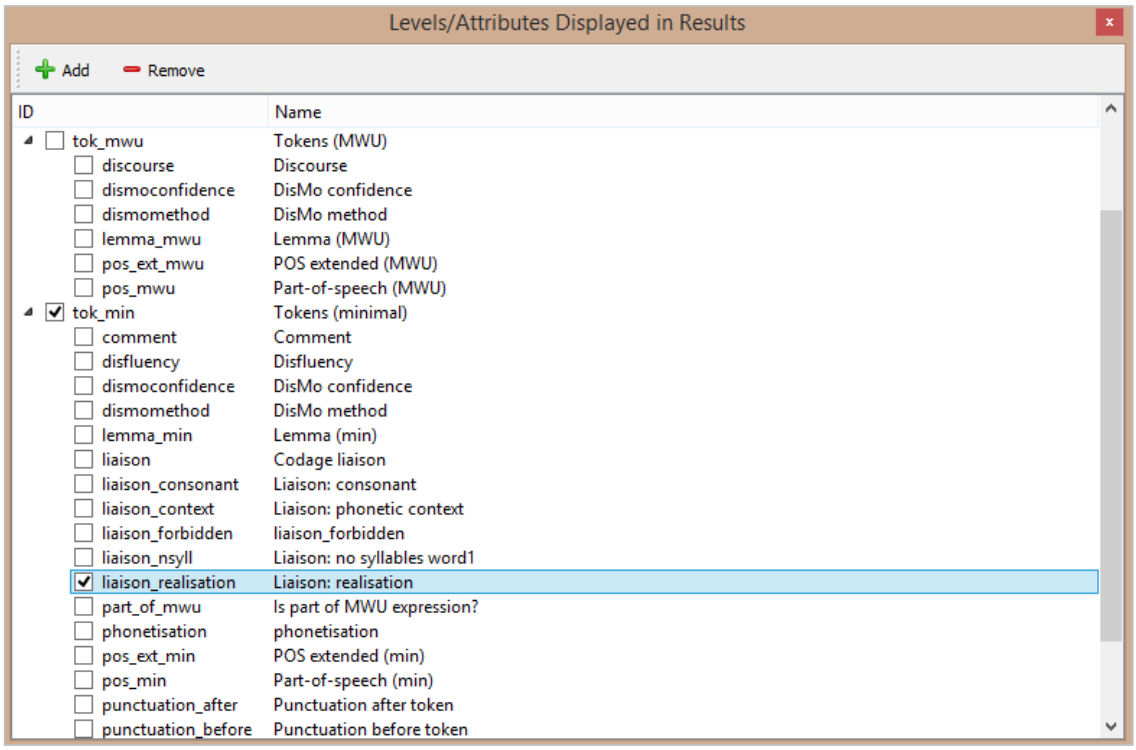


Figure 43 : Champ de sélection des attributs de recherche à afficher dans les résultats

Comme nous le verrons dans les prochaines pages, cette hiérarchie d'attributs assure une analyse de la liaison à la fois au niveau des unités lexicales minimales et des unités poly-lexicales.

Après avoir défini les attributs que l'utilisateur souhaite afficher dans les résultats des concordances, il est possible de lancer la requête en cliquant sur le bouton *Search for occurrences* (« rechercher les occurrences »). La recherche peut prendre quelques minutes à cause des dimensions importantes du corpus PFC.

2.2.2.4 Affichage des résultats

Les résultats de la requête s'affichent dans la partie inférieure de l'interface du concordancier sous forme de tableaux. Chaque ligne représente une concordance[329] et les informations concernant les attributs sélectionnés lors de l'étape précédente (cf. ci-dessus) sont organisées en colonnes.

La figure ci-dessous (cf. Figure 44) représente les résultats de la requête « PRO:per:sjt » - lemme du verbe « être » - « en » - « train » que nous avons illustrée dans les pages précédentes (cf. Figure 42) lors de la sélection des attributs de requête.

[329] Deux modalités d'affichage des résultats sont disponibles : l'une présente une concordance par ligne (*One line per occurrence (compact format)*) et l'autre utilise plusieurs lignes pour une seule occurrence (*Multiple lines per occurrence (expanded format)*) en fonction du nombre d'attributs que l'utilisateur a choisi d'afficher.

Left Context	tok_min 1	liaison_realisation 1	tok_min 2	liaison_realisation 2	tok_min 3	tok_min 4	Right Context
<all>	<all>	<all>	<all>	<all>	<all>	<all>	<all>
qu' il a dit dans sa campagne et _ maintenant	il		est	0	en	train	de vivre un monde nouveau et en vivant dans le
pas pu obtenir gain de cause _ Alors il était	il		était	0	en	train	d' entreprendre des des démarches _ et en vue d'
c' est vrai en quatre vingt dix Un peu plus	J'		étais	0	en	train	de chercher puis je coinçais _ Trois _ Et ben
Et ce j/ celui là _ cette nuit là _	il		était	0	en	train	de de travailler dans son dans son cabinet _ je
cas pour moi c' est difficile de de savoir si	ils		sont	0	en	train	de rigoler de se reprocher des choses de hum _
Alors voilà le gros problème le gros problème _ Donc	on	1	est	0	en	train	de d' essayer de faire quelque chose _ Avec le
_ tu vas les remplir _ Voilà _ Donc euh	nous	1	étions	0	en	train	de dire que _ ici les conditions d' étude sont
Ben ça dépend pour le moment on sait pas encore	on	1	est	0	en	train	de _ de réfléchir à tout ça _ Les plans
ai soulevé la question ça fait plusieurs fois _ Donc	on	1	est	0	en	train	de de d' essayer de résoudre _ ce gros gros
lui il l' aime pas trop l' école et justement	on	1	est	0	en	train	de lui chercher des stages euh on sait pas encore

Figure 44 : Affichage des résultats de la requête « PRO:per:sjt » - lemme du verbe « être » - « en » - « train »

Pour des raisons d'espace, nous n'avons choisi d'afficher que l'attribut *liaison_réalisation* des unités lexicales minimales recherchées. La présence de deux contextes de liaison dans la concordance recherchée explique la présence de deux colonnes différentes de l'attribut *liaison_réalisation*, une pour le contexte PRO:per:sjt - VER (« être ») (*liaison_réalisation 1*) et l'autre pour la séquence VER (« être ») - « en train de » (*liaison_réalisation 2*). Les valeurs (« 0 » et « 1 ») contenues dans ces colonnes font référence au codage de la liaison adopté par le protocole PFC (cf. Partie 2.Chapitre 1.1.3.3) et correspondent respectivement à des non-réalisations (valeur « 0 ») et à des réalisations de la liaison (valeur « 1 »).

Les données issues de cette requête montrent clairement la tendance déjà mise en évidence dans l'analyse fréquentielle du corpus PFC (cf. Partie 2.Chapitre 2.2.2.1), à savoir que la liaison est plus fréquente après les formes verbales du présent du verbe « être » (« est » et « sommes ») qu'après les autres temps verbaux (« était » et « étions »).

Dans les résultats on observe également la présence de colonnes contenant des informations associées aux concordances et portant sur : (a) l'identifiant de l'enregistrement (*Communication ID*) ; (b) l'identifiant de l'annotation (*Annotation ID*) ; (c) l'identifiant du locuteur (*Speaker ID*) ; (d) le contexte gauche de la concordance (*Left Context*) ; (e) le contexte droit de la concordance (*Right Context*). Ces attributs apparaissent par défaut dans les résultats sans besoin de les sélectionner préalablement. Toutes les colonnes peuvent être cachées en positionnant le curseur sur l'en-tête et en effectuant un clic droit. Il suffit de sélectionner l'option *Hide* « nom de la colonne » (« Cacher + « nom de la colonne ») pour que l'attribut disparaisse. Une petite flèche positionnée au-dessous de chaque colonne permet également de trier les concordances selon l'attribut selectionné.

Un bouton *Export* situé sur la gauche de la fenêtre des résultats permet d'exporter les concordances en format *csv*.

Dans les pages qui suivent, nous allons proposer des exemples de quelques requêtes afin d'illuster le fonctionnement du concordancier

2.2.2.5 Exemples de requêtes

Dans les exemples ci-dessous (cf. Figure 45), nous pouvons observer les résultats de la recherche de concordances de la séquence « pas-ADJ » dans laquelle nous avons combiné d'une part l'attribut *Token (min)* (correspondant à l'adverbe « pas ») et de l'autre l'attribut *Pos (min)* (correspondant à l'étiquette morphosyntaxique ADJ à savoir toutes occurrences d'un adjectif).

Results

Speaker ID	Left Context	tok_min 1	liaison_realisation 1	tok_min 2	Right Context
<all>	<all>	<all>	<all>	<all>	<all>
acama1gg_MA	les grandes euh métropoles quoi _ Mouais ici c' est	pas	0	évident	hein _ oui que vraiment de la religion _ voilà
btamp2lg_BT	les offres d' emploi mais évidemment euh _ C' est	pas	0	évident	évident hein donc c' est pour ça qu' après avoir
acama1lg_MA	puis euh pfouh j' ai arrêté parce que _ c' était	pas	0	évident	il y avait pas de ligne aussi _ jusqu' à
61abm1lg_BM	ce qu' il va se passer hein parce que c' est	pas	0	évident	on est dans des périodes _ un peu difficiles et
61abm1lg_BM	est comme ça _ Mais bon c' est peut être	pas	0	évident	non plus _ Peut être pas quand même J' espère
61abm1lg_BM	quand j' ai eu mon permis mais avant euh _	pas	0	évident	hein CM l' as eu à quel âge le permis
42acd1gg_C	j' ai été sévère avec mes enfants mais pas étriquée	pas	0	étroite	à ce point là Faut quand même pas exagérer hein
42acd1gg_C	pense que j' ai été sévère avec mes enfants mais	pas	0	étriquée	pas étroite à ce point là Faut quand même pas
75ccm1gg_CM	dis ça hein comme ça mais ça p/ ça serait	pas	0	étonnant	_ Si tu veux être un peu hein donc pas
11adp1gg_DP	deux cents heures je crois que c' était c' était	pas	0	énorme	hein mais c' était toujours ça _ Et alors on
btanh1lg_BT	_ Mais tu sais un bac en philo c' est	pas	0	énorme	euh c' est pas c' est pas super compliqué enfin
61abd1lg_HD	I/ le courant voulu _ Ben alors _ Ils sont	pas	0	élus	hein et le Gouvernement est est en route _ il

Figure 45 : Concordances issues de la requête « pas-ADJ » (attributs : Token(min)-POS(min))

Les occurrences calculées sont donnés sous forme de tableaux qui présentent en colonnes : (a) le code du locuteur (*Speaker ID*); (b) les intervalles Praat (*tMin* et *tMax)*; (c) le contexte gauche (*Left context*); (d) le mot gauche de liaison (M1) (*tok_min 1*); (e) le type de réalisation de la liaison *(liaison_réalisation 1)* ; (f) la catégorie morphosyntaxique du mot gauche (*pos_min 1*); (g) le lemme du mot gauche (*lemma_min 1*); (h) le mot droit de liaison (M2) (*tok_min 2*); (i) la catégorie morphosyntaxique du mot droit (*pos_min 2*); (j) le lemme du mot droit (*lemma_min 2*); (k) le contexte droit (*Right context*).

Les résultats de cette requête permettent à l'utilisateur de vérifier le faible pourcentage de réalisation de la liaison après l'adverbe « pas », qui l'oppose aux autres adverbes monosyllabiques (cf. Partie 2.Chapitre 2.2.2.2.5).

Il en va de même pour la recherche des cooccurrences par lemme. Dans l'exemple ci-dessous (Figure 46), nous pouvons observer les résultats de la recherche de « être-DET».

Cette requête permet d'afficher les différentes formes verbales du verbe « être » suivi d'un déterminant.

Results

Speaker ID	Left Context	tok_min 1	liaison_realisation 1	tok_min 2	Right Context
<all>	<all>	<all>	<all>	<all>	<all>
svayb1gg_YB	_ Donc euh bon le premier là c' é/ c'	était	0	une	femme qui s' était suicidée _ Ça faisait sept heures
ciasb1lg_CIASB1L	appelle _ ça c' est c' est c' est c'	est	0	une	bonne chose _ Et aussi on peut ouvrir le pays
44amr1gg_MR	émanant des syndicats chrétiens Des euh _ Et euh c'	était	0	une	euh construction euh le le le _ Construction à euh
bgaad1lg_BG	allez vis à vis des jeunes _ Hein et c'	est	1	une	fille allez elle est en en terminales comment je vais
12acr1gg_CR	sixième euh j' habitais donc euh les Costes Rouges c'	est	0	une	cité Avec euh des H.L.M on habitait en H.L.M _
acakb1lg_KB	bien sûr oui _ Voilà c' est ça _ C'	est	1	une	relation de mes parents _ je vois normal le pan
54bgc1lg_CG	heure _ que d' être après l' heure _ C'	est	0	une	impolitesse _ si vous êtes en retard _ (inspiration) vous
11atg1gg_TG	pratique (rire) Ça aide _ Ouais ouais _ Non c'	est	0	une	automatique _ pour accélérateur et frein bon moi je me
54bjl1gg_JL	euh _ C' est pas une maison neuve hein c'	est	0	une	maison qui date des années dix sept cents _ qui
btafb1gg_BT	mal de zéros je collectionnais les zéros en dictée j'	étais	0	une	vraie star quoi _ Les sciences euh j' étais pas
21amb1lg_MB	_ Mais par contre dès qu' ils le reçoivent c'	est	0	une	question d' heures presque des fois euh pour euh entre

Figure 46 : Concordances issues de la requête « être-DET » (attributs : Lemma-POS(min))

Pour terminer, l'exemple ci-dessous (Figure 47) nous permet d'observer les résultats de la recherche de « ADJ-NOM ».

Results

Speaker ID	Left Context	tok_min 1	liaison_realisation 1	tok_min 2	Right Context
<all>	<all>	<all>	<all>	<all>	<all>
11amg2gg_MG2	un dentiste avec sa femme qui ont trois enfants en	bas	1	âge	_ Voilà de la maison des enfants un peu tout
acazb1lg_E	faisaient les femmes algériennes qui travaillent avec leurs enfa...	bas	1	âge	_ Il y a des crèches quand même _ En
81aar2gg_AR2	crèmes des euh _ J' aurais beaucoup aimé avoir des	petits	1	échantillons	à vous donner _ Mais notre associatiat/ notre euh _
44ajs1lg_JS	se lever de bonne heure Hein il fallait avoir de	bonnes	1	échelles	hein _ Alors c' était là bah c' était (X)
85ajf1gg_JF	formation euh j' étais séminariste _ J' ai fait les	grandes	1	écoles	_ Euh j' étais séminariste parce que ça c' est lié
sgapb1gg_PB	étaient quand même contents que leurs gamins puissent suivr...	bonnes	1	écoles	_ Mais ce qui est venu par la suite eux
75cvl1lg_VL	sais pas si tu connais ma copine (X) _ Des	autres	1	écoles	dont (X) aussi et elle l' avait fait et elle
85amg1gg_MG	alors que il y a plein d' autres euh d'	autres	1	écoles	euh sur Paris _ Hein _ Ben il y en
61ajb1gg_JB	Et donc là ça c' est il y a un	bon	1	écoulement	là de ce côté là hein _ Oh c' est

Figure 47 : Concordances issues de la requête « ADJ-NOM » (attributs : POS(min)-POS(min))

Le concordancier permet également d'écouter les intervalles des enregistrements contenant les concordances issues des requêtes réalisées à travers la combinaison des trois attributs que nous venons d'évoquer.

3. Conclusion

Étant données les typologies d'erreurs de liaison observées dans l'ensemble des études menées sur l'acquisition de la liaison en FLE, nous avons fait l'hypothèse que le traitement de la CL comme consonne finale du M1 dans la plupart des manuels pourrait, d'une part provoquer un retard dans la mise en place de généralisations phonologiques de la liaison et d'autre part donner lieu à des réalisations orthographiques des CL. Comme le soulignent Racine et Detey (sous presse), la production de formes orthographiques pourrait sans doute être favorisée par la

place marginale qui est souvent accordée à l'enseignement des correspondances entre la graphie et la phonie dans les manuels d'enseignement du FLE.

Quant à l'enseignement et à l'acquisition de la variation de la liaison, compte tenu des résultats de notre étude (cf. Partie 4.Chapitre 2.1.6.2), nous avons suggéré que l'enseignement de la liaison variable en FLE pourrait commencer par une présentation des séquences de liaison le plus fréquentes, attestées par l'analyse fréquentielle de la liaison que nous avons menée sur le corpus PFC (cf. Partie 2.Chapitre 2.2.2.2.8). Celles-ci pourraient constituer un point de départ pour l'acquisition des usages de liaison les plus fréquents, indépendamment du style de parole.

Afin de mener une étude plus approfondie de la variation de la liaison (pendant les heures d'enseignement ou en apprentissage autonome) nous avons proposé deux ressources complémentaires (notamment une classification et un concordancier multi-niveaux), basées sur le corpus de français oral PFC (Durand, Laks et Lyche 2009), pour l'enseignement et l'acquisition de la liaison.

Conclusions générales

Cette thèse s'est consacrée, dans le cadre du projet international « InterPhonologie du Français Contemporain » (IPFC) (Detey et Kawaguchi 2008 ; Racine et al. 2010 ; Racine, Detey et Kawaguchi 2012), à l'étude longitudinale de l'acquisition de la liaison en français langue étrangère (FLE) auprès d'apprenants italophones. Compte-tenu de sa valeur sociolinguistique ainsi que de sa fréquence dans la production orale, la liaison représente un aspect crucial de l'acquisition du français langue seconde et étrangère. Néanmoins l'hétérogénéité et la variabilité qui caractérisent ce phénomène de *sandhi* externe constituent une source de complexité multiple pour l'enseignement et l'apprentissage de la liaison (en particulier en milieu hétéroglotte où l'acquisition ne bénéficie que dans une faible mesure de l'apport de l'input oral).

Les difficultés associées à l'acquisition de la liaison que nous avons observées dans notre corpus d'apprenants («corpus IPFC-italien Rome ») se situent d'une part au « niveau de la macro-planification » (Racine 2014), qui implique la maîtrise ainsi que la connaissance de la variation sociolinguistique de la liaison, et d'autre part à un « niveau de micro-planification » (Racine 2014) qui comprend les aspects phonético-phonologiques et prosodiques de réalisation de la liaison. D'un point de vue didactique, la plupart des ressources didactiques renvoient le plus souvent à une description prescriptive et normative des usages de la liaison (Fouché 1959 ; Delattre 1947) qui fait appel à des tableaux ou à des listes de règles construites sur la base de critères morphosyntaxiques (Racine et Detey sous presse). Ces classifications sont toutefois dépourvues de valeur explicative (Mallet 2008) et prédictive (Boula de Mareüil et al. 2003) de la réalisation de la liaison puisqu'elles négligent l'analyse des autres dimensions de variation de la liaison (Mallet 2008 ; Durand et al. 2011 ; Durand et Lyche sous presse).

Dès le début de ce travail en effet, la présentation d'un cadre descriptif et diachronique de la liaison (cf. Partie 1.Chapitre 1) nous a permis de mettre en évidence les limites d'une description de la liaison basée sur les seuls critères morphosyntaxiques. Si le bref parcours historique que nous avons retracé dans le premier chapitre nous a permis d'évoquer la distinction entre des liaisons morphosyntaxiques d'une part (ex. « les amis » ; « de grands amis ») et des liaisons phonologiques d'autre part (ex. « chez un ami »), nous avons vu dans les pages suivantes (cf. Partie 1.Chapitre 2) que cette opposition n'a pas de valeur explicative dans l'étude de la liaison. Autrement dit, non seulement le facteur syntaxique ne permet pas d'expliquer les comportements variables de la liaison mais il ne parvient pas non plus à distinguer les liaisons

qui sont toujours réalisées (par tout locuteur et dans tout contexte) (à savoir les liaisons catégoriques) de celles dont la réalisation est variable (à savoir les liaisons variables).

Ces observations nous ont donc amenés à consacrer la première partie de ce travail à un bref survol des multiples dimensions de variation de la liaison (cf. Partie 1.Chapitre 2), aspect qui nous a obligé à nous interroger sur les modèles de prononciation de référence et sur les approches à adopter dans la description des usages de la liaison. Un bref parcours des modèles de prononciation du français contemporain proposés au cours du XXème siècle montre que, suite aux processus de « démocratisation » et de « décentralisation » qui ont caractérisé l'évolution de la norme, les caractéristiques phonético-phonologiques (y compris la liaison) du français de référence ne sont encore aujourd'hui pas définies. Comme la norme de référence, les approches méthodologiques ont également connu le passage d'une tradition normative (Grammont 1914 ; Fouché 1959) d'étude de la liaison basée sur les intuitions des linguistes, au développement de travaux s'appuyant sur des résultats d'enquête (Ågren 1973 ; Laks 1977 ; Morin et Kaye 1982 ; Lucci 1983 ; De Jong1988 ; Encrevé 1988) de plus en plus systématiques. Parmi celles-ci, les recherches menées dans le cadre du projet « Phonologie du Français Contemporain » (PFC) (Durand, Laks et Lyche 2009) ont contribué à un renouvèlement des descriptions et des traitements classificatoires de la liaison (Durand et Lyche 2008 ; Mallet 2008 ; Durand et al. 2011) (Partie 1.Chapitre 4). De même Encrevé (1988) a été l'un des premiers à montrer l'intérêt d'une étude sur corpus, en soulignant le fait que seul un modèle capable de distinguer le comportement de la liaison de celui de l'enchaînement pouvait espérer rendre compte de la spécificité de ce phénomène. Néanmoins pendant une longue période, et parfois encore aujourd'hui, les modèles phonologiques élaborés pour l'étude de la liaison ne se sont pas enrichis des apports pluridisciplinaires issus de la linguistique du corpus ou de la psycholinguistique (cf. Chapitre VI) en faisant uniquement appel à des sources prescriptives alimentées par l'introspection des linguistes (Laks 2008) (cf. Partie 1.Chapitre 5).

En ligne avec les travaux précédents menés dans le cadre du corpus PFC (Durand et al. 2011; Laks, Calderone et Celata 2014), nous avons développé dans une deuxième partie (cf. Partie 2) de ce travail une analyse fréquentielle de la liaison au sein du corpus PFC (Durand, Laks et Lyche 2009). Cette étude, précédée d'un important travail de réannotation morphosyntaxique (Barreca et Christodoulides 2015), a abouti à une description plus fine de la variation de la liaison pouvant servir de modèle de référence pour l'étude des productions des apprenants de FLE (cf. Partie 3 ; Partie 4). D'une part, ce travail a contribué à rendre compte des idiosyncrasies internes aux contextes morphosyntaxiques de liaison dépassant ainsi les limites d'une classification basée uniquement sur les catégories morphosyntaxiques

d'appartenance des mots de liaison. D'autre part, il a mis en évidence la distribution fréquentielle (distribution de Zipf) de la liaison variable, ce qui remet en question le traitement unitaire de la liaison proposé par les approches prescriptives et par les modèles phonologiques génératifs (cf. Côté 2005). En effet, si la liaison après les conjonctions et les prépositions monosyllabiques semble être dans l'ensemble assez homogène, d'autres cas, comme la liaison après des adverbes monosyllabiques, demandent une analyse syntaxique plus profonde. D'autres types de liaison semblent en revanche principalement conditionnés par des facteurs d'ordre lexical (ex. la liaison après des formes verbales) (Côté 2013). En fin de compte, la liaison après les adjectifs épithètes, à laquelle nous avons consacré une étude supplémentaire, est sujette à des contraintes d'ordre distributionnel qui limitent le nombre d'adjectifs susceptibles d'être antéposés à l'oral (Benzitoun 2014). Cependant, ni la distribution ni la probabilité transitionnelle (Côté 2013) n'influent sur la productivité de la liaison puisque la liaison n'est pas moins fréquente après des adjectifs majoritairement antéposés qu'après des adjectifs en distribution complémentaire (Barreca et al. 2015). Ces conclusions, bien évidement partielles, représentent toutefois un enjeu d'importance pour l'acquisition et la didactique de la liaison en français langue seconde ou étrangère qui ont été au cœur des dernières sections de ce travail (cf. Partie 3 ; Partie 4 ; 0). De plus, comme nous l'avons déjà observé précédemment, les résultats de l'analyse fréquentielle de la réalisation de liaison menée sur les données du corpus PFC nous ont permis de disposer d'un modèle de référence nécessaire pour l'étude des productions des apprenants de FLE.

En effet, dans la troisième partie de ce travail, après une introduction des repères théoriques sur l'acquisition de la liaison en français L2 (cf. Partie 3), la présentation des études précédentes consacrées à la liaison (cf. 0) a été l'occasion de montrer encore une fois l'intérêt du recours à un corpus linguistique de francophones pour l'analyse des réalisations de la liaison en FLE (Racine et Detey sous presse). En effet, la plupart des travaux précédents ont évalué les productions des apprenants sur la base des descriptions prescriptives de la liaison (cf. ci-dessous) parfois éloignées des usages réels en vigueur chez les locuteurs natifs (Racine et Detey sous presse). C'est à ces mêmes sources, comme nous l'avons déjà observé précédemment, que se réfèrent la plupart des manuels pour l'enseignement de la liaison (cf. Partie 3.Chapitre 3). Néanmoins, comme le soulignent Racine et Detey (sous presse), dans certains manuels (Lauret 2007 ; Pustka 2011) la description de la liaison (au moins au niveau de la « macro-planification ») a été établie sur la base des résultats des études menées sur les corpus oraux. Il convient pourtant de souligner que malgré des avantages indéniables (Chambers 2009 ; Tyne 2009 ; Tchobanov et al. 2009 ; Detey 2010 ; Detey et Racine 2012), le recours aux corpus dans la

didactique multiplie les variétés d'usages de la liaison en ravivant le débat sur la notion de français de référence. Dans ce cadre, nous comprenons pourquoi les descriptions prescriptives représentent une tentative de « mettre de l'ordre » (Fouché 1959) puisqu'elles assurent un traitement didactique de la variation de la liaison limité aux usages de l'élite cultivée parisienne. Bien que plus économique, une telle approche renonce à tout enseignement de la variation sociolinguistique associée à la liaison. Galazzi (2008a), Detey (2010) et Detey et Racine (2012) estiment qu'un seul modèle de prononciation doit donc être présenté dans un premier stade de l'apprentissage afin d'accorder une primauté à un input unique et « standard » qui sera par la suite enrichi par l'apport des usages issus des autres variétés. L'apprenant sera ainsi confronté et sensibilisé à une variété sociolinguistique qui favorise le développement de son « acceptabilité » sociolinguistique.

Ces propositions nous paraissent apporter des réponses adéquates aux besoins de la didactique et de l'acquisition de la liaison permettant d'une part de sensibiliser les apprenants aux variétés d'usage de la liaison et de l'autre d'éviter un décalage entre les modèles de référence pour l'analyse des productions des apprenants et les usages réels des locuteurs.

C'est dans cette démarche que se situent les travaux menés au sein du projet IPFC, dont la présente étude conduite auprès d'apprenants italophones de niveau intermédiaire de FLE (cf. Partie 4) fait partie. À l'issue de ce travail, plusieurs données nous paraissent émerger dont nous proposons un bilan ci-dessous.

En ce qui concerne les liaisons catégoriques, les apprenants italophones produisent la liaison dans la plupart des contextes catégoriques dans toutes les tâches (sauf au sein des expressions figées et des pronoms clitiques) ce qui rejoint les conclusions des études précédentes (entre autres Tennant sous presse ; Falbo sous presse). Cependant un certain nombre de non-réalisations de liaisons catégoriques persistent même dans les dernières phases de la période d'observation sans doute à cause du traitement lexical (Wauquier 2009) que la liaison reçoit dans les premières étapes de l'enseignement et de l'acquisition. En effet, dans la plupart des syllabus pour l'enseignement du FLE, l'acquisition de la morphologie des différentes catégories morphosyntaxiques précède le plus souvent l'enseignement de la liaison. Dans le cadre de la théorie de la processabilité de Pienemann (1998), nous avons donc formulé l'hypothèse selon laquelle l'acquisition préalable des marques morphosyntaxiques au niveau des catégories principales serait en mesure de ralentir l'acquisition de la liaison et de la rendre redondante. Autrement dit, le passage d'un marquage au niveau catégoriel (« procédure de catégorie » *category procedure*) aux accords entre constituants d'un syntagme (« procédure intra-syntagmatique » *phrasal procedure*) ou encore entre syntagmes (« procédure inter-

syntagmatique » *S-procedure*) pourrait être ralenti par le fait que les informations morphosyntaxiques sont déjà encodées sous forme de désinences au niveau du lexème. Cela rendrait la production de la liaison redondante d'un point de vue fonctionnel. Cet effet pourrait être amplifié, en contexte d'apprentissage hétéroglotte et institutionnel, par une faible exposition à l'input oral ainsi que par l'absence d'un réel besoin communicatif. Cependant, plusieurs éléments semblent suggérer que l'apprenant est en mesure de développer des représentations phonologiques de la liaison sous forme de constructions. En effet, si le nombre de non-réalisations se réduit au cours de l'étude longitudinale (sans pourtant disparaître totalement), des formes de liaisons non attendues et de substitutions émergent surtout à partir de la deuxième séance de la période d'observation. Ces exemples montrent que l'apprenant est en mesure de dépasser un simple traitement « au cas par cas » des contextes de liaison (Wauquier 2009). En outre, une acquisition basée sur les seules stratégies lexicales présupposerait que dans les contextes DET-NOM et PRO-VER les apprenants produisent un taux plus important de liaison dans les séquences ayant des indices de cooccurrence importants et en présence d'un input écrit. Nous n'avons en revanche observé aucun effet significatif ni de fréquence de cooccurrence de la séquence de liaison ni de tâche (lecture oralisée/parole semi-spontanée) dans le cas des réalisations des contextes DET-NOM et PRO-VER. D'autres éléments semblent également appuyer l'hypothèse selon laquelle la liaison suit un itinéraire d'acquisition articulé en échelles implicationnelles. En effet, le taux élevé de non-réalisations associé aux pronoms clitiques pourrait s'expliquer par l'acquisition tardive de ces éléments (Bartning et Schlyter 2004) pour lesquels aucun schéma de réalisation de la liaison n'est accessible chez l'apprenant de niveau intermédiaire.

Quant à la réalisation des liaisons variables, les résultats de la présente étude montrent que les apprenants de niveau intermédiaire non seulement ne maîtrisent pas la variabilité sociolinguistique associée à la liaison (ce qui rejoint les résultats recueillis auprès d'autres sujets débutants et intermédiaires) (Galazzi et al. 2013 ; Falbo et al. sous presse), mais ils font également état de faibles connaissances épilinguistiques de la variation (Gombert 1990). En effet, les résultats du test de jugement montrent que si d'une part les apprenants sont conscients de la nature obligatoire des contextes de liaison catégorique, d'autre part ils ont souvent tendance dans les cas des liaisons variables à choisir comme seule variante correcte celle avec réalisation de la liaison. Ces données confirment les résultats des études précédentes (Mougeon, Nadasdi et Rehner 2002, Gadet et Tyne 2007) qui ont montré que les apprenants en milieu institutionnel sur-généralisent les variantes qu'ils jugent les plus standards.

En ce qui concerne les difficultés rencontrées par les apprenants au niveau de la « micro-planification » (Racine 2014) dans la production, notre étude s'est penchée en particulier sur les productions orthographiques des CL et sur les liaisons réalisées non-enchaînées. Les analyses menées sur la première typologie d'erreur ont relevé la présence de réalisations orthographiques non seulement dans les lectures oralisées mais également dans des tâches de production orale semi-spontanée et de repétition ainsi que du choix des variantes orthographiques dans le test de jugement. Afin de rendre compte de ces données, nous avons postulé dans le cadre du modèle interactif bimodal (Grainger et Ferrand 1996) que les représentations orthographiques et phonologiques de la liaison interagissent entre elles dans les productions des apprenants. Cela pourrait s'expliquer encore une fois par le fait qu'au début de leur acquisition les apprenants ont encodé les informations morphosyntaxiques au niveau des lexèmes (« procédure de catégorie » *category procedure*). C'est seulement au cours d'une des étapes suivantes que des échanges entre les différentes unités commencent à être mises en place entres constituants d'un syntagme (« procédure intra-syntagmatique » *phrasal procedure*) ou encore entre syntagmes différents (« procédure inter-syntagmatique » *S-procedure*). Au début de cette phase, nous pouvons supposer que les apprenants commencent à produire la liaison en s'appuyant sur les seules représentations phonologiques (ex. /de/) et orthographiques (ex. < des >) des unités dont ils disposent. Autrement dit, dans les premières phases du développement (« procédure de catégorie » *category procedure*), la liaison peut être envisagée comme le résultat de l'influence des représentations orthographiques sur les représentations phonologiques qui, en cas de non correspondance entre consonnes graphiques et CL, donnent lieu à des réalisations orthographiques de la CL (ex. « des enfants » [desɑ̃fɑ̃]) tant en répétition qu'en production (lectures et parole semi-spontanée).

Quant à la production de liaisons non-enchaînées, l'absence de formes de liaisons non-enchaînées en parole semi-spontanée semble aller dans la direction des conclusions déjà formulées en L1 (Durand et al. 2011) ainsi que par certaines recherches en L2 (Racine sous presse ; Tennant sous presse ; Andreassen et Lyche sous presse), à savoir que les non-enchaînements représenteraient un phénomène marginal dans la production de la liaison relevant de l'influence de l'input orthographique dans la lecture. En effet, le faible nombre de liaisons non-enchaînées attesté chez les apprenants italophones ne permet pas de vérifier l'hypothèse d'après laquelle les apprenants ayant une L1 caractérisée par un système syllabique ouvert produisent moins de liaisons non-enchaînées que des apprenants ayant une L1 caractérisée par un système syllabique fermé (Thomas 2002; Wauquier 2009; Tennant sous

presse). Nous ne pouvons qu'observer que les sujets italophones ne rencontrent pas de difficulté à resyllaber la CL.

Dans une démarche didactique, vu l'ensemble des difficultés rencontrées par les apprenants de FLE dans la production de la liaison (surtout la liaison variable), nous avons développé grâce aux résultats issus de l'analyse fréquentielle du corpus PFC deux types de ressources didactiques pour l'enseignement de la liaison en FLE (cf. 0). D'une part une classification multi-niveaux de la liaison réalisée par le biais de graphiques *drill-down* et d'autre part un concordancier multi-niveaux (Barreca et Christodoulides 2015) pour la recherche et l'écoute d'occurrences de liaison issues du corpus PFC (Durand, Laks et Lyche 2009).

Nous espérons que ces ressources contribueront, dans le cadre d'une démarche *data-driven learning* (apprentissage sur corpus), à renouveler l'acquisition et l′enseignement de la liaison et à réduire ainsi le décalage entre les descriptions adoptées dans l'évaluation et l'enseignement de la liaison et les usages en vigueur chez les natifs (Detey et Racine sous presse).

Références bibliographiques

ABRY Dominique et CHALARON Marie-Laure, 2011, *Les 500 exercices de phonétique : niveau B1/B2*, Paris, Hachette.

ABRY Dominique et CHALARON Marie-Laure, 2009, *Les 500 exercices de phonétique : niveau A1/A2*, Paris, Hachette.

ABRY Dominique et CHALARON Marie-Laure, 1994, *Phonétique : 350 exercices*, Paris, Hachette.

ADDA-DECKER Martine, BOULA DE MAREÜIL Philippe et LAMEL Lori, 1999, « Pronunciation Variants in French: schwa & liaison », Proceedings of 14th International Congress of Phonetic Sciences, ICPhS1999, San Francisco, États-Unis, p. 2239‑2242.

ADDA-DECKER Martine, FOUGERON Cécile, GENDROT Cédric, DELAIS-ROUSSARIE Elisabeth et LAMEL Lori, 2012, « La liaison dans la parole spontanée familière : une étude sur grand corpus », *Revue française de linguistique appliquée*, 2012, Vol. XVII, n° 1, p. 113‑128.

ADDA-DECKER Martine, NEMOTO Rena et DURAND Jacques, 2009, « Stratégies de démarcation du mot en français: une étude expérimentale sur grand corpus », Actes des Journées Linguistiques de Nantes, Journées Linguistiques de Nantes 2009, Nantes, France, p. 91‑96.

ADDA-DECKER Martine et SNOEREN Natalie D., 2011, « Quantifying temporal speech reduction in French using forced speech alignment », *Journal of Phonetics*, 2011, vol. 39, n° 3, p. 261‑270.

ÅGREN John, 1973, *Étude sur quelques liaisons facultatives dans le français de conversation radiophonique. Fréquences et facteurs*, Uppsala, Acta Universitatis Upsaliensis, 10.

ALARIO François-Xavier, 2008, « Le choix des mots: Processus psycholinguistiques de la production de langage », URL: http://lpc.univ-amu.fr/IMG/UserFiles/Images/xalario/alario2_copy.pdf, consulté le 26 août 2015.

ALARIO François-Xavier, PERRE Laetitia, CASTEL Caroline et ZIEGLER Johannes C., 2007, « The role of orthography in speech production revisited », *Cognition*, 2007, vol. 102, n° 3, p. 464‑475.

ALLEN Sidney W., 1962, *Sandhi: the theoretical, phonetic, and historical bases of word-junction in Sanskrit*, The Hague, Mouton de Gruyter.

ANDERSON Shannon, 1975, « On the Syntax of Liaison », *French Review*, 1975, vol. 48, n° 5, p. 848‑855.

ARNON Inbal et RAMSCAR Michael, 2012, « Granularity and the acquisition of grammatical gender: how order of acquisition affects what gets learned », *Cognition*, 2012, vol. 122, n° 3, p. 292‑305.

ASHBY William, 1981, « French Liaison as a Sociolinguistic Phenomenon » dans Cressey. W. et Napoli, D. (éds), *Linguistics Symposiun on Romance Languages*, Washington D.C., Georgetown University Press, p. 46‑57.

AUROUX Sylvain et WEIL Yvonne, 1991, *Dictionnaire des auteurs et des thèmes de la philosophie*, Paris, Hachette.

AVANZI Mathieu, 2014, « A Corpus-Based Approach to French Regional Prosodic Variation », *Nouveaux cahiers de linguistique française*, 2014, vol. 31, p. 309‑323.

BAAYEN Harald R., 2001, *Word frequency distributions*, Dordrecht, Kluwer Academic.

BAGOU Odile, 2006, *Euristique de segmentation aux débuts de syllabes, stratégie de segmentation prosodique: quelle place dans la segmentation de la parole en français?*, Thèse de doctorat, Université de Genève, Genève.

BAGOU Odile et LAGANARO Marina, 2009, « On the production of sandhi phenomena in French: Psycholinguistic and acoustic data », Proceedings of 10th Annual Conference of the International Speech Communication Association, INTERSPEECH2009, Brighton, Royaume Uni, p. 456‑459.

BARONI Marco, 2008, « Distributions in text » dans Lüdeling, A. et Kytö, M. (éds), *Corpus linguistics an international handbook*, Berlin, Mouton de Gruyter.

BARONI Marco et EVERT Stefan, 2014, « The zipfR package for lexical statistics: A tutorial introduction », URL: http://zipfr.r-forge.r-project.org/materials/zipfr-tutorial.pdf, consulté le 14 février 2015.

BARONI Marco et EVERT Stefan, 2007, « zipfR: Word frequency distributions in R », Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Posters and Demonstrations Sessions, ACL2007, Prague, République Tchèque, p. 29‑32.

BARONI Marco et EVERT Stefan, 2005, « Testing the extrapolation quality of word frequency models », Proceedings of Corpus Linguistics 2005, Corpus Linguistics 2005, Birmingham, Royaume Uni.

BARRECA Giulia et CHRISTODOULIDES George, 2015, « Analyse fréquentielle de la liaison dans un corpus de français parlé », 13ème Rencontres du Réseau Français de Phonologie (RFP2015), 29 juin-1 juillet 2015, Bordeaux, France.

BARRECA Giulia et CHRISTODOULIDES George, à paraître, « Analyse fréquentielle de la liaison variable dans un corpus de français parlé », dans Eychenne, J. et Laks, B. (éds), *Numero spécial de Journal of French Language Studies,* Cambridge, Cambridge University Press.

BARRECA Giulia et FLOQUET Oreste, 2015, « L'acquisition de la liaison chez des apprenants italophones : le corpus IPFC-italien (Rome) », Séminaire MoDyCo « Liaison », 16 avril 2015, Nanterre, France.

BARRECA Giulia, 2014, « Multidimensionnalité de la liaison variable et difficultés de classification. Le cas des adverbes monosyllabiques », Actes des 32èmes Journées d'Étude de la Parole, JEP2014, Le Mans, France.

BARRECA Giulia et CHRISTODOULIDES George, 2014, « Un concordancier multi-niveaux et multimédia pour des corpus oraux », Actes de la 21ème Conférence Traitement Automatique des Langues Naturelles, TALN2014, Marseille, France, p. 499‑504.

BARRECA Giulia, CHRISTODOULIDES George, DURAND Jacques, LAKS Bernard et LYCHE Chantal, 2015, « La liaison de l'adjectif antéposé dans le français parlé », Journées de l'Association of French Language Studies (AFLS 2015), Caen, France.

BARTNING Inge et SCHLYTER Suzanne, 2004, « Itinéraires acquisitionnels et stades de développement en français L2 », *French Language Studies*, 2004, vol. 14, p. 281‑199.

BASSETTI Benedetta, 2007, « Effects of hanyu pinyin on pronunciation in learners of Chinese as a foreign language » dans Guder, A. and Jiang, X. and Wan, Y. (éds), *The Cognition, Learning and Teaching of Chinese Characters*, Beijing, Beijing Language and Culture University Press, p. 156‑179.

BASSETTI Benedetta, 2006, « Orthographic input and phonological representations in learners of Chinese as a foreign language », *Written Language and Literacy*, 2006, vol. 9, nº 1, p. 95‑114.

BAUCHE Henri, 1920, *Le langage populaire: grammaire, syntaxe et dictionnaire du français tel qu'on le parle dans le peuple de Paris, avec tous les termes d'argot usuel*, Paris, Payot.

BEACCO Jean-Claude, BOUQUET Simon et PORQUIER Rémy, 2004, *Niveau B2 pour le français (utilisateur-apprenant indépendant) textes et références*, Paris, Didier.

BENZITOUN Christophe, 2014, « La place de l'adjectif épithète en français : ce que nous apprennent les corpus oraux », Actes du 4ème Congrès Mondial de Linguistique Française, CMLF2014, Berlin, Allemagne, p. 2333‑2348.

BENZITOUN Christophe, FORT Karen et SAGOT Benoit, 2012, « TCOF-POS: un corpus libre de français parlé annoté en morphosyntaxe », Actes de la conférence conjointe JEP-TALN-RECITAL 2012, JEP-TALN-RECITAL 2012, Grenoble, France.

BESSE Henri et PORQUIER Rémy, 1991, *Grammaires et didactique des langues*, Paris, Didier.

BIALYSTOK Ellen et SMITH Michael, 1985, « Interlanguage is not a state of mind: an evaluation of the construct for second-language acquisition », *Applied Linguistics*, 1985, vol. 6, nº 2, p. 101‑117.

BLANCHE-BENVENISTE Claire et CHERVEL André, 1969, *L'Orthographe*, Paris, Maspéro.

BLANC Michel et BIGGS Patricia, 1971, « L'enquete socio-linguistique sur le francais parle a Orleans (A Sociolinguistic Study of French as Spoken in Orleans) », *Francais dans le Monde*, 1971, vol. 85, p. 16‑25.

BLANC Olivier, CONSTANT Matthieu, DISTER Anne et WATRIN Patrick, 2008, « Corpus oraux et chunking », Actes de 15èmes Journées d'Études sur la Parole, JEP2008, Avignon, France.

BOERSMA Paul et WEENINK David, 2015, *Praat: doing phonetics by computer*.

BONAMI Olivier et BOYE Gilles, 2005, « Construire le paradigme d'un adjectif », *Recherches linguistiques de Vincennes*, 2005, nº 34, p. 77‑98.

BONAMI Olivier et BOYE Gilles, 2003, « La nature morphologique des allomorphies conditionnées : les formes de liaison des adjectifs en français » dans Fradin, B. et al. (éds), *Les unités morphologiques. Actes du troisième forum de morphologie*, Lille, Université de Lille 3, p. 39‑48.

BONAMI Olivier, BOYE Gilles et TSENG Jesse, 2005, « Sur la grammaire des consonnes latentes », *Langages*, 2005, nº 158, p. 89‑100.

BOOIJ Geert E, 1984, *French C/Ø-alternations, extrasyllabicity and lexical phonology*, Amsterdam, Vrije Universiteit, Vakgroep Algemene Taalwetenschap.

BOOIJ Geert E. et DE JONG Adrianus, 1987, « The domain of liaison: theories and data », *Linguistics*, 1987, vol. 25, nº 5, p. 1005‑1025.

BORDAL Guri et LYCHE Chantal, 2008, « La liaison en terre africaine », Journées Phonologie du Français Contemporain (PFC), 2008 Paris, France.

BORREL André et BILLIERES Michel, 1989, « L'évolution de la norme phonétique en français contemporain », *La Linguistique*, 1989, vol. 25, nº 2, p. 45‑62.

BOULA DE MAREÜIL Philippe, ADDA-DECKER Martine et GENDNER Véronique, 2003, « Liaisons in French: a corpus-based study using morpho-syntactic information », Proceedings of 13th International Congress of Phonetic Sciences, ICPhS2003, Barcelone, Espagne, p. 1329‑1332.

BOULA DE MAREÜIL Philippe, DIMULESCU Bianca, WOEHRLING Cécile et ADDA-DECKER Martine, 2008, « Caractérisation d'accents étrangers et régionaux en français : vers une identification automatique », *TAL*, 2008, vol. 49, nº 3, p. 135‑163.

BOULA DE MAREÜIL Philippe, ADDA-DECKER Martine et WOEHRLING Cécile, 2007, « Analysis of oral and nasal vowel realisation in northern and southern French varieties », Proceedings of 16th International Congress of Phonetic Sciences, ICPhS2007, Saarbrücken, Allemagne, p. 2221‑2224.

BOURCIEZ Edouard, 1958, *Précis historique de phonétique française*, Paris, Klincksieck.

BOURCIEZ Edouard et BOURCIEZ Jean, 1967, *Phonétique française : étude historique*, Paris, Klincksieck.

BOUTIN Beatrice Akisi, GESS Randall et GUÈYE Gabriel Marie, 2012, « French in Sénegal after three centuries: A phonological study. » dans Gess, R. S., Lyche, C. et Meisenburg, T.

(éds), *Phonological Variation in French: Illustrations from Three Continents*, Amsterdam, J. Benjamins, p. 45‑71.

BOUTIN Beatrice Akisi et LYCHE Chantal, 2014, « Ce que nous apprennent les locuteurs francophones non-lecteurs sur la liaison » dans Soum-Favaro, C., Coquillon, A. et Chevrot, J-P. (éds), *La liaison: approches contemporaines*, Berne, Peter Lang, p. 283‑307.

BOUTON Charles, 1974, *L'acquisition d'une langue étrangère : aspects théoriques et pratiques, conséquences pédagogiques essentielles*, Paris, Klincksieck.

BRANCA-ROSOFF Sonia, FLEURY Serge, LEFEUVRE Florence et PIRES Mat, 2012, « Discours sur la ville. Présentation du Corpus de Français Parlé Parisien des années 2000 (CFPP2000) », URL: http://cfpp2000.univ-paris3.fr/Corpus.html, consulté le 31 juillet 2015.

BROUGERE Gilles ET BEZILLE Hélène, 2007, « De l'usage de la notion d'informel dans le chmp de l'éducation », *Revue française de la pédagogie*, 2007, vol. 158, p. 117−160.

BRUNEAU Charles, 1931, *Manuel de phonétique pratique*, Paris, Éditions Berger-Levrault.

BUBEN Vladimír, 1935, *Influence de l'orthographe sur la prononciation du français moderne*, Bratislava, Filosofická fakulta University Komenského.

BURKI Audrey et FRAUENFELDER Uli, 2012, « Producing and recognizing words with two pronunciation variants: evidence from novel schwa words », *Quarterly journal of experimental psychology*, 2012, vol. 65, nº 4, p. 796‑824.

BURKI Audrey, FRAUENFELDER Uli et ALARIO F.-Xavier, Soumis, « On the resolution of phonological constraints in spoken productions: Acoustic and response time evidence », Soumis.

BURKI Audrey et GASKELL Gareth M., 2012, « Lexical Representation of Schwa Words: Two Mackerels, but Only One Salami », *Journal of Experimental Psychology*, 2012, vol. 38, nº 3, p. 617‑631.

BURKI Audrey, LAGANARO Marina et ALARIO F.-Xavier, 2014, « Phonologically Driven Variability: The Case of Determiners », *Journal of Experimental Psychology: Learning, Memory and Cognition*, 2014, vol. 40, nº 5, p. 1348‑1362.

BURKI Audrey et LAGANARO Marine, 2014, « La liaison: approches psycholinguistiques et neuropsycholinguistiques » dans Soum-Favaro, C., Coquillon, A. et Chevrot, J-P. (éds), *La liaison: approches contemporaines*, Berne, Peter Lang, p. 169‑182.

BUROV Yvaylo, 2012, *Les phénomènes de Sandhi dans l'espace gallo-roman*, Thèse de doctorat, Université Michel de Montaigne - Bordeaux III; Université de Sofia St-Clement d'Ohrid, Bordeaux-Sofia.

BYBEE Joan, 2005, « La liaison : effets de fréquence et constructions », *Langages*, 2005, vol. 39, nº 158, p. 24‑37.

BYBEE Joan, 2001a, « Frequency effects on French liaison » dans Bybee, J. et Hopper P. J. (éds), *Frequency and the emergence of linguistic structure*, Amsterdam, J. Benjamins, p. 337‑359.

BYBEE Joan, 2001b, *Phonology and language use*, Cambridge; New York, Cambridge University Press.

CAPPEAU Paul, 2008, « Perception et reconstruction » dans Bilger, M. (éds), *Données orales les enjeux de la transcription*, Perpignan, Presses Universitaires de Perpignan, p. 235‑247.

CARTON Fernand, 1974, *Introduction à la phonétique du français*, Paris, Bordas.

CARTON Fernand, ROSSI Mario, AUTESSERRE Denis et LEON Pierre, 1983, *Les Accents des francais*, Paris, Hachette.

CATACH Nina, 1968, *L'orthographe française à l'époque de la Renaissance. Auteurs, imprimeurs, ateliers d'imprimerie*, Genève, Droz.

CHAMBERS Angela, 2009, « Les corpus oraux en français langue étrangère : authenticité et pedagogie », *Mélanges CRAPEL*, 2009, nº 31, p. 15‑33.

CHARLIAC Lucile et MOTRON Annie-Claude, 1998, *Phonétique progressive du français : [niveau intérmédiare] avec 600 exercices*, Paris, CLE International.

CHAURAND Jacques linguiste, 1998, *Histoire de la langue francaise*, Paris, Presses Universitaires de France.

CHEVROT Jean-Pierre, 2015, « Orthographe et phonologie : réciprocité des influences et équilibration lors du développement du langage », 13ème Journées du Réseau Français de Phonologie (RFP), Bordeaux, France.

CHEVROT Jean Pierre, CHABANAL Damien et DUGUA Céline, 2007, « Pour un modele de l'acquisition des liaisons base sur l'usage: trois etudes de cas », *Journal of French Language Studies*, 2007, vol. 17, nº 1, p. 103‑128.

CHEVROT Jean-Pierre, DUGUA Céline, HARNOIS-DELPIANO Mylène, SICCARDI Anne et SPINELLI Elsa, 2013, « Liaison acquisition: debates, critical issues, future research », *Language Sciences*, 2013, vol. 39, nº 1, p. 83‑94.

CHEVROT Jean-Pierre et FAYOL Michel, 2001, « Acquisition of French Liaison and Related Child Errors » dans Almgren, M., Barrena A., Ezeizabarrena M-J., Idiazabal, I. et McWhinney, B. (éds), *Research on child language acquisition*, Somerville, Cascadilla Press, p. 760‑774.

CHEVROT Jean-Pierre et FAYOL Michel, 2000, « L'acquisition de la liaison : enjeux théoriques, premiers résultats, perspectives », *Lidil*, 2000, vol. 22, p. 11‑30.

CHEVROT Jean-Pierre et MALDEREZ Isabelle, 1999, « L'effet Buben: de la linguistique diachronique à l'approche cognitive (et retour) », *Langue Francaise*, 1999, nº 124, p. 104‑125.

CHEVROT J.-P, CHABANAL D et DUGUA C, 2007, « Pour un modele de l'acquisition des liaisons base sur l'usage: trois etudes de cas », *Journal of French Language studies*, 2007, vol. 17, n° 1, p. 103‑128.

CHOMSKY Noam et HALLE Morris, 1968, *The sound pattern of English*, New York, Harper and Row.

CHRISTODOULIDES George, 2014, « Praaline: Integrating Tools for Speech Corpus Research », Proceedings of 9th International Conference on Language Resources and Evaluation, LREC2014, Reykjavik, Islande, p. 31‑34.

CHRISTODOULIDES George, AVANZI Mathieu et GOLDMAN Jean-Philippe, 2014, « DisMo: A Morphosyntactic, Disfluency and Multi-Word Unit Annotator. An Evaluation on a Corpus of French Spontaneous and Read Speech », Proceedings of 9th International Conference on Language Resources and Evaluation, LREC2014, Reykjavik, Islande, p. 3902‑3907.

CHRISTODOULIDES George, BARRECA Giulia et AVANZI Mathieu, 2015, « DisMo: un annotateur multi-niveaux pour les corpus oraux », Actes du 22ème colloque Conférence sur leTraitement Automatique des Langues Naturelles, TALN2015, Caen, France.

CLÉDAT Léon, 1917, *Manuel de phonétique et de morphologie historique du français*, Paris, Hachette.

CLÉDAT Léon, 1890, *Précis d'orthographe et de grammaire phonétiques pour l'enseignement du français à l'étranger,* Paris, Masson.

CLEMENTS George N. et KEYSER Samuel J., 1983, *CV phonology: a generative theory of the syllable*, Cambridge, MIT Press.

COQUILLON Annalise, DURAND Jacques, EYCHENNE Julien et LYCHE Chantal, 2010, « French liaison: from global results to local varieties », Colloque international Phonlex, PHONLEX2010, Toulouse, France.

CORDER Stephen P., 1981, *Error analysis and interlanguage*, London, Oxford University Press.

CORDER Stephen P., 1978, « Language-learner language » dans Richards, J.C. (éds), *Understanding second and foreign language learning: issues and approaches*, Rowley, Newbury House Publishers, p. 71‑92.

CORDER Stephen P., 1974, « Error Analysis » dans Allen, J.L.P. et Corder, S.P. (éds), *Techniques in applied linguistics*, Oxford, Oxford University Press, p. 122‑154.

CORDER Stephen P., 1967, « The Significance of Learners' Errors », *International Review of Applied Linguistics*, 1967, vol. 5, p. 161‑169.

CORDER Stephen P., 1981, *Error analysis and interlanguage*, London; New York, Oxford University Press.

COSERIU Eugenio, 1969, *Einführung in die strukturelle Linguistik*, Tübingen, Tübinger Beiträge zur Linguistik.

COSTE Daniel, 2007, « Quelques aspects historiques et actuels de la distinction entre FLM, FLE et FLS » dans Billiez, J et Lambert, P. (éds), *Variations au coeur et aux marges de la sociolin guistique*, Paris, Harmattan, p. 215‑225.

CÔTÉ Marie-Hélène, 2015, « Liaison: résultats PFC », Ecole d'été Phonologie du Français Contemporain, PFC, Université de Vienne, Autriche.

CÔTÉ Marie-Hélène, 2013, « Understanding cohesion in French liaison », *Language Sciences*, 2013, vol. 39, p. 156‑166.

CÔTÉ Marie-Hélène, 2012, « Laurentian French (Québec): extra vowels, missing schwas and surprising liaison consonants. » dans Gess, R., Lyche, C., Meisenburg, T. (éds), *Phonological Variation in French: Illustrations from Three Continents*, Amsterdam, J. Benjamins, p. 235‑274.

CÔTÉ Marie-Hélène, 2010, « Le statut des consonnes de liaison : l'apport de données du français laurentien », Actes du 2ème Congrès Mondial de Linguistique Française, CMLF2010, La Nouvelle-Orléans, États-Unis, p. 1279‑1288.

CÔTÉ Marie-Hélène, 2005, « Le statut lexical des consonnes de liaison », *Langages*, 2005, vol. 39, nº 158, p. 66‑78.

COUSTENOBLE Hélène Nathalie et ARMSTRONG Lilias Eveline, 1934, *Studies in French intonation,* Cambridge, W. Heffer et Sons.

CUQ Jean-Pierre, 1991, *Le français langue seconde : origines d'une notion et implications didactiques*, Paris, Hachette.

CUTLER Anne, MEHLER Jacques, NORRIS Dennis et SEGUI Juan, 1986, « The syllable's differing role in the segmentation of French and English », *Journal of Memory and Language*, 1986, vol. 25, p. 385‑400.

CYLWIK Natalia, WAGNER Agnieszka et DEMENKO Grażyna, 2009, « The EURONOUNCE corpus of non-native Polish for ASR-based Pronunciation Tutoring System », *Archives of Acoustics*, 2009, vol. 35, nº 3, p. 309‑329.

DABENE L, 1990, *Variations et rituels en classe de langue*, Paris, Hatier.

DAMIAN Markus F. et BOWERS Jeffrey S., 2003, « Effects of orthography on speech production in a form-preparation paradigm », *Journal of Memory and Language*, 2003, vol. 49, nº 1, p. 119‑132.

DAMOURETTE Jacques et PICHON Edouard, 1968, *Des mots à la pensée : essai de grammaire de la langue française*, Paris, Editions d'Artrey.

DEBAISIEUX Jeanne-Marie, 2009, « Des documents authentiques oraux aux corpus : un defi pour la didactique du FLE », *Mélanges CRAPEL*, 2009, vol. 31, p. 35‑56.

DEBROCK Mark et FLAMENT-BOISTRANCOURT Danièle, 1996, « Le corpus LANCOM : bilan et perspectives », *Reviews of Applied Linguistics*, 1996, vol. 111-112, p. 1‑36.

DEJEAN DE LA BATIE Bernadette, 1993, *Word boundary ambiguity in spoken French*, Thèse de doctorat, Monash University, Victoria.

DE JONG Adrianus, 1988, *Sociolinguistic aspects of French liaison*, Thèse de doctorat, Free University, Amsterdam.

DE KEYSER Robert M., 2000, « The Robustness of Critical Period Effects in Second Language Acquisition. », *Studies in Second Language Acquisition*, 2000, vol. 22, nº 4, p. 499‑533.

DE LA CLERGERIE Eric, SAGOT Benoit, LIONEL Nicolas et GUENOT Marie-Laure, 2009, « FRMG : évolutions d'un analyseur syntaxique TAG du français », Actes des Journées ATALA, ATALA 2009, Paris, France.

DELAIS-ROUSSARIE Elisabeth, 1995, *Pour une approche parallèle de la structure prosodique étude de l'organisation prosodique et rythmique de la phrase française*, Thèse de doctorat, Université de Toulouse II, Toulouse.

DELAIS-ROUSSARIE Elisabeth et YOO Hiyon, 2010, « The COREIL corpus: a learner corpus designed for studying phrasal phonology and intonation. » dans Dziubalska-Kolaczyk, K. Wrembel, M. et Kul, M. (éds), *Proceedings of New Sounds 2010 - Sixth International Symposium on the Acquisition of Second Language Speech*, Poznan, Adam Mickiewicz University, p. 100‑105.

DELATTRE Pierre, 1966, *Studies in French and comparative phonetics: selected papers in French and English*, The Hague, Mouton de Gruyter.

DELATTRE Pierre, 1956, « La Fréquence des liaisons facultatives en français », *The French Review*, 1956, vol. 30, nº 1, p. 48‑54.

DELATTRE Pierre, 1955, « Les Facteurs de la Liaison Facultative en Français », *The French Review*, 1955, vol. 29, nº 1, p. 42‑49.

DELATTRE Pierre, 1947, « La Liaison en Français, Tendances et Classification », *The French Review*, 1947, vol. 21, nº 2, p. 148‑157.

DELL François, 1973, *Les règles et les sons*, Paris, Hermann.

DELL François, 1970, *Les regles phonologiques tardives et la morphologie derivationnelle du français Topics in French phonology and derivational morphology*, Thèse de doctorat, MIT, Cambridge.

DELL François, HIRST Daniel et VERGNAUD Jean-Roger, 1984, « De la Syllabicité » dans Dell, F., Hirst, D. et J.-R. Vergnaud. (éds), *Forme sonore du langage: structure des représentations en phonologie*, Paris, Hermann, p. 123‑155.

DELL Gary S., 1986, « A Spreading-Activation Theory of Retrieval in Sentence Production », *Psychological Review*, 1986, vol. 93, nº 3, p. 283‑321.

DE MORAS Nadine, 2011, *Acquisition de la liaison et de l'enchaînement en français L2: le rôle de la fréquence*, Thèse de doctorat, Western University, Canada.

DENIS Pascal et SAGOT Benoit, 2009, « Coupling an annotated corpus and a morphosyntactic lexicon for state-of-the-art POS tagging with less human effort », Proceedings of 23rd Pacific Asia Conference on Language, Information and Computation, PACLIC2009, Hong Kong, Chine, p. 110‑119.

DESROCHERS Richard, 1994, « Les liaisons dangereuses: le statut équivoque des erreurs de liaison », *Lingvisticae Investigationes*, 1994, vol. 18, nº 2, p. 243‑284.

DETEY Sylvain, DURAND Jacques, LAKS Bernard et LYCHE Chantal, 2010, *Les variétés du français parlé dans l'espace francophone: ressources pour l'enseignement*, Paris, Ophrys.

DETEY Sylvain, 2010, « Normes pédagogiques et corpus oraux en FLE: le curseur apprenabilité / acceptabilité et la variation phonético-phonologique dans l'espace francophone » dans Bertrand, O. et Schaffner, I. (éds), *Quel français enseigner ? La question de la norme dans l'enseignement/apprentissage*, Paris, Éditions de l'École polytechnique, p. 155‑168.

DETEY Sylvain, 2005, *Interphonologie et représentations orthographiques : du rôle de l'écrit dans l'enseignement-apprentissage du français oral chez des étudiants japonais*, Thèse de doctorat, Université de Toulouse II, Toulouse.

DETEY Sylvain et KAWAGUCHI Yoshiya, 2008, « Interphonologie du Français Contemporain (IPFC): récolte automatisée des données et apprenants japonais », Journées Phonologie du français contemporain: variation, interfaces, cognition (PFC), Paris, France.

DETEY Sylvain et RACINE Isabelle, 2012, « Les apprenants de français face aux normes de prononciation: quelle(s) entrée(s) pour quelle(s) sortie(s) ? », *Revue française de linguistique appliquée*, 2012, vol. 17, nº 1, p. 81‑96.

DI CRISTO Albert et HIRST Daniel, 1993, « Rythme syllabique; rythme mélodique et représentation hiérarchique de la prosodie du français », *Travaux de l'Institut de Phonétique d'Aix*, 1993, vol. 15, p. 9‑24.

DITTMAR Norbert et KLEIN Wolfgang, 1979, *Developing grammars*, Berlin.

DOMERGUE Urbain, 1805, *Manuel des étrangers amateurs de la langue françoise*, Paris, La Librairie économique.

DOSTIE Gaétane, 2004, *Pragmaticalisation et marqueurs discursifs: analyse sémantique et traitement lexicographique*, Bruxelles, De Boeck.

DUBOIS Jean, 1965, *Grammaire structurale du français*, Paris, Larousse.

DUGUA Céline, 2006, *Liaison, segmentation lexicale et schémas syntaxiques entre 2 et 6 ans: un modèle développemental basé sur l'usage*, Thèse de doctorat, Université Stendhal, Grenoble, France.

DULAY Heidi et BURT Marina, 1979, « Natural sequences in child second language acquisition », *Language Learning*, 1979, vol. 24, p. 37‑53.

DULAY Heidi C., BURT Marina K. et KRASHEN Stephen D., 1982, *Language two*, New York, Oxford University Press.

DUPOUX Emmanuel, FUSHIMI Takao, KAKEHI Kazuhiko et MEHLER Jacques, 1999, « Prelexical Locus Of An Illusory Vowel Effect In Japanese », Proceeedings of 6th European Conference on Speech Communication and Technology, Eurospeech99, Budapest, Hongrie.

DUPOUX Emmanuel, KAKEHI Kazuhiko, HIROSE Yuki, PALLIER Christophe et MEHLER Jacques, 1998, « Epenthetic Vowels in Japanese: a Perceptual Illusion? », *Journal of Experimental Psychology: Human Perception and Performance*, 1998, vol. 25, p. 1568‑1578.

DUPOUX Emmanuel, PALLIER Christophe, KAKEHI Kazuhiko et MEHLER Jacques, 2001, « New Evidence for Prelexical Phonological Processing in Word Recognition », *Language and Cognitive Processes*, 2001, vol. 16, p. 491‑505.

DURAND Jacques, LAKS Bernard, CALDERONE Basilio et TCHOBANOV Atanas, 2011, « Que savons-nous de la liaison aujourd'hui ? », *Langue française*, 2011, nº 169, p. 103‑136.

DURAND Jacques, LAKS Bernard et LYCHE Chantal, 2014, « French phonology from a corpus perspective : the PFC programme » dans Durand, J., Gut, U. et Kristoffersen, G. (éds), *The Oxford Handbook of Corpus Phonology*, Oxford, Oxford University Press, p. 486-497.

DURAND Jacques, LAKS Bernard et LYCHE Chantal, 2009, « Le projet PFC: une source de données primaires structurées. » dans *Phonologie, variation et accents du français*, Paris, Hermes.

DURAND Jacques et LYCHE Chantal, 2008, « French liaison in the light of corpus data », *Journal of French Language Studies*, 2008, vol. 18, nº 1, p. 33‑66.

DURAND Jacques et LYCHE Chantal, sous presse, « Approaching variation in PFC: the liaison level. » dans Detey, S., Durand, J., Laks, B. et Lyche, C. (éds), *Varieties of Spoken French*, Oxford, Oxford University Press.

ECKMAN Fred R., 2008, «Typological Markedness and Second Language Phonology », dans Hansen Edwards, J. G. et Zampini, M. L. (éds), *Phonology and Second Language Acquisition*, Amsterdam, J. Benjamins, p. 95-115.

ECKMAN Fred R., 1984, « Universals, typologies and interlanguage » dans Rutherford, W.E. (éds), *Language Universals and Second Language Acquisition*, Amsterdam, J. Benjamins, p. 79‑105.

ECKMAN Fred R., 1981, « On the naturalness of interlanguage phonological rules », *Language Learning*, 1981, vol. 31, nº 1, p. 195‑216.

ECKMAN Fred R., 1977, « Markedness and the Contrastive Analysis Hypothesis », *Language Learning*, 1977, vol. 27, nº 2, p. 315‑330.

ELLIS Rod, 1994, *The study of second language acquisition*, Oxford, Oxford University Press.

ENCREVÉ Pierre, SCHEER Tobias et WAUQUIER Sophie, 2015, « Autosgmental news from h haspiré and liaison without enchaînement », 13ème Rencontres du Réseau Français de Phonologie (RFP2015), 29 juin-1 juillet 2015, Bordeaux, France.

ENCREVÉ Pierre, 1988, *La liaison avec et sans enchaînement : phonologie tridimensionnelle et usages du français*, Paris, Seuil.

ERDENER Vahit Dogu et BURNHAM Denis K., 2005, « The role of audiovisual speech and orthographic information in nonnative speech production », *Language Learning*, 2005, vol. 5, no 2, p. 191‑228.

ESHKOL Iris, TELLIER Isabelle, TAALAB Samer et BILLOT Sylvie, 2010, « Etiqueter un corpus oral par apprentissage automatique à l'aide de connaissances linguistiques », Actes des 10èmes Journées internationales d'Analyse statistique des Données Textuelles, JADT2010, Rome, Italie, p. 1‑12.

EVERT Stefan, 2004, « A simple LNRE model for random character sequences », Actes des 7èmes Journées internationales d'Analyse statistique des Données Textuelles, JADT2004, Louvain-la-Neuve, Belgique, p. 411‑422.

EYCHENNE Julien, 2011, « La liaison en français et la théorie de l'optimalité », *Langue française*, 2011, no 1, p. 79‑101.

EYCHENNE Julien, LYCHE Chantal, DURAND Jacques et COQUILLON Annelise, 2014, « Quelles données pour la liaison en français: la question des corpus » dans Soum-Favaro, C., Coquillon, A. et Chevrot, J-P. (éds), *La liaison : approches contemporaines*, Berne, Peter Lang, p. 33‑60.

EYCHENNE Julien et PATERNOSTRO Roberto, sous presse, « Analyzing transcribed speech with Dolmen » dans Detey, S., Durand, J. et Lyche, C. (éds), *Varieties of Spoken French: a source book*, Oxford, Oxford University Press.

EYCHENNE Julien et WALKER Douglas C., 2010, « La prononciation du français en Amérique du Nord : éléments de synthèse » dans Detey, S., Durand, J., Laks, B. et Lyche, C. (éds), *Les variétés du français parlé dans l'espace francophone: ressources pour l'enseignement*, Paris, Ophrys, p. 131‑144.

FALBO Caterina, JANOT Pascale, MURANO Michela et PATERNOSTRO Roberto, sous presse, « Gran[d] émoi à l'Unio[n] européenne": studenti italofoni di FLE alle prese con la liaison », *L'apprentissage de la liaison en français par des locuteurs non natifs: éclairage des corpus oraux, Bulletin VALS-ASLA*, sous presse, vol. 201.

FLEDGE James E., 1987, « The production of "new" and "similar" phones in a foreign language : evidence for the effect of equivalence classification », *Journal of Phonetics*, 1987, vol. 15, p. 47‑65.

FOUCHE Pierre, 1959, *Traité de prononciation française*, Paris, Klincksieck.

FOUCHE Pierre, 1939, « Les diverses sortes du français au point de vue phonétique », *Le français moderne*, 1939, vol. 4, p. 199‑216.

FOUGERON Cécile, GOLDMAN Jean-Philippe, DART Alicia, GUELAT Laurence et JEAGER Clémentine, 2001, « Influence de facteurs stylistiques, syntaxiques et lexicaux sur la réalisation de la liaison en français », Actes de la 8ème Conférence Traitement Automatique des Langues Naturelles, TALN2001, Tours, France.

FOUGERON Cécile, GOLDMAN Jean-Philippe et FRAUENFELDER Uli, 2001, « Liaison and schwa deletion in French: an effect of lexical frequency and competition? », Proceedings of 7th European Conference on Speech Communication and Technology 2nd INTERSPEECH, INTERSPEECH, Aslborg, Denmark.

FREI Henri, 1929, *La grammaire des fautes*, Paris, Geuthner.

FROST Ram, 1998, « Toward a strong phonological theory of visual word recognition: True issues and false trails. », *Psychological Bulletin Psychological Bulletin*, 1998, vol. 123, nº 1, p. 71‑99.

GADET Françoise, 2007, *La variation sociale en français*, Paris, Ophrys.

GADET Françoise, 1989, *Le français ordinaire*, Paris, Colin.

GADET Françoise et TYNE Henry, 2007, « Le style en sociolinguistique: ce que nous apprend l'acquisition », *Pratiques*, 2007, vol. 135-136, p. 91‑99.

GALAZZI Enrica, 2015a, « L'humour des sons: ap-prendre la parole en découvrant le plaisir des différences » dans Paissa, P., Rigat, F. et Vittoz, M.-B. (éds), *Dans l'amour des mots. Chorale(s) pour Mariagrazia*, Alessandria, Edizioni dell'Orso, p. 29-39.

GALAZZI Enrica, 2015b, « La prononciation du français standard à l'épreuve du troisième millénaire », *Repères DoRiF,* 2015, vol. 8, URL: http://www.dorif.it/ezine/ezine_articles.php?id=235, consulté le 9 septembre 2015.

GALAZZI Enrica, 2014, « "Poisson sans boisson est poison": de l'articulation à la diction. Manuels de prononciation et apprentissage de la langue parlée (1880-1910) », *Documents pour l'histoire du français langue étrangère ou seconde*, 2014, vol. 52, p. 63‑81.

GALAZZI Enrica, 2008a, « Accent natif où tu es-tu? », *Rassegna Italiana di Linguistica Applicata*, 2008, vol. 3, p. 115‑128.

GALAZZI Enrica, 2008b, « 1950: où est passée l'oralité? Français Fondamental, corpus oraux, contenus d'enseignement 50 ans de travaux et d'enjeux », *Le français dans le monde. Recherches et Applications*, 2008, vol. 43, p. 12‑26.

GALAZZI Enrica, 2002a, *Le son à l'école: phonétique et enseignement des langues : fin XIXe siècle-début XXe siècle*, Brescia, La scuola.

GALAZZI Enrica, 2002b, « L'enseignement des langues: le français en Italie », *Les Cahiers de l'ASDIFLE*, 2002, vol. 13, p. 37‑62.

GALAZZI Enrica, FALBO Caterina, JANOT Pascale, MURANO Michela et PATERNOSTRO Roberto, 2013, « Autour d'un corpus d'apprenants italophones de FLE: présentation du projet Interphonologie du Français Contemporain – italien », *Repères DoRiF*, 2013, vol. 3.

GALAZZI Enrica et MOLINARI Chiara, 2007, *Les français en émergence*, Berne, Peter Lang.

GAONAC'H Daniel, 1991, *Théories d'apprentissage et acquisition d'une langue étrangère*, Paris, Hatier.

GASKELL Gareth M., SPINELLI Elsa et MEUNIER Fanny, 2002, « Perception of resyllabification in French. », *Memory and Cognition*, 2002, vol. 30, p. 798‑810.

GASS Susan M. et MACKEY Alison, 2002, « Frequency Effects and Second Language Acquisition: A Complex Picture? », *Studies in Second Language Acquisition*, 2002, vol. 24, nº 2, p. 249‑260.

GASS Susan M. et SELINKER Larry, 1992, *Language transfer in language learning*, Amsterdam, J. Benjamins.

GERMAIN Claude, 1993, *Évolution de l'enseignement des langues: 5000 ans d'histoire*, Paris, CLE international.

GILQUIN Gaëtanelle, DE COCK Sylvie et GRANGER Sylviane, 2010, « Louvain International Database of Spoken English Interlanguage », URL: https://www.uclouvain.be/en-cecl-lindsei.html, consulté le 1 août 2015.

GIVÓN Talmy, 1979, *On understanding grammar*, New York, Academic Press.

GOÏ Cécile et HUVER Emmanuelle, 2012, « FLE, FLS, FLM: continuum ou interrelations ? », *Le Français aujourd'hui*, 2012, nº 1, p. 25‑35.

GOLDBERG Adele E., 1995, *Constructions: a construction grammar approach to argument structure*, Chicago, University of Chicago Press.

GOLDSMITH John A., 1979, *Autosegmental phonology*, Thèse de doctorat, Garland Pub., New York.

GOMBERT Jean Émile, 1990, *Le développement métalinguistique*, Paris, Presses universitaires de France.

GOUGENHEIM Georges, 1938, *Système grammatical de la langue française*, Paris, Éditions d'Artrey.

GOUGENHEIM Georges, 1935, *Éléments de phonologie française; étude descriptive des sons du français au point de vue fonctionnel*, Paris, Les Belles Lettres.

GRAINGER Jonathan et FERRAND Ludovic, 1996, « Masked Orthographic and Phonological Priming in Visual Word Recognition and Naming: Cross-Task Comparisons », *Journal of Memory and Language*, 1996, vol. 35, nº 5, p. 623‑647.

GRAMMONT Maurice, 1933, *Traité de phonétique*, Paris, Delagrave.

GRAMMONT Maurice, 1914, *Traité pratique de prononciation française*, Paris, Delagrave.

GREVISSE Maurice, 1936, *Le bon usage: grammaire française*, Paris, Duculot.

GROSJEAN François et YERSIN-BESSON Carole, 1996, « L'effet de l'enchaînement sur la reconnaissance des mots dans la parole continue », *L'année psychologique*, 1996, vol. 96, nº 1, p. 9‑30.

GUIRAUD Pierre, 1965, *Le francais populaire*, Paris, Presses universitaires de France.

GUT Ulrike, 2009, *Non-native speech a corpus-based analysis of phonological and phonetic properties of L2 English and German*, Frankfurt am Main, Peter Lang.

HALLÉ Pierre A., CHÉREAU Céline et SEGUI Juan, 2000, « Where Is the /b/ in "absurde" [apsyrd]? It Is in French Listeners' Minds », *Journal of Memory and Language*, 2000, vol. 43, nº 4, p. 618‑639.

HALLÉ Pierre A., SEGUI Juan, FRAUENFELDER Uli et MEUNIER Christine, 1998, « Processing of illegal consonant clusters: A case of perceptual assimilation? », *Journal of Experimental Psychology Human Perception & Performance*, 1998, vol. 24, nº 2, p. 592‑608.

HANSEN EDWARDS Jette G., 2008, « Social factors and variation in production in L2 phonology» dans Hansen Edwards, J. G. et Zampini, M. L. (éds), *Phonology and Second Language Acquisition*, Amsterdam, J. Benjamins, p. 251‑279.

HARNOIS-DELPIANO Mylène, CAVALLA Cristelle et CHEVROT Jean-Pierre, 2012, « L'acquisition de la liaison en L2: étude longitudinale chez des apprenants coréens de FLE et comparaison avec enfants francophones natifs », Actes du 3ème Congrés Mondial de Linguistique Française, CMLF2012, Lyon, France, p. 1575‑1589.

HERNANDEZ Nicolas et BOUDIN Florian, 2013, « Construction d'un large corpus écrit libre annoté morpho-syntaxiquement en français », Actes de 20ème la Conférence Traitement Automatique des Langues Naturelles, TALN2013, Les Sables d'Olonne, France, p. 160‑173.

HOWARD Martin, 2013, « La liaison en français langue seconde: une étude longitudinale préliminaire », *Language, Interaction and Acquisition*, 2013, vol. 4, nº 2, p. 190‑231.

HOWARD Martin, 2005, « L'acquisition de la liaison en français langue seconde », *Cognition, représentation, langage*, 2005, nº HS-1.

IOUP Georgette, 2008, « Exploring the role of age in the acquisition of a second language phonology » dans Hansen Edwards, J. G. et Zampini, M. L. (éds), *Phonology and Second Language Acquisition*, Amsterdam, J. Benjamins, p. 41-62.

IOUP Georgette et WEINBERGER Steven, 1987, *Interlanguage phonology: the acquisition of a second language sound system*, Cambridge, Newbury House Publishers.

JACOBS Bob, 1988, « Neurobiological Differentiation of Primary and Secondary Language Acquisition. », *Studies in Second Language Acquisition*, 1988, vol. 10, nº 3, p. 303‑337.

JOHNS Tim, 1988, « Implications et applications des logiciels de concordance dans la salle de classe », *Les Langues modernes*, 1988, vol. 5.

KAHN Daniel, 1980, *Syllable-based generalizations in English phonology*, Thèse de doctorat, Garland Pub., New York.

KEMPEN Gerard et HOENKAMP Edward, 1987, « An Incremental Procedural Grammar for Sentence Formulation », *Cognitive Science*, 1987, vol. 11, nº 2, p. 201‑258.

KILANI-SCHOCH, Marianne, 1983, « Troncation ou insertion dans les liaisons francaises: quelques donnees aphasiques comme indices externes », *Folia Linguistica*, vol. 17, p. 445‑461.

KLAUSENBURGER Jurgen, 2001, *Features and Interfaces in Romance Essays in honor of Heles Contreras*, Amsterdam, J. Benjamins.

KLAUSENBURGER Jürgen, 1984, *French liaison and linguistic theory*, Stuttgart, F. Steiner Verlag Wiesbaden.

KLAUSENBURGER Jurgen, 1974, « Rule inversion, opacity, conspiracies: French liaison and elision », *Lingua*, 1974, vol. 34, nº 2–3, p. 167‑179.

KLEIN Wolfgang, 1998, « The contribution of second language acquisition research », *Language Learning*, 1998, vol. 48, nº 4, p. 527‑550.

KLEIN Wolfgang, 1989, *L'acquisition de langue étrangère*, Paris, Colin.

KLEIN Wolfgang et PERDUE Clive, 1997, « The Basic Variety (or: Couldn't natural languages be much simpler?) », *Second Language Research*, 1997, vol. 13, nº 4, p. 301‑347.

KLEIN Wolfgang et PERDUE Clive, 1992, *Utterance structure developing grammars again*, Amsterdam, J. Benjamins.

KOUIDER Sid et DUPOUX Emmanuel, 2001, « A functional disconnection between spoken and visual word recognition: evidence from unconscious priming », *Cognition*, 2001, vol. 82, nº 1, p. 35‑49.

KRASHEN Stephen D., 1981, *Second language acquisition and second language learning*, Oxford; New York, Pergamon Press.

KRASHEN Stephen D., SFERLAZZA Victoria, FELDMAN Lorna et FATHMAN Ann K., 1976, « Adult performance on the slope test: more evidence for a natural sequence in adult second language acquisition », *Language Learning*, 1976, vol. 26, nº 1, p. 145‑151.

LACHERET-DUJOUR Anne, 1999, *La prosodie du français*, Paris, CNRS Éditions.

LADO, Robert, 1957, *Linguistics across cultures*, Ann Arbor, University of Michigan Press.

LAGANARO Marina, BAGOU Odile, MICHEL Violaine, DAYER Géraldine et SCHNEIDER Laurence, 2010, « On the production of sandhi phenomena in French aphasic speakers with impaired phonological/phonetic encoding », *Procedia social and behavioral sciences*, 2010, vol. 6, p. 76‑77.

LAKS Bernard, 2014, « Diachronie de la liaison en français contemporain : le cas de la parole publique (1999-2011) » dans Durand. J., Gjeerts, K. et Laks, B. (éds), *La phonologie du français : normes périphéries modélisation : Mélanges pour Chantal Lyche*, Paris, Presses Universitaires de Paris Ouest, p. 333‑375.

LAKS Bernard, 2009, « Les hommes politiques français et la liaison (1908-1999) » dans Baronian, L. et Martineau, F. (éds), *Mélanges offerts à Yves Charles Morin*, Québec, Presses de l'Université de Laval, p. 237‑269.

LAKS Bernard, 2008, « Pour une phonologie de corpus », *Journal of French Language Studies*, 2008, vol. 18, n° 01, p. 3‑32.

LAKS Bernard, 2005a, « Phonologie et construction syntaxique : la liaison, un test de cohésion et de figement syntaxique », *Revue des linguistes de l'université Paris X Nanterre*, 2005, n° 53, p. 155‑171.

LAKS Bernard, 2005b, « La liaison et l'illusion », *Langages*, 2005, vol. 39, n° 158, p. 101‑125.

LAKS Bernard, 2005c, « Dynamiques de la liaison en français: synchronie et diachronie », Munich, 2005.

LAKS Bernard, 2002, « Description de l'oral et variation : la phonologie et la norme », *L'Information Grammaticale*, 2002, vol. 94, n° 1, p. 5‑10.

LAKS Bernard, 1983, « Langage et pratiques sociales. Étude sociolinguistique d'un groupe d'adolescents », *Actes de la recherche en sciences sociales*, 1983, vol. 46, n° 1, p. 73‑97.

LAKS Bernard, 1980, *Différenciation linquistique et différenciation sociale: quelques problèmes de sociolinguistique française*, Thèse de doctorat, Université Paris VIII, Vincennes.

LAKS Bernard, 1977, « Contribution empirique a l'analyse socio-differentielle de la chute des /r/ dans les groupes consonantiques finals », *Langue Francaise*, 1977, vol. 34, p. 109‑125.

LAKS Bernard, CALDERONE Basilio et CELATA Chiara, 2014, « French liaison and the lexical repository » dans Celata, C., Calamai, S. (éds), *Advances in Sociophonetics*, Amsterdam, J. Benjamins, p. 29‑54.

LANDRON Simon, PAILLEREAU, Nikola, NAWAFLEH Ahmad, EXARE Christelle, ANDO Hirofumi et GAO Jiyain, 2010, « Le corpus PhoDiFLE : un corpus commun de français langue étrangère pour une étude phonétique des productions de locuteurs de langues maternelles plurielles », *Cahiers de praxématique*, 2010, vol. 54-55, p. 73‑86.

LANGACKER Ronald W., 1987, *Foundations of cognitive grammar*, Stanford, Stanford University Press.

LANGLARD Henri, 1928, *La liaison dans le français*, Paris, Champion.

LARSEN-FREEMAN Diane et LONG Michael H., 1991, *An introduction to second language acquisition research*, London, Longman.

LAURET Bertrand, 2007, *Enseigner la prononciation du français: questions et outils*, Paris, Hachette.

LENNEBERG Eric H., CHOMSKY Noam et MARX Otto, 1967, *Biological foundations of language*, New York, Wiley.

LEON Monique, 2003, *Exercices systématiques de prononciation française*, Paris, Hachette.

LEON Monique et LEON Pierre R., 1997, *La prononciation du français*, Paris, Nathan.

LEON Pierre R., 2007, *Phonétisme et prononciations du français : avec travaux pratiques d'application et corrigés*, Paris, Colin.

LEON Pierre R., 1968, « Aspects phonostylistiques des niveaux de langue », *Le français dans le monde*, 1968, nº 57, p. 68‑72.

LE PRIEULT Henri, 2006, *Grammaticalité : traditions et modernités*, Toulouse, Presses universitaires du Mirail.

LEON Pierre R., 1966, *Prononociation du français standard. Aide mémoire d'orthoépie. À l'usage des étudiants étrangers* (édition revue et argumentée de : Aide mémoire d'Orthoépie, CLA Besançon 1961), Paris, Didier.

LEVELT W. J. M, 1989, *Speaking from intention to articulation*, Cambridge, MIT Press.

LIBERMAN Mark, 1979, *The intonational system of English*, Thèse de doctorat, Garland Pub., New York.

LIBERMAN Mark et PRINCE Alan, 1977, « On Stress and Linguistic Rhythm », *Linguistic Inquiry*, 1977, vol. 8, nº 2, p. 249‑336.

LOPORCARO Michele, 1997, *L'origine del raddoppiamento fonosintattico: saggio di fonologia diacronica romanza*, Basel, Francke.

LOTT David, 1983, « Analysing and Counteracting Interference Error », *English Language Teaching Journal*, 1983, vol. 37, nº 3, p. 256‑261.

LUA Kim Teng, 1994, « Frequency-rank curves and entropy for Chinese characters and words », *Computer Processing of Chinese and Oriental Languages*, 1994, vol. 8, nº 1, p. 37‑52.

LUCCI Vincent, 1983, *Etude phonétique du français contemporain à travers la variation situationnelle: débit, rythme, accent, intonation, [schwa] muet, liaisons, phonèmes*, Grenoble, Université des langues et lettres de Grenoble.

LÜDI Georges et PY Bernard, 1986, *Être bilingue*, Berne, Peter Lang.

LYCHE Chantal, 2010, « Le français de référence : éléments de synthèse » dans Detey, S., Durand, J., Laks, B. et Lyche, C. (éds), *Les variétés du français parlé dans l'espace francophone : ressources pour l'enseignement*, Paris, Ophrys, p. 143‑165.

MAJOR, Roy C., 2008, « Transfer in second language phonology », dans Hansen Edwards, J. G. et Zampini, M. L. (éds), *Phonology and Second Language Acquisition*, Amsterdam, J. Benjamins, p. 63-94.

MALECOT André, 1977, *Introduction à la phonétique française*, The Hague, Mouton de Gruyter.

MALECOT André, 1975, « French Liaison as a Function of Grammatical, Phonetic and Paralinguistic Variables », *Phonetica*, 1975, vol. 32, nº 3, p. 161‑179.

MALLET Géraldine-Mary, 2008, *La liaison en français : descriptions et analyses dans le corpus PFC*, Thèse de doctorat, Université Paris Ouest Nanterre La Défense, Nanterre.

MALMBERG Bertil, 1974, *Manuel de phonétique générale*, Paris, Picard.

MALMBERG Bertil, 1969, *Phonétique française*, Malmö, Hermonds.

MANDELBROT Benoit B., 1957, *Linguistique statistique macroscopique. Théorie mathématique de la loi de Zipf (basée sur les méthodes de la théorie de l'information et de la thermodynamique)*, Paris, Institut Henri Poincaré, Seminaire de calcul des probabilites.

MARSLEN-WILSON William, 1987, « Functional parallelism in spoken word-recognition », *Cognition*, 1987, vol. 25, nº 1-2, p. 71‑102.

MARSLEN-WILSON William et WELSH Alan, 1978, « Processing Interactions and Lexical Access during Word Recognition in Continuous Speech », *Cognitive Psychology*, 1978, vol. 10, p. 29‑73.

MARTINET André, 1971, *La prononciation du français contemporain*, Genève, Droz.

MARTINET André, 1964, « Pour un dictionnaire de la prononciation française » dans Abercrombie, D. et al. (éds), *In Honour of Daniel Jones*, Londres, Longmans, p. 129‑131.

MARTINET André et WALTER Henriette, 1973, *Dictionnaire de la prononciation française dans son usage réel*, Paris, France-Expansion.

MARTINON Philippe, 1913, *Comment on prononce le français: traité complet de prononciation pratique avec les noms propres et les mots étrangers*, Paris, Larousse.

MASTROMONACO Silvana M., 1999, *Liaison in French as a second language*, Thèse de doctorat, Université de Toronto, Canada.

MCQUEEN James, CUTLER Anne et NORRIS Dennis, 2006, « Phonological abstraction in the mental lexicon », *Cognitive Science*, 2006, vol. 30, p. 1113‑1126.

MEHLER Jacques, DOMMERGUES Jean-Yves, FRAUENFELDER Uli et SEGUI Juan, 1981, « The syllable's role in speech segmentation », *Journal of Verbal Learning and Verbal Behavior*, 1981, nº 20, p. 298‑305.

MEIGRET Louis, 1542, *Traité touchant le commun usage de l'escriture françoise*, Genève, Slatkine Reprints.

MEISEL Jürgen M., CLAHSEN Harald et PIENEMANN Manfred, 1981, « On determining developmental stages in natural second language acquisition », *Studies in Second Language Acquisition*, 1981, vol. 3, nº 2, p. 109-135.

MILLER George A. et NEWMAN Edwin B., 1958, « Tests of a statistical explanation of the rank-frequency relation for words in written English », *The American journal of psychology*, 1958, vol. 71, nº 1, p. 209‑218.

MILNER Jean-Claude, 1967, « French truncation rule », Quartely Progress Report of the Research Laboratory of Electronics, Cambridge, États-Unis, MIT Press.

MILROY Lesley, 1980, *Language and social networks*, Oxford, Blackwell.

MORAIS Jose, CASTRO Sao Luis, SCLIAR-CABRAL L., KOLINSKY Régine et CONTENT Alain, 1987, « The effects of literacy on the recognition of dichotic words », *Quaterly Journal of Experimental Psychology*, 1987, vol. 39, nº 3, p. 451‑465.

MORIN Yves-Charles, 2005, « La liaison releve-t-elle d'une tendance a eviter les hiatus ? Reflexions sur son evolution historique », *Langages*, 2005, vol. 39, nº 158, p. 8‑23.

MORIN Yves-Charles, 2003, « Remarks on prénominal liaison consonants in French » dans Ploch, S. (éds), *Living on the edge. 28 papers in honour of Jonathan Kaye*, Berlin, Mouton de Gruyter, p. 385‑400.

MORIN Yves-Charles, 2000, « Le français de référence et les normes de prononciation », *Cahiers de l'Institut de linguistique de Louvain*, 2000, vol. 26, nº 1, p. 91‑135.

MORIN Yves-Charles, 1992, « Le mot, les mots, les bons mots » dans Mel′čuk, I., Clas, A., Cormier, M., Collignon, D. (éds), *Le Mot, les mots, les bons mots: hommage à Igor A. Mel'čuk par ses amis, collègues et élèves à l'occasion de son soixantième anniversaire*, Montréal, Presses de l'Université de Montréal, p. 233‑250.

MORIN Yves-Charles, 1986, « On the morphologization of word-final consonant deletion in French » dans Andersen, H. (éds), *Sandhi Phenomena in the Languages of Europe*, Berlin, Mouton de Gruyter, p. 167‑210.

MORIN Yves-Charles et KAYE Jonathan D., 1982, « The syntactic bases for French liaison », *Journal of Linguistics*, 1982, vol. 18, nº 2, p. 291‑330.

MOUGEON Raymond, NADASDI Terry et REHNER Katherine, 2002, « État de la recherche sur l'appropriation de la variation par les apprenants avancés du FL2 ou FLE », *Acquisition et interaction en langue étrangère*, 2002, nº 17, p. 7‑50.

MYLES Florence et MITCHELL Rosamund, 2007, *French Learner Language Oral Corpora (FLLOC)*, http://www.flloc.soton.ac.uk/, 2007, consulté le 10 juin 2015.

NERI Ambra, CUCCHIARINI Catia et STRIK Helmer, 2006, « Selecting Segmental Errors in Non-Native Dutch for Optimal Pronunciation Training », *International Review of Applied Linguistics in Language Teaching (IRAL)*, 2006, vol. 44, nº 4, p. 357‑404.

NESPOR Marina, 1993, *Fonologia*, Bologna, Il Mulino.

NEW Boris, 2006, « Lexique 3 : Une nouvelle base de données lexicales », Actes de la 13ème Conférence Traitement Automatique des Langues Naturelles, TALN2006, Louvain-la-Neuve, Belgique.

NGUYEN Noël, WAUQUIER-GRAVELINES Sophie, LANCIA Leonardo et TULLER Betty, 2007, « Detection of liaison consonants in speech processing in French: Experimental data and theoretical implications » dans Prieto P., Mascarò J. et Solé M. J. (éds), *Segmental and Prosodic Issues in Romance Phonology*, Amsterdam, J. Benjamins, p. 3‑23.

NIEDERMANN Max, 1953, *Précis de phonétique historique du latin*, Paris, Klincksieck.

NOYAU Colette, 1980, « Étudier l'acquisition d'une langue non-maternelle en milieu naturel », *Langages*, 1980, nº 57, p. 73‑86.

NYROP Kristoffer, 1934, *Manuel phonétique du français parlé*, Copenhagen, Gyldendalske Boghande.

ODLIN Terence, 1989, *Language transfer: cross-linguistic influence in language learning*, Cambridge, Cambridge University Press.

OLLER John W. et ZIAHOSSEINY Seid M, 1970, « The contrastive analysis hypothesis and spelling errors », *Language Learning*, 1970, vol. 20, p. 183‑189.

PAGLIANO Claudine, 2003, *L'épenthèse consonantique en français*, Thèse de doctorat, Université de Nice-Sophia Antipolis, Nice.

PARETO Vilfredo, 1896, *Cours d'économie politique professé à l'Universiţté de Lausanne*, Lausanne, Pichon.

PASSY Paul, 1887, *Les Sons du français,* Paris, Firmin Didot.

PATERNOSTRO Roberto, 2015, « La réalisation de la liaison et la gestion de la variation chez les apprenants italophones », Ecole d'été Phonologie du Français Contemporain, PFC, Université de Vienne, Autriche.

PATERNOSTRO Roberto, 2014, *L'intonation des jeunes en région parisienne: aspects phonétiques et sociolinguistiques, implications didactiques*, Thèse de doctorat, Università degli studi di Brescia; Université Paris Ouest Nanterre La Défense, Nanterre.

PELLET CHENEVAL Pauline, MICHEL Violaine, PYTHON Grégoire et LAGANARO Marina, 2013, « Contextual phonological errors and sandhi phenomenon as un index of ahead phonological planning in aphasia », *Procedia social and behavioral sciences*, 2013, vol. 94, nº 4, p. 51‑52.

PERDUE Clive, 1993, *Adult language acquisition: cross-linguistic perspectives*, Cambridge, Cambridge University Press.

PERLMUTTER David, 1998, « Interfaces: explanation of allomorphy and the architecture of grammars » dans Lapointe, S., Brentani, D. et Farrell, P. (éds), *Morphology and its relation to phonology and syntax*, Stanford, CSLI Publications, p. 307‑338.

PICHON Edouard, 1938, « Genre et questions connexes. (Sur les pas de Mlle Durand) », *Le français moderne*, 1938, vol. 6, p. 107‑126.

PIENEMANN Manfred, 2005, *Cross-linguistic aspects of processability theory*, Amsterdam, J. Benjamins.

PIENEMANN Manfred, 1998, *Language processing and second language development processability theory*, Amsterdam, J. Benjamins.

PIENEMANN Manfred, 1985, « Learnability and syllabus construction » dans Hyltenstam, K. et al. (éds), *Modelling and assessing second language acquisition*, Clevedon, Multilingual Matters, p. 171‑201.

PIENEMANN Manfred, 1984, « Psychological Constraints on the Teachability of Languages », *Studies in Second Language Acquisition*, 1984, vol. 6, n° 2, p. 186‑214.

PIERREHUMBERT Janet B., 2001, « Exemplar dynamics: Word frequency, lenition and contrast » dans Bybee J. et Hopper P. (éds), *Frequency and the emergence of linguistic structure*, Amsterdam, J. Benjamins, p. 137‑158.

PILLOT-LOISEAU Claire, AMELOT Angélique et FREDET Florentina, 2010, « Contributions of experimental phonetics to the didactics of the pronunciation of the French as a Foreign language: stage 1: reflection around the establishment of a speaking materials » dans Dziubalska-Kolaczyk, K. Wrembel, M. et Kul, M. (éds), *Proceedings of New Sounds 2010 - Sixth International Symposium on the Acquisition of Second Language Speech*, Poznan, Adam Mickiewicz University, p. 343‑348.

PORQUIER Rémy, 1974, « Progression didactique et progression d'apprentissage : quels critères », *Études de linguistique appliquée*, 1974, vol. 16, p. 105‑111.

PRINCE Alan et SMOLENSKY Paul, 1993, *Optimality theory: constraint interaction in generative grammar*, New Brunswick, Rutgers Center for Cognitive Science, Rutgers, the State University of New Jersey.

PULGRAM Ernst, 1970, *Syllable, word, nexus, cursus*, The Hague, Mouton de Gruyter.

PUREN Christian, 1988, *Histoire des méthodologies de l'enseignement des langues*, Paris, CLE international.

PUSTKA Elissa, 2011, *Einführung in die Phonetik und Phonologie des Französischen*, Berlin, Erich Schmidt Verlag.

PUSTKA Elissa, sous presse, « Die Liaison im Fremdspracherwerb: Eine Pilotstudie zu Münchner Lehramtsstudenten », *L'apprentissage de la liaison en français par des locuteurs non natifs: éclairage des corpus oraux, Bulletin VALS-ASLA*, sous presse, vol. 201.

PY Bernard, 1980, « Quelques réflexions sur la notion d'interlangue », *Travaux neuchâtelois de linguistique*, 1980, vol. 1, p. 31-54.

RACINE Isabelle, 2008, *Les effets de l'effacement du Schwa sur la production et la perception de la parole en français*, Thèse de doctorat, Université de Genève, Genève.

RACINE Isabelle, sous presse, « La liaison chez les apprenants hispanophones avancés de FLE », *L'apprentissage de la liaison en français par des locuteurs non natifs: éclairage des corpus oraux, Bulletin VALS-ASLA*, sous presse, vol. 201.

RACINE Isabelle, BÜRKI Audrey et SPINELLI Elsa, 2014, « The implication of spelling and frequency in the recognition of phonological variants: evidence from pre-readers and readers », *Language, Cognition and Neuroscience*, 2014, vol. 29, nº 7, p. 893‑898.

RACINE Isabelle et DETEY Sylvain, sous presse, « Corpus oraux, liaison et locuteurs non-natifs: de la recherche en phonologie à l'enseignement du français langue étrangère », *L'apprentissage de la liaison en français par des locuteurs non natifs: éclairage des corpus oraux, Bulletin VALS-ASLA*, sous presse, vol. 201.

RACINE Isabelle, DETEY Sylvain, BUHLER Nathalie, SCHWAB Sandra, ZAY Françoise et KAWAGUCHI Yuji, 2010, « The production of French nasal vowels by advanced Japanese and Spanish learners of French : a corpus-based evaluation study » dans Dziubalska-Kolaczyk, K. Wrembel, M. et Kul, M. (éds), *Proceedings of New Sounds 2010 - Sixth International Symposium on the Acquisition of Second Language Speech*, Poznan, Adam Mickiewicz University, p. 367‑372.

RACINE Isabelle, DETEY Sylvain et KAWAGUCHI Yuji, 2012, « Des atouts d'un corpus multitâches pour l'étude de la phonologie en L2: l'exemple du projet « Interphonologie du français contemporain (IPFC) » dans Kamber, A., Skupiens, C. (éds), *Recherches récentes en FLE*, Berne, Peter Lang, p. 1‑19.

RACINE Isabelle et GROSJEAN Francs, 2005, « Le coût de l'effacement du schwa lors de la reconnaissance des mots en francais », *Canadian journal of experimental psychology /Revue canadienne de psychologie expérimentale*, 2005, vol. 59, nº 4, p. 240.

RACINE Isabelle, ZAY Françoise, DETEY Sylvain et KAWAGUCHI Yuji, 2011, « De la transcription de corpus à l'analyse interphonologique : enjeux méthodologiques en FLE », *Travaux Linguistiques du CerLiCO*, 2011, vol. 24, p. 13‑30.

RAPP Brenda, BENZING Lisa et CARAMAZZA Alfonso, 1997, « The Autonomy of Lexical Orthography », *Cognitive Neuropsychology Cognitive Neuropsychology*, 1997, vol. 14, nº 1, p. 71‑104.

RENOU Louis, 1975, *Grammaire sanscrite*, Paris, Adrien-Maisonneuve.

REZEAU Joseph, 2007, « L'apport du concordancier à l'analyse et à la remédiation des erreurs des apprenants dans les forums de discussion en ligne », *Apprentissage des Langues et Systèmes d'information et de Communication*, 2007, vol. 10, nº 2, p. 27‑43.

REZEAU Joseph, 1997, « Applications des concordanciers à l'enseignement de la grammaire anglaise en DEUG », *Groupe d'étude et de recherche en anglais de spécialité*, 1997, vol. 15-18, p. 117‑132.

RICHARDS Jack, 1971, « A Non-Contrastive Approach to Error Analysis », *English Language Teaching Journal*, 1971, vol. 25, p. 204‑219.

RIVENC Paul, 2003, *Apprentissage d'une langue étrangère seconde 3*, Bruxelles, De Boeck.

ROELOFS Ardi, 2006, « The influence of spelling on phonological encoding in word reading, object naming, and word generation », *Psychonomic bulletin and review*, 2006, vol. 13, nº 1, p. 33‑37.

RUTHERFORD William E, 1984, *Language Universals and Second Language Acquisition*, Amsterdam, J. Benjamins.

RUVOLETTO Samantha et WAUQUIER Sophie, 2015, « Liaison en français : de l'oral à l'écrit, c'est d' la phonologie », 13ème Rencontres du Réseau Français de Phonologie (RFP2015), 29 juin-1 juillet 2015, Bordeaux, France.

RUVOLETTO Samantha, 2014, « Liaison, élision et enchaînement : le rôle de la phonologie et du lexique chez les enfants au début de l'école primaire », Actes du 4ème Congrès Mondial de Linguistique Française, CMLF2014, Berlin, Allemagne, p. 1579‑1590.

SAFFRAN Jenny, 2003, « Statistical Language Learning : Mechanisms and Constraints », *Current Directions in Psychological Science*, 2003, vol. 12, nº 4, p. 110‑114.

SAMPSON Rodney, 2001, « Liaison, nasal vowels and productivity », *Journal of French Language Studies*, 2001, vol. 11, nº 2, p. 241-258.

SCHANE Sanford A., 1974, « There is no French Truncation Rule » dans Campbell, R. J., Goldin G. M., Wang, C. M. (éds), *Linguistic studies in Romance languages*, Washington, Georgetown University Press, p. 89‑100.

SCHANE Sanford A., 1968, *French phonology and morphology*, Cambridge, MIT Press.

SCHANE Sanford A., 1965, *The phonological and morphological structure of French*, Thèse de doctorat, MIT, Cambridge.

SCHMID Helmut, 1994, « Probabilistic Part-of-Speech Tagging Using Decision Trees », Proceedings of 12th International Conference on New Methods in Language Processing, NeMLaP1994, Manchester, Royaume Uni.

SEGUI Juan, 2003, « La perception du langage parlé » dans Bonnet, C, Ghiglione, R. et Richard, J-F. (éds), *Traité de psychologie cognitive: perception, action, langage*, Paris, Dunod, p. 199‑234.

SEIDENBERG M. S., 1997, « Language Acquisition and Use: Learning and Applying Probabilistic Constraints », *Science*, 1997, vol. 275, nº 5306, p. 1599‑1603.

SELINKER Larry, 1972, « Interlanguage », *International Review of Applied Linguistics in LanguageTeaching*, 1972, vol. 10, nº 3, p. 209‑231.

SERIANNI Luca et CASTELVECCHI Alberto, 1988, *Grammatica italiana: italiano comune e lingua letteraria, suoni, forme, costrutti*, Torino, UTET.

SHARWOOD SMITH Michael et KELLERMAN Eric, 1986, « Crosslinguistic influence in second language acquisition: an introduction » dans Kellerman, E. et Sharwood, S. (éds), *Crosslinguistic influence in second language acquisition*, New York, Pergamon Institute of English, p. 1‑9.

SHOEMAKER Ellenor M., 2010, « The Exploitation of Fine Phonetic Detail in the Processing of L2 French » dans VanPatten, B. et Jegerski, J. (éds), *Reserach in Second Language Processing and Parsing*, Amsterdam, J. Benjamins, p. 259‑354.

SHOEMAKER Ellenor M. et BIRDSONG David, 2008, « La résolution de la liaison en français par des locuteurs natifs et non-natifs », *Acquisition et interaction en langue étrangère*, 2008, vol. 27, p. 43‑62.

SIMON Ellen, CHAMBLESS Della et KICKHOFEL ALVES Ubirata, 2010, « Understanding the role of orthography in the acquisition of a non-native vowel contrast », *Language Sciences*, 2010, vol. 32, nº 3, p. 380‑394.

SINGLETON David M. et RYAN Lisa, 2004, *Language acquisition the age factor*, Clevedon, Multilingual Matters.

SOUM-FAVARO Christiane, COQUILLON Annelise et CHEVROT Jean-Pierre (éds), 2014, *La liaison : approches contemporaines*, Berne, Peter Lang.

SPINELLI Elsa, CUTLER Anne et MCQUEEN James, 2003, « Processing resyllabified words in French », *Journal of Memory and Language*, 2003, vol. 48, p. 233‑254.

SPINELLI Elsa et FERRAND Ludovic, 2005, *Psychologie du langage: l'écrit et le parlé, du signal à la signalisation*, Paris, Colin.

SPINELLI Elsa, MCQUEEN James et CUTLER Anne, 2003, « Processing resyllabified words in French », *Journal of Memory and Language*, 2003, vol. 48, p. 233‑254.

SPINELLI Elsa, MCQUEEN James et CUTLER Anne, 2002, « Resolution of liaison for lexical access in French », *Revue française de linguistique appliquée*, 2002, vol. 7, nº 1, p. 83‑96.

SPLENDIDO Frida, 2014, *Le développement d'aspects phonético-phonologiques du français chez des enfants bilingues simultanés et successifs : Le VOT et la liaison dans une étude de cas multiples*, Thèse de doctorat, Université de Lund, Lund.

STERIADE Donca, 1999, « Lexical Conservatism in French Adjectival Liaison » dans Authier, J-M., Bullock, B.E. et Reed, L.A. (éds), *Formal perspectives on romance linguistics*, Amsterdam, J. Benjamins, p. 243-270.

STEVENS Vance, 1991, « Concordance-based vocabulary exercises: A viable alternative to gap-fillers » dans Tim, J. et King, P. (éds), *English Language Research Journal*, Birmingham, University of Birmingham, p. 47‑61.

STEWART Guy, BERNARDINI Silvia et ASTON Guy, 2004, « Introduction: ten years of TaLC » dans Aston, G., Bernardini, S. et Aston, G. (éds), *Corpora and language learners*, Amsterdam, J. Benjamins, p. 1‑18.

STRIDFELDT Monika, 2005, *La perception du français oral par des apprenants suédois*, Thèse de doctorat, University of Umeå, Umeå.

TARONE Elaine, 1987, « Some influences on syllable structure of interlanguage phonology » dans Ioup, G. et Weinberger, S. (éds), *Interlanguage phonology: the acquisition of a second language sound system*, Cambridge, Newbury House Publishers, p. 233‑247.

TARONE Elaine, 1979, « Interlanguage as Chameleon », *Language Learning*, 1979, vol. 29, n° 1, p. 181‑191.

TCHOBANOV Atanas, LAKS Bernard, DURAND Jacques, DETEY Sylvain et LYCHE Chantal, 2009, « Ressources phonologiques au service de la didactique de l'oral : le projet PFC-EF », *Mélanges CRAPEL*, 2009, vol. 31, p. 223‑236.

TENNANT Jeff, sous presse, « Canadian Anglophone Learners' Realization of French Liaison », *L'apprentissage de la liaison en français par des locuteurs non natifs: éclairage des corpus oraux, Bulletin VALS-ASLA*, sous presse, vol. 201.

TERNES Elmar, 1977, « Konsonantische Anlautverânderungen in den keltischen und romanischen Sprachen », *Romanistisches Jahrbuch*, 1977, vol. 28, p. 19‑53.

THOMAS Alain, 2002, « La variation phonétique en français langue seconde au niveau universitaire avancé », *Acquisition et interaction en langue étrangère*, 2002, vol. 17, p. 101‑121.

THOMAS Alain, 2004, « Phonetic norm versus usage in advanced French as a second language », *International Review of Applied Linguistics in Language Teaching*, 2004, vol. 42, n° 4, p. 365-382.

THOMAS Alain, 1998, « La liaison et son enseignement: des modèles orthoépiques à la réalité linquistique », *Canadian Modern Language Review/ La Revue canadienne des langues vivantes*, 1998, vol. 54, n° 4, p. 543‑552.

THUROT Charles, 1881, *De la prononciation française depuis le commencement du XVIe siècle, d'après les témoignages des grammairiens*, Genève, Slatkine Reprints.

TOUTANOVA Kristina, KLEIN Dan, MANNING Christhopher et SINGER Yoram, 2003, « Feature-Rich Part-of-Speech Tagging with a Cyclic Depende ncy Network », Proceedings of 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, HLT-NAACL2003, New York, Etats-Unis, p. 252‑259.

TRANEL Bernard, 2000, « Aspects de la phonologie du français et la théorie de l'optimalité », *Langue Française*, 2000, n° 126, p. 39‑72.

TRANEL Bernard, 1995a, « Current Issues in French Phonology: Liaison and Position Theories » dans Goldsmith, J. (éds), *The handbook of phonological theory*, Cambridge, Blackwell, p. 798‑816.

TRANEL Bernard, 1995b, « French final consonants and nonlinear phonology », *Lingua*, 1995, vol. 95, n° 1-3, p. 131‑167.

TRANEL Bernard, 1990, « On suppletion and French liaison », *Probus*, 1990, vol. 2, nº 2, p. 169‑208.

TRANEL Bernard, 1981, *Concreteness in generative phonology: evidence from French*, Berkeley, University of California Press.

TRANEL Bernard, 1976, « A note of final consonant deletion in modern French », *Lingua Lingua*, 1976, vol. 39, nº 1-2, p. 53‑68.

TRANEL Bernard, 1974, *The phonology of nasal vowels in modern French*, Thèse de doctorat, University of California, San Diego.

TREMBLAY Annie, 2011, « Learning to Parse Liaison-Initial Words: An Eye-Tracking Study », *Bilingualism: Language and Cognition*, 2011, vol. 14, nº 3, p. 257‑279.

TREMBLAY Annie et SPINELLI Elsa, 2014, « Indices acoustico-phonétiques dans la reconnaissance des mots » dans Soum-Favaro, C., Coquillon, A. et Chevrot, J-P. (éds), *La liaison: approches contemporaines*, Berne, Peter Lang, p. 93‑116.

TROUBETZKOY Nikolaï S., 1949, *Principes de phonologie*, Paris, Klincksieck.

TRUDEAU Danielle, 1992, *Les inventeurs du bon usage: 1529-1647*, Paris, Editions de Minuit.

TYNE Henry, 2009, « Corpus oraux par et pour l'apprenant », *Mélanges CRAPEL*, 2009, nº 31, p. 91‑111.

VALDMAN Albert, 2000, « Comment gérer la variation dans l'enseignement du français langue etrangere aux Etats-Unis », *French Review*, 2000, vol. 73, p. 648‑709.

VALDMAN Albert, 1989, « The elaboration of pedagogical norms for second language learners in conflictual diglossia situation » dans Gass, S., Madden, D., Preston, D. et Selinker, L. (éds), *Variation in second language acquisition*, Clevedon, Multilingual Matters, vol.1, p. 15‑34.

VALDMAN Albert, 1986, « Variation et norme pédagogique dans l'enseignement du français langue étrangère » dans Giacomi, A., Véronique, D. (éds), *Acquisition d'une langue étrangère. Perspectives et recherches*, Université de Provence, p. 723‑733.

VALETOPOULOS Freiderikos, sous presse, « Analyse de la liaison dans le discours non spontané des apprenants hellénophones chypriotes de niveau intermédiaire/avancé », *L'apprentissage de la liaison en français par des locuteurs non natifs: éclairage des corpus oraux, Bulletin VALS-ASLA*, sous presse, vol. 201.

VAUGELAS Claude F., 1647, *Remarques sur la langue françoise*, Versailles, Pierre Le Petit.

VERDELHAN-BOURGADE Michèle, 2007, *Le français langue seconde: un concept et des pratiques en évolution*, Bruxelles, De Boeck.

VERONIQUE Daniel, 1995, « Le developpement des connaissances grammaticales en français langue 2. Implications pour une évaluation » dans Chaudenson R. (éds), *Vers un outil d'évaluation des compéteneces linguistiques en fraçais dans l'expace francophone*, Paris, CIRELA/ACCT, p. 29‑45.

VERONIQUE Daniel, 1992, « Recherches sur l'acquisition des langues secondes : un etat des lieux et quelques perspectives », *Acquisition et interaction en langue étrangère*, 1992, n° 1, p. 5‑36.

VIGNER Gérard, 2009, *Le français langue seconde: comment apprendre le français aux élèves nouvellement arrivés*, Paris, Hachette éducation.

VIHMAN Marilyn et CROFT William, 2007, « Phonological development: toward a "radical" templatic phonology », *Linguistics*, 2007, vol. 45, n° 4, p. 683‑725.

VINAY Jean-Paul, 1970, *Traduction automatique et analyse des faits prosodiques*, Montréal, Didier.

VISCEGLIA Tanya, TSENG Chiu-yu, KONDO Mariko, MENG Helen et SAGISAKA Yoshinori, 2009, « Phonetic aspects of content design in AESOP (Asian English Speech cOrpus Project) », Proceedings of 2009 International Committee for the Co-ordination and Standardisation of Speech Databases and Assesment Techniques, Oriental COCOSDA2009, Beijing, China, p. 60‑65.

VOGEL Klaus, 1995, *L'interlangue : la langue de l'apprenant*, Toulouse, Presses universitaires du Mirail.

WALTER Henriette et MARTINET André, 1982, *Enquête phonologique et variétés régionales du français*, Paris, Presses universitaires de France.

WARNANT Léon, 1987, *Dictionnaire de la prononciation française dans sa norme actuelle*, Paris, Duculot.

WARTBURG Walther von, 1988, *Evolution et structure de la langue française*, Berne, Francke.

WAUQUIER GRAVELINES Sophie, 1996, *Organisation phonologique et traitement de la parole continue : contraintes prosodiques et phonologiques de l'accès lexical*, Thèse de doctorat, Université Paris VII, Paris.

WAUQUIER-GRAVELINES Sophie et BRAUD Virginie, 2005, « Proto-determinant et acquisition de la liaison obligatoire en francais », *Langages*, 2005, n° 158, p. 53‑65.

WAUQUIER Sophie, 2009, « Acquisition de la liaison en L1 et L2 : stratégies phonologiques ou lexicales ? », *Acquisition et interaction en langue étrangère*, 2009, p. 93‑130.

WAUQUIER Sophie et SHOEMAKER Ellenor M., 2015, « Processing liaison in L2 French: the case of non-traditional learners », 12th Old World Conference in Phonology (OCP12), Barcelone, Espagne.

WAUQUIER Sophie et SHOEMAKER Ellenor M., 2013, « Convergence and divergence in the acquisition of French liaison by native and non-native speakers: A review of existing data and avenues for future research », *Language, Interaction and Acquisition*, 2013, vol. 4, nº 2, p. 161‑189.

WEINREICH, Uriel, 1953, *Languages in contact,* New York, Linguistic Circle of New York.

WETZELS Leo, 1987, « The Timing of Latent Consonants in Modern French » dans Neidle Carol, Nufiez Cedefio Rafael A. (éds), *Studies in Romance Languages*, Dordrecht, Foris, p. 283‑317.

WHITNEY William Dwight, 1889, *Sanskrit grammar: including both the classical language, and the older dialects, of Veda and Brahmana*, Cambridge, Harvard University Press.

WIDDOWSON Henry, 1998, « Context, community, and authentic language », *TESOL Quarterly,* 1998, vol. 32, p. 705‑716.

YAMAGUCHI Naomi, 2015, « L'acquisition phonologique, de Jackobson aux modèles fréquentiels », *Languages,* 2015, vol. 198, nº 2, p. 31‑50.

YOUNG-SCHOLTEN, Marta, 1998, « Second language syllable simplification: Deviant development or deviant input? » dans Leather, J. et Allen, J. (éds), *New souds 97: Proceedings of the Third International Symposium on the Acquisition of Second-Language Speech*, Amsterdam, University of Amsterdam, p. 351‑360.

YOUNG-SCHOLTEN, Martha, 2002, « Orthographic input in L2 phonological development » dans Burmeister B., Piske t. et a Rohde A. (éds), *An Integrated View of Language Development: Papers in Honour of Henning Wode*, Trier, Wissenschaftlicher Verlag Trier, p. 263‑279.

YOUNG-SCHOLTEN, Martha, AKITA Mamiko. et CROSS Naomi, 1999, « Focus on form in phonology: Orthographic exposure as a promoter of epenthesis » dans Robinson,P. et Jungheim, N.O. (éds), *Pragmatics and Pedagogy. Proceedings of the Third PAcSLRF*, Tokyo, Aoyama Gakuin University, vol. 2, p. 227‑233.

YULE Udny G., 1924, *A mathematical theory of evolution: based on the conclusions of Dr. J.C. Willis, F.R.S.*, London, Printed and published for the Royal Society of London by Harrison and Sons.

ZESIGER Pascal et DE PARTZ Marie-Pierre, 1994, « Perturbations du langage écrit: les dyslexies et les dysgraphies » dans Seron, X. et Jeannerod, M. (éds), *Neuropsychologie humaine*, Liège, Mardaga, p. 419‑437.

ZIEGLER Johannes C., MUNEAUX Mathilde et GRAINGER Jonathan, 2003, « Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation », *Journal of Memory and Language*, 2003, vol. 48, nº 4, p. 779‑793.

ZINK Gaston, 2006, *Phonétique historique du français*, Paris, Presses universitaires de France.

ZIPF George K., 1949, *Human behavior and the principle of least effort; an introduction to human ecology*, Cambridge, Addison-Wesley Press.

Index des notions

A

Acquisition
- Acquisition en milieu institutionnel, 220
- Acquisition guidée, 219, 220, 227, 229, 239, 267, 271
- Acquisition L1, 125, 237, 244, 363, 413, 414
- Acquisition L2, 177, 265, 268, 363, 410, 420
- Acquisition non guidée, 219, 220

Annotation, 155, 158, 160 - 164, 166 - 170, 178, 179, 297, 299, 331, 441 - 448
- Annotation morphosyntaxique, 155, 160 - 163, 169, 445
- Couche d'annotation, 163, 166, 167, 169, 170

Apprentissage sur corpus, 272, 440, 443, 359

C

Codage, 59, 74, 159, 160, 255, 266, 297, 299, 300, 303, 304, 331, 333, 339, 358, 362, 441, 446, 448

Concordance, 431, 436, 440 - 445, 447 - 451, 459, 463, 475, 483, 485

Construction, 103, 107 - 109, 123 - 126, 141, 179, 185 - 188, 190, 195 - 199, 205, 206

Corpus
- Corpus InterPhonologie du Français Contemporain (IPFC), 293, 295 - 297, 300, 304, 305, 308, 310, 312, 314, 327, 328, 331, 333, 334
- Corpus Phonologie du Français Contemporain (PFC) 155 - 164, 167 - 169, 171, 172, 177 - 179, 188 - 190, 192, 196, 197, 201 - 204, 206

D

Distribution
- Distribution de Zipf, 172, 176, 204, 455
- Probabilité distributionnelle, 203

E

Épenthèse, 83, 84, 103, 105 - 107, 109, 116, 123, 128, 159, 160, 190, 205, 221, 303, 358, 433

Épilinguistique, 3, 5, 308, 309, 321, 325, 334, 418, 422, 428, 431, 435, 457

Étiquette, 161 - 164, 166 - 169, 171, 178, 179, 203, 301, 333, 334, 440, 443, 445, 446, 449

F

Flottance
- Flottance squelettale, 120
- Flottance syllabique, 119

Français
- Français de référence, 4, 63, 68, 274, 278, 288, 454, 456, 478, 480
- Français langue étrangère, 293, 305, 306, 310, 331, 334, 344, 348, 355, 356, 362, 366, 399, 409, 410, 411
- Français langue maternelle, 218, 271
- Français langue seconde, 209, 217, 218, 229, 239, 271, 280, 291
- Français standard, 64, 66 - 68, 271, 273, 274, 278, 287, 436, 473, 478
- Français standardisé, 64, 67, 271, 273

Fréquence
- Analyse fréquentielle, 151, 153, 155, 171, 178, 181, 188, 196, 204, 205, 206
- Fréquence de cooccurrence, 2, 43, 44, 45, 60, 123, 124, 197, 317, 346, 347, 373, 415, 457

G

Groupe rythmique, 15, 16, 49, 50, 71, 80, 81, 90, 124, 158, 280, 282, 283

H

Hétérogénéité, 77, 78, 91, 97, 98, 109, 151

Hiatus, 13, 25, 33, 37, 122, 123, 189, 480

I

Idiosyncrasie, 181, 182, 183, 204, 377, 388, 440, 454

Interlangue, 210 - 213, 215, 222, 223, 224, 226, 228 - 239, 253, 260

Itinéraire acquisitionnel, 4, 210, 226, 228, 229, 232, 237, 293, 294, 306, 412, 415, 416, 423, 424

L

Lemme, 136, 171, 187, 190 - 193, 233, 234, 424, 441 - 449
Lexique mental, 129, 131, 136, 138, 139, 141 - 144, 146, 147, 150, 205, 206
Liaison
- Fausse liaison, 80, 82, 83, 106, 116
- Liaison catégorique, 78 - 80, 87, 88, 91 - 93, 95, 97, 333, 334, 337 - 343
- Liaison enchaînée, 33, 59, 98, 159, 160, 339, 356, 357, 406
- Liaison erratique, 78, 80 - 82, 87, 94, 96, 259, 297, 314, 317, 319, 321, 331, 334, 340, 366, 388, 399
- Liaison non attendue, 82, 83, 300, 303, 334, 366 - 368
- Liaison non-enchaînée, 34 -36, 258 - 260, 262, 264, 266, 267, 268
- Liaison variable, 78, 79, 86, 87, 91, 94, 96 - 98, 171, - 183, 338 - 341
- Réalisation orthographique, 251 - 253, 257, 267, 268, 282, 358 - 362, 369, 374, 376, 377, 382, 385,
- Substitution, 244, 251 - 253, 260, 262, 303, 323, 358, 359, 363 -366, 374, 379, 380, 385, 392

N

Non-réalisation, 111 - 115, 244, 245, 251 - 253, 262, 339 -356, 362, 370 - 374, 376, 377

O

Orthographe, 111, 131, 133, 135, 136, 143, 144, 146 - 151, 158, 239, 264, 266, 267, 268, 279, 281, 282

P

Phonologie
- Crible phonologique, 217
- Généralisation phonologique, 127, 251, 306, 409, 420, 421, 426, 431 - 434, 450

R

Règle de troncation, 104, 111, 112, 113, 115, 116, 119
Représentation
- Représentation orthographique, 263, 264, 307, 308, 413, 419 - 421, 423, 425, 428
- Représentation phonologique, 242, 262, 263, 267, 308, 413, 419, 431, 433
Ressource pédagogique, XI, 6, 279, 280, 287, 306
Resyllabation, 1, 5, 13, 14, 16, 17, 21, 33, 34, 36, 101, 102, 114, 115, 124

S

Sandhi
- *Sandhi* externe, 2, 9, 10, 12, 16, 28, 33, 101, 288, 307, 453
- *Sandhi* interne, 9, 11
Stratégie
- Stratégie d'apprentissage, 210, 212, 235, 236, 263
- Stratégie de communication, 212, 235, 236
Syntaxe
- Cohésion syntaxique, 1, 16, 37, 40 - 42, 44, 60, 97, 108, 117, 124, 196, 197, 343, 415

T

Transcription, 158 - 160, 162, 178, 255, 280, 284, 294, 297, 298, 299, 312, 324, 327, 331, 362, 434, 441, 442, 466, 483
Transfert, 212, 213, 214, 215, 224, 225, 252, 346

Index des noms

Å

Ågren, 32, 35 - 37, 44 - 46, 49, 50, 57, 60, 72, 73, 89, 90, 91, 92

A

Andreassen, 254 - 258, 260, 267
Ashby, 53, 54, 91
Auroux, 69, 462
Avanzi, 161 - 163, 178

B

Bagou, 131 - 133, 138, 140, 141, 142
Baroni, 173 - 175, 178, 180
Bartning, 4, 226, 229, 230, 232, 233, 245, 306
Bassetti, 221, 222, 267
Beacco, 236, 274, 276
Benzitoun, 93, 161, 162, 164, 166, 204
Bertrand, 278
Billières, 64, 67, 75, 271
Blanche-Benveniste, 27
Bonami, 96, 101, 105, 188, 189
Bordal, 57
Borrel, 64, 67
Boula de Mareüil, 2, 59, 79, 99, 171, 176, 177, 273, 278
Bourciez, 24, 25, 27
Boutin, 57
Bowers, 148, 149
Boyé, 96, 101, 105, 189
Burki, 138 - 142
Burov, 1, 4, 9, 10, 11, 12, 16
Burt, 213, 226
Bybee, 2, 38, 44, 60, 94, 108, 109, 123, 124, 125, 127, 143, 177, 188, 191 - 197

C

Cappeau, 297
Carton, 64, 67, 68, 271, 273, 281, 283
Catach, 27
Chambers, 286, 288,
Charliac, 281, 282
Chevrot, 2, 36, 38, 44, 125 - 129, 196 - 198, 239, 243, 261 - 263, 266, 268, 307
Chomsky, 110, 113, 114, 116, 117, 216
Christodoulides, 161, 162, 164, 178, 440, 441 - 463
Coquillon, 56, 129
Corder, 211, 212, 214, 223
Coseriu, 51
Coste, 218
Côté, VII, 2, 16, 17, 38, 40, 41, 56, 57, 77, 83, 84, 101 - 109, 128, 129, 155, 177, 185, 186, 192, 196 - 205
Coustenoble, 34
Cuq, 218

D

Dabene, 218
Damian, 148, 149
Damourette, 1, 38, 60, 196
De Jong, 2, 46, 47, 53, 60, 70 - 75, 244, 353, 454
De Moras, 2, 239, 240, 246 - 250, 266, 320, 333, 357, 360, 399
Debaisieux, 286, 468
Delattre, 12, 14, 17, 18, 20, 32, 33, 34, 35, 37, 38, 40 - 44, 47 - 50, 57, 59, 60, 70 - 72, 78, 79 - 82, 84, 85, 87, 88 - 96, 109
Dell, 104, 114, 115, 117, 119, 135, 136, 138, 469
Detey, 134, 135, 145 - 147, 244, 250, 254, 255, 256 - 258, 261, 268, 274, 275, 277 - 280, 282, 285 - 288, 293, 295, 297
Domergue, 25, 82
Dostie, 167
Dugua, 38, 125, 126, 128, 196, 243, 262
Dulay, 213, 226
Dupoux, 134, 145, 147, 148
Durand, 15, 18, 31, 35, 36, 43, 47, 53, 54, 55, 56, 57, 59, 70, 74, 75, 79, 80, 83, 84, 89, 94 - 99

E

Eckman, 216, 260

Encrevé, 16, 33 - 36, 43, 47, 50, 52, 54, 56, 70 - 94, 98, 102, 104 - 128
Estienne, 25, 26, 27
Evert, 173, 174, 178, 180, 462
Eychenne, 56, 73, 74, 84, 123, 279, 331

F

Falbo, 254 - 261, 263, 266, 304, 307, 366, 374, 410, 417, 421, 428, 456, 457
Fayol, 125, 126, 262
Ferrand, 132, 133, 146, 147, 149, 150, 151, 206, 419, 421, 428
Floquet, 5, 304, 305
Fouché, 13, 14, 16, 17, 18, 20, 33, 43, 50, 57, 65, 68 - 75, 79, 80, 82, 84 - 87, 101, 102, 109, 110
Frauenfelder, 139, 151, 235
Frei, 83

G

Gadet, 51, 52, 53, 58, 80, 81, 83, 275, 276, 277
Galazzi, 34, 64, 68, 217, 273, 274, 275 - 278, 280, 304, 309, 311, 320
Gaskell, 137, 139
Gass, 216, 222
Givón, 228, 229
Goldberg, 108, 123, 126
Goldsmith, 117
Gombert, 321, 325, 418, 422, 428, 435
Gougenheim, 24, 25, 40, 108
Grainger, 133, 149 - 151, 206, 419, 421, 428
Grammont, 13, 15, 16, 32, 42, 44, 50, 54, 60, 65, 68, 75, 81
Grevisse, 68, 69, 82, 84, 85, 86, 109, 110
Grosjean, 137, 142, 147
Gut, 294, 471

H

Hallé, 147, 475
Harnois-Delpiano, 2, 3, 239, 240, 243, 244, 248, 251, 252, 261, 262, 263
Hernandez, 189, 475
Howard, 2, 239, 240, 244, 245, 246 - 250, 263, 266, 333, 357, 360

K

Kawaguchi, 244, 250, 254 - 258, 261, 280, 293, 295, 422, 453
Kaye, 16, 35, 49, 50, 70, 75, 81, 84, 89, 99, 106, 107, 108, 124, 177, 188, 190, 191
Klausenburger, 104, 106, 107, 116
Klein, 210, 219, 220, 226, 227, 228, 230
Krashen, 213, 216, 220

L

Labov, 78
Lacheret-Dujour, 15, 476
Laganaro, 138, 139, 140, 141, 142
Laks, 1, 2, 3, 4, 12, 15, 18, 27, 28, 33, 36, 42 - 45, 53 - 56, 59 - 64, 68, 70, 73, 75, 80, 84, 90, 93, 94, 97, 106, 109, 110, 121, 125
Langacker, 108, 123
Lauret, 277, 281, 285, 288
Le Prieult, 26, 61, 63, 68, 110
Léon, 17, 18, 50, 64, 66, 70, 71, 72, 81, 82, 90, 101, 102, 241, 271, 273, 281 - 285
Levelt, 135, 136, 138, 233
Liberman, 117
Long, 191, 193, 222, 226, 237
Loporcaro, 10
Lucci, 35, 49, 50, 58, 70, 72, 73, 75
Lyche, 31, 43, 47, 53, 55 - 57, 59, 64, 70, 75

M

Malécot, 18, 33, 53, 57, 65, 70, 72, 90, 91, 92, 245, 247, 281, 284
Mallet, 2, 18, 31, 35, 39, 47, 52, 53, 54, 58, 59, 75, 77, 78, 80, 83, 89, 94, 96, 98
Marslen-Wilson, 132, 134
Martinet, 1, 12, 13, 21, 25, 66, 67, 273
Mastromonaco, 2, 239, 244, 245, 246 - 253, 259, 263, 266, 333, 357, 360, 363
Mehler, 134, 265, 468, 471
Molinari, 275, 276, 304
Montron, 432
Morin, 33, 35, 49, 50, 64 - 68, 70, 75, 81, 84, 89, 99, 101, 106 - 108, 110, 124, 126
Mougeon, 274, 275, 422, 435

N

Nespor, 11, 12, 307
New, 317
Nguyen, 142
Niedermann, 13, 21
Noyau, 211, 220, 226
Nyrop, 72, 80, 81, 92

P

Passy, 14, 15, 34, 280
Paternostro, 275, 286, 304, 331
Perdue, 226, 228, 230, 294
Pichon, 1, 38, 60, 65, 196
Pienemann, 212, 226, 227, 233 - 235, 412, 413, 423, 424, 427
Pierrehumbert, 126
Porquier, 211, 212, 235, 236, 274
Prince, 117, 121, 294
Pulgram, 13, 15, 16, 21
Pustka, 254, 256, 258, 259, 261, 262, 285, 288

R

Racine, 129 - 133, 143, 147, 148, 149, 250, 251, 254, 255 - 261, 264, 266 - 268, 274, 275, 277, 279, 282, 285 - 288, 293, 295, 297 - 299
Renou, 9,
Ruvoletto, 126, 262, 410

S

Schane, 104, 111 - 115, 122, 189
Scheer, 36, 110, 128
Schlyter, 4, 226, 229, 230, 232, 233, 306, 412, 415, 423 - 427
Schmid, 161, 162, 164, 172
Selinker, 211, 212, 222, 235, 237
Selkirk, 11, 12, 15
Shoemaker, 125, 126, 240, 241, 242, 243, 261, 262, 263, 265
Spendido, 243
Spinelli, 132, 133, 137, 138, 142, 146, 147, 242
Steriade, 19, 33, 189
Stridfeldt, 2, 239, 241, 242

T

Tarone, 216, 236
Tchobanov, 286, 287, 288
Tennant, 254, 256, 258 - 260, 263, 264, 266, 267, 307, 308
Thomas, 239, 240, 244 - 250, 252, 253, 259, 260, 261, 263, 266 - 268, 279
Tranel, 101, 102, 104 - 106, 116, 117, 118, 121 - 123
Tremblay, 138, 142, 240 - 242
Troubetzkoy, 11, 217
Tyne, 275, 276, 286, 288

V

Valdman, 277, 285, 287
Valetopoulos, 254, 255, 256, 258, 259, 266, 362
Véronique, 229, 235
Vogel, 11, 12, 212

W

Wauquier, 3, 36, 125 - 128, 137, 142, 241, 243, 253, 259, 260 - 265, 267, 307, 410, 411, 414, 416, 417, 420
Wetzels, 118, 189

Y

Young-Scholten, 221, 267

Z

Ziegler, 133 - 135, 150
Zink, 21 - 24
Zipf, 46, 47, 172 - 174, 176, 178, 180, 204

Table des illustrations

Tableau 1 : Contextes morphosyntaxiques de liaison potentielle (extraits du tableau détaillé de Delattre 1947) 40
Tableau 2 : Points d'enquête du corpus PFC 157
Tableau 3 : Comparaison des jeux d'étiquettes utilisés par les différents annotateurs morphosyntaxiques 166
Tableau 4 : Schéma d'annotation des disfluences proposé par *DisMo* 168
Tableau 5 : Exemple construit de liste de fréquence 172
Tableau 6 : Exemple construit de classification rang-fréquence 173
Tableau 7 : Exemple construit de spectre de fréquence 173
Tableau 8 : Types de liaison variables réalisées les plus fréquents (n. tokens > 100) .. 179
Tableau 9 : Résultats de l'analyse d'ajustement (*goodness-of-fit*) du modèle statistique paramétrique LNRE aux données observées 180
Tableau 10 : Comparaison entre les types de liaison variable les plus fréquents (n. tokens > 100) (en gras) et les types issus des mêmes classes morphosyntaxiques 182
Tableau 11 : Taux de réalisation de la liaison par conjonction des types « CON:sub-PRO:per:sjt » et « CON:coo-PRO:per:sjt » 183
Tableau 12 : Taux de réalisation de la liaison par préposition du type « PRP-DET:ind » 184
Tableau 13 : Taux de réalisation de la liaison par forme verbale des types VER:pres-DET:ind, VER:impf-DET:ind et VER:ppas-DET:ind 185
Tableau 14 : Taux de réalisation de la liaison par préposition des types « PRP-NOM:com » et « PRP-NOM:pro » 185
Tableau 15 : Taux de réalisation de la liaison par préposition des types « ADV:deg-ADJ » et « ADV:neg-ADJ » 186
Tableau 16 : Taux de réalisation de la liaison par préposition des types « VER:pres-PRP » et « VER :impf-PRP » 186
Tableau 17 : Taux de réalisation de la liaison par préposition du type « ADJ-NOM » 187
Tableau 18 : Constructions internes aux types de liaison variable les plus fréquents (n. tokens > 100) 188
Tableau 19 : Réalisations de la liaison après adjectifs singuliers (lemme) du type ADJ-NOM 191

Tableau 20 : Réalisations de la liaison après adjectifs pluriels (lemme) du type ADJ-NOM 192
Tableau 21 : Caractéristiques distributionnelles (antéposition / postposition par rapport au nom) tous contextes confondus des adjectifs « gros » et « long »..... 193
Tableau 22 : Calcul de l'indice de la fréquence d'une séquence de liaison d'après Chevrot et al. 2007 (tiré de Côté 2013)..... 197
Tableau 23 : Résultats des réalisations de la liaison après des mots invariables monosyllabiques (Côté 2013)..... 198
Tableau 24 : Probabilités transitionnelles des adverbes monosyllabiques (Côté 2013)200
Tableau 25 : Hiérarchie implicationnelle de la théorie de la processabilité de Pienemann (1998) 234
Tableau 26 : Bilan des études sur l'acquisition de la liaison en L2 241
Tableau 27 : Comparaison des taux de réalisation de la liaison par type de liaison..... 249
Tableau 28 : Études sur l'acquisition de la liaison en L2 issues du projet IPFC 255
Tableau 29 : Comparaison entre les taux de réalisation (catégorique et variable) par tâche (lecture et parole) extraits des études du projet IPFC 256
Tableau 30 : Comparaison entre les taux de réalisation des liaisons non-enchaînées extraits des études du projet IPFC..... 258
Tableau 31 : Séances de l'étude longitudinale IPFC-italien (Rome)..... 305
Tableau 32 : Profil socio-culturel des locuteurs du corpus IPFC-italien (Rome)..... 310
Tableau 33 : Profil linguistique des locuteurs du corpus IPFC-italien (Rome)..... 311
Tableau 34 : Protocole enquête IPFC-italien (Rome)..... 314
Tableau 35 : Types et séquences de liaison issues de la tâche de répétition des séquences isolées..... 315
Tableau 36 : Types et séquences de liaison issus de la tâche de lecture des séquences isolées..... 317
Tableau 37 : Types et séquences de liaison issus de la tâche de lecture du texte PFC. 319
Tableau 38 : Types et séquences de liaison issus de la tâche de lecture du texte journalistique..... 322
Tableau 39 : Types et séquences de liaison issus du test de jugement 325
Tableau 40 : Séquences de liaison de la tâche de dénomination d'images..... 327
Tableau 41 : Répartition des sujets en deux groupes (écrit et oral) pendant la tâche de reformulation..... 329
Tableau 42 : Types et séquences de liaison de la tâche de reformulation 329

Tableau 43 : Types de liaison potentielle catégorique et variable (étude longitudinale) 333
Tableau 44 : Comparaison longitudinale des taux de production des liaisons catégoriques et variables conformes à la CL cible par type de liaison 335
Tableau 45 : Résultats du test statistique non paramétrique Kruskal-Wallis mené sur les données de production des liaisons catégoriques conformes à la CL cible (étude longitudinale) 335
Tableau 46 : Résultats du test statistique non paramétrique Kruskal-Wallis mené sur les données de production des liaisons variables conformes à la CL cible (étude longitudinale)336
Tableau 47 : Résultats totaux de l'étude longitudinale – réalisation des liaisons catégoriques et variables 338
Tableau 48 : Résultats totaux - taux de non-réalisation des liaisons catégoriques et variables (étude longitudinale) 339
Tableau 49 : Taux de non-réalisation des liaisons catégoriques et variables par type de liaison (étude longitudinale) 340
Tableau 50 : Taux de non-réalisation et des réalisations de la liaison du type DET-NOM par M1 (étude longitudinale) 341
Tableau 51:Taux de non-réalisation et des réalisationd de la liaison des types PRO-AUX, PRO-PRO, PRO-VER et VER-PRO par M1 (étude longitudinale) 343
Tableau 52 : Taux de non-réalisation et des réalisations de la liaison des types EXF-EXF (étude longitudinale) 344
Tableau 53 : Taux de non-réalisation et des réalisations de la liaison du type NUM-NOM par M1 (étude longitudinale) 344
Tableau 54 : Taux de non-réalisation et des réalisations de la liaison des paires de séquences ayant de indices de cooccurrence différents (lecture-étude longitudinale) 345
Tableau 55 : Comparaison entre le taux de non-réalisation de la liaison variable chez les apprenants FLE (corpus IPFC-italien Rome) et chez les natifs (corpus PFC) 346
Tableau 56 : Taux de non-réalisation et des réalisations de la liaison du type ADJ-NOM par M1 (étude longitudinale) 347
Tableau 57 : Taux de non-réalisation et des réalisations de la liaison du type ADV-ADJ, ADV-ADV, ADV-DET, ADV-INF, ADV-NOM, ADV-PPA, ADV-PRO, ADV-PRP par M1 (étude longitudinale) 348
Tableau 58 : Taux de non-réalisation et des réalisations de la liaison des types NOM-ADJ, NOM-ADV, NOM-AUX, NOM-PRP et NOM-VER par M1 (étude longitudinale) 350

Tableau 59 : Taux de non-réalisation et des réalisations de la liaison des types PRP-DET, PRP-NAM, PRP-NOM et PRP-PRO par M1 (étude longitudinale) 351

Tableau 60 : Taux de non-réalisation et des réalisations de la liaison des types VER-ADJ, VER-ADV, VER-DET, VER-INF, VER-NOM, VER-PDE et VER-PRP par M1 (étude longitudinale) 352

Tableau 61 : Taux de non-réalisation et des réalisations de la liaison des types AUX-ADV, AUX-INF, AUX-PPA et AUX-PRP par M1 (étude longitudinale) 353

Tableau 62 : Taux de non-réalisation et des réalisations de la liaison des types INF-ADV, INF-DET et INF-PRP par M1 (étude longitudinale) 354

Tableau 63 : Résultats totaux - taux d'enchaînement des liaisons catégoriques et variables (étude longitudinale) 354

Tableau 64 : Résultats totaux - taux de non-enchaînement des liaisons catégoriques et variables (étude longitudinale) 355

Tableau 65 : Résultats totaux - taux de pause-hésitation des liaisons catégoriques et variables (étude longitudinale) 355

Tableau 66 : Résultats totaux - taux de non-compréhensions des liaisons catégoriques et variables (étude longitudinale) 356

Tableau 67 : Résultats totaux (liaison catégorique et variable) ‑ nature de la CL (étude longitudinale) 356

Tableau 68 : Résultats totaux ‑ nature de la CL par typologie de liaison (étude longitudinale) 357

Tableau 69 : Taux des réalisations orthographiques de la CL par type de liaison (étude longitudinale) 359

Tableau 70 : Taux des réalisations non conformes de la CL mais présentes dans le M1 par type de liaison (étude longitudinale) 361

Tableau 71 : Réalisations non conformes avec substitution de la CL par type (étude longitudinale) 362

Tableau 72 : Réalisations non conformes avec substitution de la CL par type de consonne (étude longitudinale) 363

Tableau 73 : Liaisons erratiques par type et token de liaison (étude longitudinale) 364

Tableau 74 : Liaisons non attendues par type de liaison (étude longitudinale) 365

Tableau 75 : Liaisons non attendues par type et token de liaison (étude longitudinale) 366

Tableau 76 : Comparaison longitudinale des taux de réponses totales des liaisons catégoriques issus du test de jugement 369
Tableau 77 : Comparaison longitudinale des taux de réponses totales des liaisons variables issus du test de jugement 370
Tableau 78 : Types de liaison potentielle catégorique et variable (étude transversale) 373
Tableau 79 : Résultats totaux - réalisations de la liaison (tâche de dénomination d'image) 374
Tableau 80 : Taux de réalisation orthographique de la CL par type de liaison (tâche de dénomination d'image) 374
Tableau 81 : Comparaison entre les taux de réalisation orthographique de la CL issus de la lecture des séqueces isolées de liaison (étude longitudinale) et de la tâche de dénomination d'image (étude transversale) 375
Tableau 82 : Résultats totaux - taux de réalisation par type de liaison (tâche reformulation) 376
Tableau 83 : Groupe oral - taux de réalisation par type de liaison (tâche reformulation) 377
Tableau 84 : Groupe écrit - taux de réalisation par type de liaison (tâche reformulation) 377
Tableau 85 : Résultats totaux – nature de la CL (tâche reformulation) 378
Tableau 86 : Groupe oral - nature de la CL (tâche reformulation) 379
Tableau 87 : Groupe écrit - nature de la CL (tâche reformulation) 380
Tableau 88 : Types de liaison potentielle catégorique et variable (tâche de répétition) 382
Tableau 89 : Résultats totaux – Réalisations par type de liaison (tâche de répétition). 383
Tableau 90 : Résultats totaux - nature de la CL par type de liaison (tâche de répétition) 384
Tableau 91 : Types de liaison potentielle catégorique et variable (tâches de lecture).. 385
Tableau 92 : Résultats totaux – Réalisations par type de liaison (tâches de lecture) ... 389
Tableau 93 : Résultats totaux - Nature de la CL par type de liaison (tâches de lecture) 393
Tableau 94 : Types de liaison potentielle catégorique et variable (tâches de parole semi-spontanée) 396
Tableau 95 : Résultats totaux - Réalisations par type de liaison (tâches de parole semi-spontanée) 401

Tableau 96 : Résultats totaux - Nature de la CL par type de liaison (tâches de parole semi-spontanée)..... 404
Tableau 97 : Comparaison des taux d'enchaînement des liaisons catégoriques entre les trois tâches de production orale..... 404
Tableau 98 : Comparaison des taux d'enchaînement des liaisons variables entre les trois tâches de production orale..... 404
Tableau 99 : Comparaison des taux de non-enchaînement des liaisons totales entre les trois tâches de production orale..... 405
Tableau 100 : Comparaison des taux de réalisation des CL orthographiques totales entre les trois tâches de production orale..... 405
Tableau 101 : Comparaison des taux d'enchaînement des types DET-ADJ et DET-NOM entre les trois tâches de production orale..... 406
Tableau 102 : Comparaison des taux d'enchaînement des types PRO-VER et PRO-AUX entre les trois tâches de production orale..... 406
Tableau 103 : Itinéraire d'acquisition de la liaison catégorique en FLE..... 422

Table des exemples

Exemple 1 : Énoncé annoté en liaison par le biais du codage liaison PFC 160

Exemple 2 : Annotation du pronom personnel objet « les » réalisée par le biais de l'annotateur *DisMo* 167

Exemple 3 : Annotation du marqueur discursif « tu vois » réalisée par le biais de l'annotateur *DisMo* 167

Exemple 4 : Annotation du faux départ « les ac/ » réalisée par le biais de l'annotateur *DisMo* 168

Exemple 5 : Annotation du marqueur discursif « bon » réalisée par le biais de l'annotateur *Treetagger* 168

Exemple 6 : Annotation du marqueur discursif « bon » réalisée par le biais de l'annotateur *DisMo* 168

Exemple 7 : Annotation de l'adverbe composé « tout de suite » réalisée par le biais de l'annotateur *DisMo* 169

Exemple 8 : Annotation de l'adverbe composé « petit à petit » réalisée par le biais de l'annotateur *Treetagger* 169

Exemple 9 : Annotation de l'adverbe composé « petit à petit » réalisée par le biais de l'annotateur *DisMo* 169

Exemple 10 : Excercices pour l'acquisition de la liaison (Léon et Léon 1997) 284

Exemple 11 : Exemple de séquence de liaison codée selon le codage alphanumérique liaison IPFC 300

Table des figures

Figure 1: Tableau simplifié de la classification de Delattre (1947 ; 1966)..... 89
Figure 2 : Les statuts lexicaux des consonnes de liaison (CL) (Côté 2010, p. 1280)... 103
Figure 3: Règle de Troncation (Schane 1965) 112
Figure 4 : Règle de Suppression (Schane 1974) 113
Figure 5 : Règle de Suppression de la CL (Schane 1978) 114
Figure 6 : Règle de Liaison (Dell 1970) 114
Figure 7 : Règle de Troncation (Dell 1970)..... 115
Figure 8 : Règle de dénasalisation de la voyelle (Tranel 1976)..... 117
Figure 9 : Flottance syllabique de la CL..... 119
Figure 10 : Extrasyllabicité de la CL du mot « petit » (Clements et Keyser 1983)..... 119
Figure 11 : Convention universelle d'association des codas flottantes (Encrevé 1988) 121
Figure 12 : Construction à schéma ouvert de la réalisation de la liaison dans le type « nom-adjectif » (Bybee 2005) 125
Figure 13 : Construction à schéma ouvert de la non-réalisation de liaison dans le type « nom-adjectif » (Bybee 2005)..... 125
Figure 14 : Hypothèse de l'Indépendence entre système phonologique et orthographique (*orthographic autonomy hypothesis*) (tiré de Rapp, Benzing et Caramazza 1997) 144
Figure 15 : Hypothèse de la médiation phonologique (*phonological mediation hypothesis*) (tiré de Rapp, Benzing et Caramazza 1997, p. 75)..... 146
Figure 16 : Modèle interactif bimodal (Grainger et Ferrand 1996 ; Ziegler, Muneaux et Grainger 2003) 150
Figure 17 : Comparaison des courbes rang-fréquence représentées dans un repère bilogarithmique (tirées de Baroni 2008, p. 12‑13) 175
Figure 18 : Distribution de la fréquence des réalisations de la liaison variable dans la parole semi-spontanée du corpus PFC 179
Figure 19 : Comparaison des deux spectres de fréquences issus du modèle LNRE (*fZM model*) et des données observées (*observed)*..... 181
Figure 20 : Caractéristiques distributionnelles (antéposition / postposition par rapport au nom) tous contextes confondus des adjectifs singuliers observés en contexte de liaison..... 193
Figure 21 : Fréquence relative des liaisons non-réalisées par adjectif 194

Figure 22 : Analyse syntaxique de l'énoncé « ils ont mal, ils arrivent plus à marcher » (Corpus PFC, Grenoble, 38aca1gg) réalisée par le biais de l'analysateur syntaxique *FRMG Wiki* (De la Clergerie et al. 2009) 202

Figure 23 : Analyse syntaxique de l'énoncé « C'est plus intéressant d'enseigner au lycée » (Corpus PFC, Toulouse, 31afb1lg) réalisée par le biais de l'analysateur syntaxique *FRMG Wiki* (De la Clergerie et al. 2009)..... 202

Figure 24 : Analyse syntaxique de l'énoncé « ma grand-mère était placée aussi quand elle était plus jeune » (Corpus PFC, Grenoble 38agp1lg) réalisée par le biais de l'analysateur syntaxique *FRMG Wiki* (De la Clergerie et al. 2009) 203

Figure 25 : Analyse syntaxique de l'énoncé « ils parlent pas parfaitement mais on les comprend » (Corpus PFC, Toulouse 31atc1gg) réalisée par le biais de l'analysateur syntaxique *FRMG Wiki* (De la Clergerie et al. 2009)..... 203

Figure 26 : Interactions entre représentations phonologiques et orthographiques dans la production de la liaison (Grainger et Ferrand 1996) 206

Figure 27 : Interface pour la passation automatique du test de jugement..... 323

Figure 28 : Exemple d'une diapositive de la tâche de dénomination d'images 327

Figure 29 : Exemple d'une diapositive de la tâche de reformulation 328

Figure 30 : Comparaison des taux de production des liaisons catégoriques conformes à la CL cible entre les quatre séances de l'étude longitudinale 336

Figure 31 : Comparaison des taux de production des liaisons variables conformes à la CL cible entre les quatre séances de l'étude longitudinale 337

Figure 32 : Comparaison longitudinale des taux de réponses correctes des liaisons catégoriques issus du test de jugement..... 367

Figure 33 : Comparaison longitudinale des taux de réponses correctes des liaisons variables issus du test de jugement 368

Figure 34 : Interaction entre représentations phonologiques et orthographiques des mots de liaison (Grainger et Ferrand 1996) dans la production des réalisations orthographiques de la liaison chez les apprenants FLE 417

Figure 35 : Développement des généralisations phonologiques de la CL chez les apprenants de FLE. Persistance interactions entre représentations phonologiques et orthographiques des mots de liaison (Grainger et Ferrand 1996) 419

Figure 36: Premier niveau de la classification multi-niveaux 435

Figure 37 : Affichage icône des indices de fréquence de réalisation (niveau 1) 436

Figure 38 : Sélection de l'histogramme des liaisons réalisées (niveau 1) 436

Figure 39 : Deuxième niveau de la classification multi-niveaux du type ADJ-NOM.. 437
Figure 40 : Deuxième niveau de la classification multi-niveaux du type ADV-ADJ .. 438
Figure 41 : Trois champs de l'interface du concordancier multi-niveaux et multimédia 442
Figure 42 : Champ de requête des concordances..... 443
Figure 43 : Champ de sélection des attributs de recherche à afficher dans les résultats 445
Figure 44 : Affichage des résultats de la requête « PRO:per:sjt » - lemme du verbe « être » - « en » - « train » 446
Figure 45 : Concordances issues de la requête « pas-ADJ » (attributs : Token(min)-POS(min)) 447
Figure 46 : Concordances issues de la requête « être-DET » (attributs : Lemma-POS(min)) 448
Figure 47 : Concordances issues de la requête « ADJ-NOM » (attributs : POS(min)-POS(min)) 448

Table des matières

Sommaire V
Remerciements VII
Sigles et abréviations IX
Résumé XI
Abstract XIII
Abstract XV
Introduction 1

Partie 1. La liaison : origines, classifications, traitements et représentations 7

Chapitre 1. La liaison : un cas particulier de *sandhi* externe 9

1. Le phénomène de sandhi 9
2. Les phénomènes de sandhi externe en français : l'enchaînement et la liaison 12

2.1. L'enchaînement 12
2.2. Le français langue cursus 15
2.3. La liaison : un cas particulier de sandhi externe 16

3. Consonnes finales : consonnes fixes et consonnes de liaison 17
4. Typologies des CL 18
5. Évolution diachronique des consonnes finales en français 21
6. Etablissement d'une norme orthographique et d'usage de la liaison 25
7. Conclusion 28

Chapitre 2. Variables linguistiques d'étude de la liaison 31

1. Liaisons morphosyntaxiques et phonologiques 31
2. Analyse des variables linguistiques 33

2.1. Variables phonologiques 33

2.1.1. Liaisons enchaînées ou non-enchaînées 33

2.2. Variables phonétiques 37
2.3. Variables syntaxiques 37

2.3.1. Degrés de cohésion syntaxique 37
2.3.2. Liaisons antéposées et postposées à une tête syntaxique 42

2.4. Variables lexicales 43

2.4.1. Cohésion syntagmatique comme fréquence de cooccurrence des mots de liaison 44

2.4.2. Fréquence des mots de liaison 45

2.4.3. Longueur des mots de liaison 47

2.5. Variables prosodiques 48

2.5.1. La longueur du groupe syntaxique auquel appartiennent le M1 et le M2 48

2.5.2. L'intonation de la phrase 49

2.5.3. L'accent d'insistance 50

2.5.4. Pauses 50

2.5.5. Débit de parole 50

2.6. Variables sociolinguistiques 51

2.6.1. Variation inter-locuteur 52

2.6.1.1 Variation diastratique 52

2.6.1.2 Variation diachronique 54

2.6.1.3 Variation diatopique 56

2.6.2. Variation intra-locuteur 57

2.6.2.1 Variation diaphasique 57

2.6.2.2 Variation diamésique 58

3. Conclusion 60

Chapitre 3. Normes de prononciation et approches méthodologiques pour l'étude de la liaison 63

1. Quel français de référence pour la description de la liaison ? 63

1.1. Le français de la « bonne société parisienne » 64

1.2. Le français des medias : « le français standard » 66

1.3. Le français « qui passe inaperçu » : « le français standardisé » 67

2. Approches méthodologiques d'étude de la liaison 68

2.1. Approches normatives 68

2.2. Approches descriptives 70

2.2.1. Premières études ... 70
2.2.2. Etudes systématiques de liaison ... 73
3. Conclusion ... 74
Chapitre 4. Typologies et classifications de la liaison ... 77
1. Types et tokens de liaison ... 77
2. Tripartition des contextes de liaison ... 78
2.1. Différentes dénominations ... 78
2.2. Liaisons catégoriques ... 79
2.3. Liaisons variables ... 79
2.4. Entre erreurs de performance et « fausses liaisons » ... 80
2.4.1. Liaisons catégoriques non réalisées ... 80
2.4.2. Liaisons erratiques ... 80
2.4.2.1 Raisons prosodiques : liaison erratique entre deux groupes rythmiques et après la conjonction « et » ... 81
2.4.2.2 Raisons euphoniques ... 81
2.4.2.3 Raisons « d'intelligibilité » : liaison erratique avec des « termes disjoints » ou des citations ... 82
2.4.2.4 Contraintes étymologiques : devant « h » aspiré ... 82
2.4.3. « Fausses liaisons » chez les natifs ... 82
3. Classifications ... 84
3.1. Classifications normatives ... 85
3.1.1. Grevisse (1936) ... 85
3.1.2. Fouché (1959) ... 86
3.2. Classifications descriptives ... 87
3.2.1. Delattre (1947 ; 1966) ... 87
3.2.2. Léon (1966) ... 90
3.2.3. Encrevé (1988) ... 90
3.2.3.1 La variabilité de la liaison après les mots invariables (adverbes et prépositions) monosyllabiques ... 91
3.2.3.2 Premières données orales sur la liaison après adjectifs épithètes antéposés ... 92

3.2.4. Durand et Lyche (2008) 94

3.2.4.1 La variabilité de la liaison après la construction impersonnel « c'est » 94

3.2.4.2 Données supplémentaires pour l'étude de la liaison après adjectifs épithètes antéposés 94

3.2.4.3 Données supplémentaires pour l'étude de la variabilité de la liaison après prépositions monosyllabiques 95

3.2.4.4 La liaison entre noms pluriels et verbe comme contexte de liaison erratique 96

3.2.5. Mallet (2008) 96

3.2.6. Durand et al. (2011) 97

4. Conclusion : hétérogénéité de la liaison et difficultés de prédiction des réalisations 98

Chapitre 5. Traitement et statut de la consonne de liaison : approches linguistiques 101

1. Consonnes de liaison et consonnes fixes : deux statuts et deux traitements ? 101

2. Statut des consonnes de liaison 102

2.1. Consonnes finales 103

2.1.1. Analyse par troncation 104

2.1.2. Analyse par supplétion : allomorphie 104

2.1.3. Analyses autosegmentales : la notion de flottance 105

2.2. Consonnes épenthétiques 106

2.3. Consonnes initiales 107

2.4. Consonnes morphémiques 107

2.4.1. Analyse préfixale 107

2.4.2. Analyse suffixale 108

2.5. Consonnes intégrées à des constructions 108

2.6. Un statut hétérogène 109

3. Traitements des consonnes de liaison 109

3.1. Les modèles de la phonologie générative linéaire 109

3.1.1. Des consonnes finales latentes : application de la règle de troncation (Schane 1965) 111

3.1.1.1 Schane (1974) 113

3.1.1.2 Schane (1978) 113

3.1.2. Des consonnes finales enchaînées : application de la règle de métathèse (Dell 1970) 114

3.1.2.1 Dell (1973) 115

3.1.3. Des consonnes épenthétiques : application de la règle d'épenthèse 116

3.1.3.1 Klausenburger (1974) 116

3.1.3.2 Tranel (1974) 116

3.1.3.3 Tranel (1976 ; 1981) 116

3.2. Les modèles de la phonologie multilinéaire : des consonnes instables .. 117

3.2.1. Consonnes flottantes 118

3.2.2. Des consonnes flottantes par rapport au squelette 118

3.2.2.1 Wetzels (1987), Tranel (1990) 118

3.2.3. Des consonnes flottantes par rapport à la syllabe 118

3.2.3.1 Clements et Keyser (1983) 118

3.2.4. Des consonnes flottantes par rapport au squelette et à la syllabe 119

3.2.4.1 Encrevé (1988) 119

3.3. Des consonnes non intrinsèques : les modèles basés sur la théorie de l'optimalité 121

3.3.1. Tranel (2000) 122

3.4. Les modèles de la phonologie exemplariste : des consonnes appartenant à des constructions 123

3.5. Les recherches en acquisition L1 125

3.6. Conclusion 129

Chapitre 6. Représentations de la liaison : approches psycholinguistiques 131

1. Modèles perceptifs et productifs : repères théoriques 131

1.1. Modèles de reconnaissance des mots parlés 131

1.1.1. Modèle Cohorte 132
1.1.2. Modèle TRACE 133
1.1.3. Limites 133
1.2. Modèles de reconnaissance des mots écrits 134
1.2.1. Modèle Cohorte 134
1.2.2. Modèle à deux voies en cascades (Dual-Route Cascaded Model, DRC) 135
1.3. Modèles de production orale 135
1.3.1. Levelt 135
1.3.2. Dell 136
1.4. Bilan 136
1.5. Reconnaissance des séquences de liaison 137
1.6. Production orale des séquences de liaison 138
2. Représentations de la liaison dans le lexique mental 142
2.1.1. Hypothèses abstractionnistes 142
2.1.2. Hypothèses exemplaristes 143
3. Représentation de la liaison : entre phonologie et orthographe 143
3.1. Indépendance entre système oral (phonologie) et écrit (orthographe) (orthographic autonomy hypothesis) 144
3.2. Implication inter-tâches entre système oral (phonologie) et écrit (orthographique) (phonological mediation hypothesis) 146
3.2.1. Influence de l'orthographe dans la reconnaissance des mots parlés 147
3.2.2. Influence de l'orthographe dans la production orale 148
3.3. Bilan 149
3.4. Modèle interactif bimodal 150
4. Conclusion 151
Partie 2. Analyse fréquentielle de la liaison dans un corpus de français parlé. 153
Chapitre 1. Le corpus « Phonologie du Français Contemporain » (PFC) : un corpus de français parlé codé en liaison et annoté en morphosyntaxe 155
1. Brève présentation du projet PFC 155

1.1. Objectifs 156
1.2. Points d'enquêtes 156
1.3. Méthodologies 157
1.3.1. Protocole d'enquête 157
1.3.2. Traitement et transcription des données 158
1.3.3. Codage et analyse de la liaison 159
2. Annotations morphosyntaxiques pour l'étude de la liaison 160
2.1. Première annotation morphosyntaxique à l'aide de l'étiqueteur Treetagger (Schmid 1994) 161
2.2. Un travail de ré-annotation morphosyntaxique à l'aide de l'étiqueteur DisMo 162
2.2.1. Jeu d'étiquettes morphosyntaxiques 164
2.3. Conclusion 167
Chapitre 2. Analyse de la distribution de la fréquence de la liaison variable dans la parole semi-spontanée 171
1. Études précédentes 171
1.1. Loi de Zipf 172
1.2. Distribution de Zipf de la fréquence des réalisations totales de la liaison 176
2. Présente étude 177
2.1. Corpus et méthodologies 178
2.2. Résultats 178
2.2.1. Distribution de la fréquence de la liaison variable 178
2.2.2. Idiosyncrasies de la liaison variable 181
2.2.2.1 Conjonctions suivies d'un pronom (CON-PRO) 183
2.2.2.2 Prépositions suivies d'un déterminant (PRP-DET) 183
2.2.2.3 Verbes suivis d'un déterminant (VER-DET) 184
2.2.2.4 Prépositions suivies d'un nom commun et propre (PRP-NOM) 185
2.2.2.5 Adverbes suivis d'un adjectif (ADV-ADJ) 185
2.2.2.6 Verbes suivis d'une préposition (VER-PRP) 186
2.2.2.7 Adjectifs suivis d'un nom (ADJ-NOM) 187

2.2.2.8 Bilan 187

2.2.3. Étude supplémentaire de la liaison après adjectifs épithètes 188

2.2.3.1 Approches théoriques et traitement de la liaison après adjectif épithète 189

2.2.3.2 Objectifs et corpus 190

2.2.3.3 Résultats 190

2.2.3.4 Bilan 195

2.3. Discussion 196

2.3.1. Cohésion et productivité des constructions spécifiques : entre syntaxe, fréquence et probabilité transitionnelle 196

2.3.1.1 Hypothèse exemplariste 197

2.3.1.2 Hypothèse basée sur la probabilité transitionnelle 197

3. Conclusion 204

Partie 3. Enseignement et acquisition de la liaison en français langue seconde et étrangère 207

Chapitre 1. Repères théoriques sur l'acquisition d'une langue seconde et étrangère 209

1. Introduction 209

2. Le système linguistique des apprenants : l'interlangue 210

3. Influence des facteurs internes et externes sur le développement de l'interlangue 213

3.1. Des facteurs internes 213

3.1.1. Le rôle de la L1 213

3.1.1.1 Le transfert dans la phonologie L2 214

3.1.2. La saillance linguistique 215

3.2. Des facteurs externes 216

3.2.1. L'âge 216

3.2.2. Le contexte d'acquisition d'une L2 : langue seconde et langue étrangère 217

3.2.3. Modalité d'acquisition : acquisition guidée et non guidée 219

3.2.4. Quelques caractéristiques de l'input 221
3.2.4.1 L'input écrit 221
3.2.4.2 La fréquence de l'input 222
4. Approches de l'interlangue 223
4.1. Analyse des erreurs 223
4.1.1. Erreurs interlinguistiques : le transfert interlangue 224
4.1.2. Erreurs intralinguistiques 225
4.1.3. Erreurs inter et intralinguistiques 225
4.2. Limites 226
4.3. Itinéraires, profils et stratégies d'acquisition 226
4.3.1. Itinéraires acquisitionnels et profils des apprenants 226
4.3.1.1 Le projet « Heidelberger Forschungsprojekt Pidgin Deutsch » (HPD) (Dittmar et Klein 1979) 227
4.3.1.2 Le programme « Zweitspracherwerb Italienischer und Spanischer Arbeiter » (ZISA) (Meisel, Clahsen et Pienemann 1981) 227
4.3.1.3 Le programme « European Science Foundation » (ESF) 228
4.3.1.4 Programmes d'études sur les itinéraires acquisitionnels et profils grammaticaux du français L2 229
Itinéraire acquisitionnel de l'accord sujet-verbe au pluriel 232
Itinéraire acquisitionnel des pronoms objet 232
4.3.1.5 Théorie de la Processabilité 233
4.3.2. Stratégies de l'apprenant 235
4.3.3. Limites 236
5. Conclusion 237
Chapitre 2. L'acquisition de la liaison en français langue seconde et étrangère 239
1. Etudes précédentes 239
1.1. Traitement cognitif et perception de la liaison 241
1.2. Production de la liaison 243
1.3. Contextes de liaison étudiés 248
1.4. Typologies d'erreurs de liaison 251
2. Études issues du projet InterPhonologie du Français Contemporain (IPFC) 254

2.1. Production de la liaison 256
2.2. Typologies d'erreurs de liaison 257
3. Modèles développementaux d'acquisition de la liaison en L2 261
3.1. Comparaison entre processus d'acquisition en L1 et L2 263
4. Le rôle de la L1 dans l'acquisition de la liaison 265
5. Le rôle de l'orthographe dans l'acquisition de la liaison en L2 267
6. Conclusion 268

Chapitre 3. Normes pédagogiques et matériaux didactiques dans l'enseignement de la liaison en français langue étrangère 271

1. Quelle variété de référence pour l'enseignement de la liaison en FLE? 271
1.1. Français parisien cultivé : une tentative de « mettre de l'ordre » 272
1.2. Recherche d'une nouvelle norme de prononciation : entre nivellement des usages et intégration de la variation sociolinguistique 273
1.3. Acquisition et enseignement de la variation sociolinguistique 274
1.3.1. L'« acceptabilité » sociolinguistique dans les normes pédagogiques 277
2. Matériaux didactiques 279
2.1. Traités de phonétique 280
2.2. Manuels de phonétique 281
2.3. Corpus oraux pour l'enseignement de la liaison 286
3. Conclusion 288

Partie 4. Une étude longitudinale de l'acquisition de la liaison chez des apprenants italophones de français langue étrangère (FLE) : l'enquête IPFC-italien (Rome) 291

Chapitre 1. L'enquête « Interphonologie du Français Contemporain - italien » (IPFC-italien) (Rome) : méthodologies, corpus et traitement des données 293

1. Corpus oraux en L2 294
2. Le projet InterPhonologie du Français Conteporain (IPFC) 295
2.1. Objectifs 295
2.2. Points d'enquête 296

2.3. Protocole 296
2.4. Traitement des données 297
2.4.1. Transcription 297
2.4.1.1 Traitement des formes ambiguës 298
2.4.2. Codage liaison 299
2.4.2.1 Champs 1 : nature de la CL cible 300
2.4.2.2 Champs 2 et Champs 3 : catégories syntaxiques du mot liaisonnant (M1) et du mot qui suit (M2) 301
2.4.2.3 Champ 4 : Nombre de syllabes et nature de la voyelle du mot liaisonnant (M1) 302
2.4.2.4 Champ 5 : Réalisation de la liaison 302
2.4.2.5 Champ 6 : Type de CL réalisée 303
2.4.2.6 Champ 7 : Présence d'une pause / hésitation / coup de glotte ... 303
3. L'équipe IPFC-italien 304
3.1. Points d'enquête 304
4. L'enquête IPFC-italien (Rome) 305
4.1. Objectifs 306
4.1.1. Objectif 1 : Développement de représentations phonologiques de la CL 306
4.1.1.1 Hypothèse 1 306
4.1.2. Objectif 2 : Influence de la L1 dans l'acquisition de la liaison 307
4.1.2.1 Hypothèse 2 307
4.1.3. Objectif 3 : Influence de l'input écrit dans l'acquisition de la liaison 308
4.1.3.1 Hypothèse 3 308
4.1.4. Objectif 4 – Maîtrise et connaissance de la variabilité de la liaison 308
4.1.4.1 Hypothèse 4 309
4.1.5. Objectif 5 : enjeux pédagogique 309
4.2. Sujets 309

4.3. Méthodologies 312
4.3.1. Étude longitudinale 314
4.3.1.1 Répétition des séquences isolées de liaison 314
4.3.1.2 Lecture des séquences isolées de liaison 316
4.3.1.3 Lecture de la liste PFC 317
4.3.1.4 Lecture texte PFC 318
4.3.1.5 Lecture du texte journalistique adapté 320
4.3.1.6 Test de jugement 322
4.3.1.7 Entretien guidé 325
4.3.2. Étude transversale 325
4.3.2.1 Dénomination d'images 326
4.3.2.2 Répétition et lecture d'une liste spécifique de mots (IPFC-italien) 327
4.3.2.3 Tâche de reformulation 327
4.4. Traitement des données 330
4.4.1. Traitement des données de liaison 330
Chapitre 2. Résultats de l'enquête IPFC-italien (Rome) 331
1. Résultats de l'étude longitudinale 332
1.1. Résultats totaux 332
1.2. Réalisation de la liaison 337
1.2.1. Non-réalisations de la liaison 338
1.2.1.1 Contextes catégoriques 340
Après déterminant 340
Avant et après pronoms personnels et clitiques 341
Dans des expressions figées 343
Après les adjectifs numéraux 344
1.2.1.2 Le rôle de la fréquence de cooccurrence 344
1.2.1.3 Contextes variables 345
Après adjectif 347
Après adverbe 347
Après nom 349

Après préposition 350
Après verbe 351
Après verbe auxiliaire 353
Après verbe infinitif 354
1.2.2. Réalisations 354
1.2.2.1 Enchaînement 354
1.2.2.2 Non-enchaînement 355
1.2.2.3 Pauses-hésitations 355
1.2.2.4 Non-compréhensions 355
1.3. Nature de la CL réalisée 356
1.3.1. Liaison conforme à la CL cible 357
1.3.2. CL presque conforme à la cible mais avec différence phonétique 357
1.3.3. CL non conforme à la CL présente dans le M1 360
1.3.4. CL non conforme avec substitution de la CL 361
1.4. Liaisons erratiques 364
1.5. Liaisons non attendues 364
1.6. Résultats du test de jugement 367
1.6.1. Jugements liaisons catégoriques 369
1.6.2. Jugements liaisons variables 370
1.7. Bilan de l'étude longitudinale 371
2. Résultats de l'étude transversale 373
2.1. Types de liaison potentielle 373
2.1.1. Dénomination d'image 374
2.1.1.1 Réalisation de la liaison 374
2.1.1.2 Nature de la CL réalisée 374
2.1.2. Tâche de reformulation 375
2.1.2.1 Réalisation de la liaison 375
2.1.2.2 Nature de la CL réalisée 377
2.2. Bilan de l'étude transversale 380
3. Résultats par type d'oral 380

3.1. Répétition 381
3.1.1. Type de liaison potentielle 381
3.1.2. Réalisation 382
3.1.3. Nature de la CL réalisée 383
3.2. Lecture oralisée 384
3.2.1. Type de liaison potentielle 384
3.2.2. Réalisation 385
3.2.3. Nature de la CL réalisée 389
3.3. Parole semi-spontanée 393
3.3.1. Type de liaison potentielle 393
3.3.2. Réalisation 397
3.3.3. Nature de la CL réalisée 401
3.4. Bilan des résultats par tâche 404
Chapitre 3. Discussion et aspects d'un itinéraire d'acquisition de la liaison en français langue étrangère 407
1. Discussion 407
1.1. Hypothèse 1 407
1.1.1. Non-réalisation des liaisons catégoriques 407
1.1.1.1 La liaison comme marque redondante dans l'acquisition du français langue étrangère (FLE)? 409
1.1.2. Vers des représentations phonologiques de la liaison 411
1.1.3. L'acquisition de la liaison après les pronoms clitiques 413
1.2. Hypothèse 2 414
1.3. Hypothèse 3 415
1.3.1. Interactions entre représentations phonologiques et orthographiques 416
1.4. Hypothèse 4 419
2. Entre stades post-initiaux et stades intermédiaires 420
3. Itinéraire d'acquisition de la liaison catégorique en français langue étrangère (FLE) 421

3.1. Stade initial - procédure par lemme et par catégorie 422
3.2. Stade post-initial - procédure intra-syntagmatique 423
3.3. Stade intermédiaire - procédure inter-syntagmatique 424
3.4. Stades avancés - procédure de phrase subordonnée 424
4. Conclusion 425

Partie 5. Des atouts d'un corpus de natifs dans l'acquisition de la liaison en francais langue etrangère 427

Chapitre 1. Ressources pour l'acquisition de la liaison en français langue étrangère 429
1. Quel traitement de la CL dans l'enseignement du FLE ? 429
1.1. L'étude des correspondances phonographiques 432
2. L'acquisition de la variation de la liaison au-délà des règles classificatoires 433
2.1. Proposition d'une classification multi-niveaux de la liaison 435
2.2. Un concordancier multi-niveaux et multimédia pour l'étude de la liaison 438
2.2.1. Typologies 439
2.2.2. Un concordancier pour une étude multi-niveaux et multimédia de la liaison 440
2.2.2.1 Illustration 442
2.2.2.2 Définition de la requête 442
2.2.2.3 Sélection des attributs du corpus à afficher dans les résultats.... 443
2.2.2.4 Affichage des résultats 445
2.2.2.5 Exemples de requêtes 447
3. Conclusion 448

Conclusions générales 451
Références bibliographiques 459
Index des notions 489
Index des noms 491
Table des illustrations 495
Table des exemples 501

Table des figures 503
Table des matières 507
Annexes 527
Annexe 1 529
Préposition-Nom (PRP-NOM) 529
PRP-NOM 529
Conjonction-Pronom (CON-PRO) 529
CON : sub-PRO 529
Préposition-Déterminant (PRP-DET) 529
PRP-DET 529
Verbe-Déterminant (VER-DET) 529
VER : impf-DET 529
VER : ppas-DET 529
VER : ppre-DET 530
VER : pres-DET 530
VER : subp-DET 530
Verbe-Verbe (VER-VER) 530
VER : cond-VER 530
VER : impf-VER 530
VER : ppas-VER 530
VER : pres-VER 530
VER : pres:aux-VER 530
Verbe - Préposition (VER-PRP) 530
VER : pres-PRP 530
VER : impf- PRP 531
VER : inf-PRP 531
VER : ppas-PRP 531
Adverbe-Adjectif (ADV-ADJ) 531
Adjectif-Nom (ADJ-NOM) 531
Adverbe-Verbe (ADV-VER) 531

Adverbe-Préposition 531
Préposition-Verbe 531
Verbe-Adjectif (VER-ADJ) 532
VER : impf-ADJ 532
VER : pres-ADJ 532
Pronom-Préposition (PRO-PRP) 532
PRO : ind-PRP 532
Préposition-Pronom (PRP-PRO) 532
Verbe-Adverbe (VER-ADV) 532
Verbe-Nom (VER-NOM) 532
Adverbe-Déterminant (ADV-DET) 532
Adverbe-Adverbe (ADV-ADV) 533
Nom-Adjectif (NOM-ADJ) 533
Conjonction-Déterminant (CON-DET) 533
Pronom-Pronom (PRO-PRO) 533
PRO : rel-PRO 533
Adverbe-Nom (CON-NOM) 533
Pronom-Adjectif (PRO-ADJ) 533
Conjonction-Verbe (CON-VER) 534
Adverbe-Pronom (CON-PRO) 534
Nom-Conjonction (NOM-CON) 534
Pronom-Adverbe (PRO-ADV) 534
Conjonction-Adverbe (CON-ADV) 534
Annexe 2 533
Annexe 3 539
Fiche ségnalétique 539
Questionnaire 540
Consentement de participation 544
Consenso per il trattamento dei dati personali 545
Annexe 4 547
Étude longitudinale 547

Tâche de répétition des séquences de liaison isolées 547
Liste des stimuli d'entraînement 547
Liste des stimuli 547
Tâche de lecture de la liste de mots PFC 549
Liste des mots d'entraînement 549
Liste des mots cible selon l'ordre de soumission 549
Tâche de lecture des séquences de liaison 552
Liste des séquences d'entraînement 552
Liste des séquences cibles 552
Tâche lecture texte PFC 554
Le Premier Ministre ira-t-il à Beaulieu ? 554
Tâche lecture texte journalistique 555
Ghettos, sweet ghettos 555
Test de jugement 557
Liste des stimuli 557
Distracteurs 559
Étude transversale 561
Tâche de dénomination d'image 561
Image d'entraînement 561
Liste des images : 561
Répétition et lecture d'une liste spécifique de mots (IPFC-italien) 563
Liste des mots d'entraînement 563
Liste des mots cibles 563
Partie commune à toutes les équipes IPFC 563
Liste spécifique à l'équipe IPFC-italien : 564
Tâche de reformulation 566
Liste des séquences d'entraînement 566
Liste des séquences cibles 566

ANNEXES

Annexe 1

Extraits d'énoncés des types de liaison variable les plus fréquents (n. tokens >100).

Préposition-Nom (PRP-NOM)

PRP-NOM

« […] elle ne venait vers nous on lui parlait *en allemand* mais elle nous répondait en français ». (Corpus PFC_Neuchâtel_scajb1gg)

Conjonction-Pronom (CON-PRO)

CON : sub-PRO

« […] *quand ils* ont hérité cette maison et bien ils ont éd/ décidé de faire des travaux, et de me garder en même temps à l'entretien, puisque même s'ils ne viennent pas tous les mois ». (Corpus PFC_Douzens _11amg2gg)

Préposition-Déterminant (PRP-DET)

PRP-DET

« Maintenant, il se rend *dans un* bar après avoir vendu ses bœufs ». (Corpus PFC_Burkina Faso_bfasb1lg)

Verbe-Déterminant (VER-DET)

VER : impf-DET

« Ensuite on prenait le deuxième cycle aussi trois ans, et la dernière époque *était un* régent ». (Corpus PFC_Nyon_svajb1gg)

VER : ppas-DET

« Ouais elle a *fait un* an dans un euh high school euh aux Etats-Unis ». (Corpus PFC_ Québec ville_cqbpt1gg)

VER : ppre-DET

« N'est-ce pas. Le, oui le, le franglais *étant un* mélange ». (Corpus PFC_Nyon _svarv1lg)

VER : pres-DET

« […] euh, elle va avoir tendance à rire de l'Ancien Testament bon elle qui *est une* scientifique ». (Corpus PFC_Québec ville_ cqags1gg)

VER : subp-DET

« Faudrait que les gens *fassent un* effort ». (Corpus PFC_RCA Bangui_ rcadt1gg_DT1)

Verbe-Verbe (VER-VER)

VER : cond-VER

« Mais le graphisme me faisait penser que ça *pourrait être* plus ». (Corpus PFC_Nyon_ svarv1lg)

VER : impf-VER

« […] et elle *allait étudier* ». (Corpus PFC_Douzens_ 11aml1lg)

VER : ppas-VER

« Elle s'est *fait insulter* ». (Corpus PFC_Québec_cqbgt1lg)

VER : pres-VER

« Elle *doit être* de mai parce que je l'ai rencontrée ». (Corpus PFC_Nyon_svaab1lg)

VER : pres:aux-VER

« Donc ma soeur *est arrivée* deux années ». (Corpus PFC_Pouteaux-Courbevoie_92aaf2gg)

Verbe - Préposition (VER-PRP)

VER : pres-PRP

« Aiguevives qui *est à* neuf kilomètres d'ici ». (Corpus PFC_ Douzens_11aal1gg)

VER : impf- PRP

« […] ils étaient att/ euh, ils *étaient en* vélo ». (Corpus PFC_ Douzens_11ajp1gg)

VER : inf-PRP

« Ça me. Ça me disait euh enfin j'avais envie de revenir un peu *m'installer à* Paris ». (Corpus PFC_Paris Centre Ville_75ccm1gg)

VER : ppas-PRP

« elle s'était *mis en* tête euh d'être euh égyptologue ». (Corpus PFC_Marseille Centre Ville_13aag1gg)

Adverbe-Adjectif (ADV-ADJ)

« Elle me dit Anne, "c'est très, *très intéressant* ». (Corpus PFC_ Douzens_11aml1lg)

Adjectif-Nom (ADJ-NOM)

« Bon ben maintenant bah je vaque à mes *petites occupations* ». (Corpus PFC_ Brunoy_ 91asl1lg)

Adverbe-Verbe (ADV-VER)

« Mais euh, après quelques minutes ça, finirait par *bien aller*, je pense ». (Corpus PFC_ Québec_ cqajr1gg)

Adverbe-Préposition

« […] je pense qu'au niveau de la presse ils préféreraient ne *pas en* parler maintenant parce qu'ils ont pas vraiment la solution ». (Corpus PFC_ Neuchâtel _ scapy1lg)

Préposition-Verbe

« Enfin il était. *Sans avoir* jamais été à l'école, il était, je peux dire c'est pas parce que c'est mon papa, mais il était très intelligent ». (Corpus PFC_ Rodez_ 12atp1gg)

Verbe-Adjectif (VER-ADJ)

VER : impf-ADJ

« Sa mère mais autrement mon grand-père *était italien* ». (Corpus PFC_ Nyon_ svarb2gg)

VER : pres-ADJ

« Oui, oui, oui, oui il était d'ici enfin euh de Lecumberry oui, oui et maman *est originaire* d'ici, du (XX) ouais, voilà ». (Corpus PFC_ Biarritz_64ama1gg)

Pronom-Préposition (PRO-PRP)

PRO : ind-PRP

« Vas demander ta copie, maintenant tu as plus *rien à* perdre de toute façon ». (Corpus PFC_ Domfrontais_61ahl1lg)

Préposition-Pronom (PRP-PRO)

« *Chez eux*, c'est relativement plat ». (Corpus PFC_ Brunoy_91acs2lg)

Verbe-Adverbe (VER-ADV)

« Ben, les enfants euh, oui, ils *sont encore* à la maison, disons ils sont domiciliés à la maison, c'est sûr ». (Corpus PFC_ Vendée_85amm1lg)

Verbe-Nom (VER-NOM)

« On, on l'avait gardé euh, d'ailleurs il *est ingénieur* maintenant ». (Corpus PFC_ Roanne_42ayd1lg)

Adverbe-Déterminant (ADV-DET)

« C'est *pas un* apprentissage total ». (Corpus PFC_ Chlef_acaana1gg)

Adverbe-Adverbe (ADV-ADV)

« […] comme l'Université n'a pas, l'année n'a *pas encore* écoulée, il attend ». (Corpus PFC_ RCA Bangui_rcarn1gg)

Nom-Adjectif (NOM-ADJ)

« Hum, dans le village, les, les, il y a que des *personnes âgées*, elles parlent toutes patois ». (Corpus PFC_ Toulouse_31aec1gg)

Conjonction-Déterminant (CON-DET)

« [..] c'est pas les mêmes tarifs *quand une* personne elle prend son euh, son enfant à la crèche ». (Corpus PFC_ Chlef_acazb1lgg)

Pronom-Pronom (PRO-PRO)

PRO : rel-PRO

« […] la façon dont ils, *dont ils* conçoivent, l'enseignement, en général ». (Corpus PFC_Chlef_acaab1lg)

Adverbe-Nom (CON-NOM)

« Pas *très envie*. J'ai pas envie non j'en ai vraiment marre ». (Corpus PFC_ParisCentre Ville_75ccr1lg)

Pronom-Adjectif (PRO-ADJ)

« Moins souvent dans le Courrier de l'Escaut que que *certains autres* ». (Corpus PFC_Tournai_btapm1gg)

Conjonction-Verbe (CON-VER)

« […] puis *quand arrivait* le printemps et, et je me souviens quand nous étions jeunes parce que moi je suis l'aînée de cinq, si nous avions des devoirs à, à faire à la maison ou alors, enfin (XX) devoir à faire si on demandait à mon père qu'on craignait un peu ». (Corpus PFC_Lacaune_81arc1gg)

Adverbe-Pronom (CON-PRO)

« Ça dépend, tout de *comment on* prend les, comment on prend les choses, hein? ». (Corpus PFC_Nyon_svaab1gg)

Nom-Conjonction (NOM-CON)

« Et mes parents sont nés en territoire, euh, algérien-français quoi. Et, euh, et nous aussi, tous mes *frères et* mes sœurs ». (Corpus PFC_Rodez_12ags1gg)

Pronom-Adverbe (PRO-ADV)

« […] une fois par an on allait une fois par an au restaurant *tous ensemble*, mais euh c'était des sandwichs et il fallait pas que ce soit trop lourd non plus pour euh ». (Corpus PFC_Gembloux_bgamm1gg)

Conjonction-Adverbe (CON-ADV)

« Du mieux possible vous allez me dire d'accord *mais encore* ». (Corpus PFC_Marseille Centre Ville_13aas1lg).

Annexe 2

Le tableau ci-dessous résume les résultats issus de l'analyse fréquentielle de la liaison variable (types par catégorie morphosyntaxique générale ex. NOM ; ADV etc.). En colonne, respectivement : (a) le nombre total d'occurrences de liaisons non-réalisées (dorénavant LNR) ; (b) le nombre total d'occurrences de laisons réalisées (dorénavant LR) ; (c) le nombre total de liaisons potentielles (dorénavant TOT) ; (d) le taux de réalisation de la liaison (dorénavant %LR).

	LNR	**LR**	**TOT**	**%LR**
Total général	18272	4296	22568	19%
PRP-NOM	35	653	688	95%
CON-PRO	1076	637	1713	37%
PRP-DET	201	553	754	73%
VER-DET	3160	486	3646	13%
VER-VER	1974	411	2385	17%
VER-PRP	3589	282	3871	7%
ADV-ADJ	267	276	543	51%
ADJ-NOM	46	211	257	82%
ADV-VER	791	104	895	12%
ADV-PRP	848	92	940	10%
PRP-VER	15	91	106	86%
VER-ADJ	416	90	506	18%
PRO-PRP	33	68	101	67%
PRP-PRO	85	56	141	40%
VER-ADV	1065	54	1119	5%
VER-NOM	374	50	424	12%
ADV-DET	424	37	461	8%
ADV-ADV	427	36	463	8%
NOM-ADJ	327	33	360	9%
CON-DET	100	21	121	17%
PRO-PRO	31	16	47	34%
ADV-NOM	151	8	159	5%
PRO-ADJ	2	7	9	78%
CON-VER	39	6	45	13%
ADV-PRO	819	5	824	1%
NOM-CON	599	5	604	1%
PRO-ADV	17	5	22	23%
CON-ADV	449	3	452	1%

Le tableau ci-dessous résume les résultats détaillés issus de l'analyse fréquentielle de la liaison variable (types par catégorie morphosyntaxique générale ex. NOM : com ; ADV : neg etc.). En colonne, respectivement : (a) le nombre total d'occurrences de liaisons non-réalisées (dorénavant LNR) ; (b) le nombre total d'occurrences de laisons réalisées (dorénavant LR) ; (c) le nombre total de liaisons potentielles (dorénavant TOT) ; (d) le taux de réalisation de la liaison (dorénavant %LR).

	LNR	**LR**	**TOT**	**%LR**
Total général	18272	4296	22568	19%
CON:sub-PRO:per:sjt	205	627	832	75%
PRP-DET:ind	159	553	712	78%
VER:pres-DET:ind	1476	435	1911	23%
PRP-NOM:com	10	386	396	97%
VER:pres:aux-VER:ppas	735	336	1071	31%
PRP-NOM:pro	23	263	286	92%
ADV:deg-ADJ	46	251	297	85%
VER:pres-PRP	1152	220	1372	16%
ADJ-NOM:com	46	209	255	82%
ADV-VER:ppas	174	74	248	30%
VER:pres-ADJ	254	71	325	22%
PRO:ind-PRP	30	63	93	68%
ADV-PRP	390	59	449	13%
VER:pres-ADV	522	50	572	9%
PRP-PRO:per:ton	6	49	55	89%
PRP-VER:ppre	0	44	44	100%
VER:impf-DET:ind	976	34	1010	3%
NOM:com-ADJ	325	33	358	9%
VER:pres-VER:inf	453	31	484	6%
ADV-DET:ind	171	29	200	15%
VER:pres-NOM:pro	61	27	88	31%
PRP-VER:inf	5	25	30	83%
ADV-ADJ	115	24	139	17%
VER:impf-PRP	922	21	943	2%
ADV:deg-PRP	50	20	70	29%
ADV-ADV	196	18	214	8%
VER:pres-PRP:det	179	18	197	9%
VER:impf:aux-VER:ppas	331	17	348	5%
CON:sub-DET:def	0	15	15	100%
VER:pres-NOM:com	82	15	97	15%
ADV:neg-ADV	193	14	207	7%
VER:pres-VER:ppas	29	14	43	33%
VER:ppas-PRP	341	13	354	4%
ADV:deg-VER:ppas	34	12	46	26%
PRO:rel-PRO:per:sjt	8	12	20	60%
PRP-VER:pres	0	12	12	100%
VER:impf-ADJ	110	12	122	10%

VER:ppas-DET:ind	285	12	297	4%
ADV:neg-PRP	284	8	292	3%
ADV:neg-DET:ind	222	7	229	3%
PRO:ind-ADJ	2	7	9	78%
PRP-PRO:per:sjt	76	7	83	8%
CON:sub-PRO:dem	0	6	6	100%
ADV:deg-NOM:com	12	5	17	29%
PRO:ind-PRP:det	2	5	7	71%
VER:pres:aux-ADJ	15	5	20	25%
ADV:deg-ADV	21	4	25	16%
ADV-VER:pres:aux	3	4	7	57%
CON:sub-DET:ind	10	4	14	29%
NOM:com-CON:coo	549	4	553	1%
PRO:ind-ADV	9	4	13	31%
PRP-VER:pres:aux	1	4	5	80%
VER:impf-NOM:com	104	4	108	4%
ADV:deg-PRP:det	10	3	13	23%
ADV:neg-VER:ppas	299	3	302	1%
ADV-VER:inf	44	3	47	6%
CON:coo-ADV	444	3	447	1%
CON:sub-VER:pres	0	3	3	100%
PRP-NOM:com:aux	0	3	3	100%
VER:impf-ADV	295	3	298	1%
VER:impf-PRP:det	181	3	184	2%
VER:ppas-VER:inf	18	3	21	14%
VER:pres:aux-PRP	4	3	7	43%
ADJ-NOM:pro	0	2	2	100%
ADV:deg-PRO:per:sjt	22	2	24	8%
ADV:int-PRO:per:sjt	65	2	67	3%
ADV:int-VER:pres	6	2	8	25%
ADV:neg-NOM:com	51	2	53	4%
ADV-PRP:det	67	2	69	3%
ADV-VER:pres	7	2	9	22%
CON:sub-DET:pos	0	2	2	100%
CON:sub-PRO:per:ton	0	2	2	100%
CON:sub-VER:pres:aux	1	2	3	67%
PRO:ref-PRO:per:obji	0	2	2	100%
PRO:rel-PRO:ind	1	2	3	67%
PRP-VER:impf	0	2	2	100%
PRP-VER:inf:aux	9	2	11	18%
VER:impe-PRP	23	2	25	8%
VER:impf-VER:inf	163	2	165	1%
VER:inf-DET:ind	283	2	285	1%
VER:pres:aux-VER:ppas:aux	35	2	37	5%
VER:pres-VER:inf:aux	10	2	12	17%
ADV:deg-DET:ind	17	1	18	6%
ADV:deg-VER:pres	1	1	2	50%
ADV:neg-ADJ	106	1	107	1%

ADV:neg-VER:inf	145	1	146	1%
ADV:neg-VER:pres	0	1	1	100%
ADV-NOM:pro	27	1	28	4%
ADV-PRO:per:sjt	579	1	580	0%
ADV-VER:impf	3	1	4	25%
CON:coo-PRO:per:sjt	857	1	858	0%
CON:sub-PRO:per:obji	0	1	1	100%
CON:sub-VER:impf	0	1	1	100%
NOM:pro-CON:coo	35	1	36	3%
PRO:dem-ADV	6	1	7	14%
PRP-NOM:acr	2	1	3	33%
PRP-VER:impf:aux	0	1	1	100%
PRP-VER:ppre:aux	0	1	1	100%
VER:cond-VER:inf	41	1	42	2%
VER:impe:aux-VER:ppas	2	1	3	33%
VER:impf-DET:pos	0	1	1	100%
VER:inf-PRP	552	1	553	0%
VER:ppas-ADJ	3	1	4	25%
VER:ppas-ADV	79	1	80	1%
VER:ppas-NOM:com	18	1	19	5%
VER:ppas-NOM:pro	11	1	12	8%
VER:ppre-ADJ	1	1	2	50%
VER:ppre-DET:ind	7	1	8	13%
VER:ppre-PRP	9	1	10	10%
VER:pres:aux-NOM:com	5	1	6	17%
VER:pres:aux-VER:inf	2	1	3	33%
VER:pres-VER:pres	1	1	2	50%
VER:subi-NOM:pro	0	1	1	100%
VER:subp-DET:ind	26	1	27	4%

Annexe 3

Documents enquête IPFC-italien (Rome)

Fiche ségnalétique

NOM, prénom(s) :

Date de naissance :

Lieu de naissance :

Domiciles successifs (en nombre d'années)

Domicile actuel :

Études (préciser jusqu'à quel âge et quel type d'études) :

Langues parlées :

Père de l'informateur, année de naissance :

Lieu d'origine :

Profession

Études :

Langues parlées (étrangères ou régionales) :

Mère de l'informateur, année de naissance :

Lieu d'origine :

Profession :

Études :

Langues parlées (étrangères ou régionales) :

Personnes ayant joué un rôle important au moment de l'apprentissage du français:

Activités culturelles, loisirs, voyages :

Questionnaire[330]

Nom - Prénom : Sexe : Féminin □

Date de naissance : Masculin □

1. Combien de langues parle-t-on dans votre famille ?

1 □ 2 □ plus de 2 □

Si deux ou plus, quelle(s) langue(s) parle-t-on ?

2. Quelle était votre première langue vivante au collège ?

3. Quelles autres langues avez-vous étudiées au collège ?

4. Quelles langues avez-vous étudiées au lycée ?

5. Quelles langues étudiez-vous à l'université ?

6. Combien de cours concernant la langue et la culture française avez-vous choisies ? (nombre d'heures par semaine au total)

- 1ère année de Licence (Triennale):
- 2ème année de Licence (Triennale):
- 3ème année de Licence (Triennale) :

7. Quelle note avez-vous obtenue au cours de lectorat français ?

- 1ère année de Licence (Triennale):
- 2ème année de Licence (Triennale):
- 3ème année de Licence (Triennale) :

8. Avez-vous effectué des séjours en France, ou dans un pays où l'on parlait français ?

[330] Nous remercions Mylène Harnois-Delpiano de nous avoir permis d'utiliser ce modèle de questionnaire.

Oui □ Non □

Si oui, à quel âge et combien de temps ?

9. Avez-vous pris des cours à l'Alliance Française ?

Oui □ Non □

Si oui, précisez le nom de chaque cours et sa durée (combien d'heures par semaine / combien de semaines).

10. Avez-vous pris des cours particuliers de français ?

Oui □ Non □

Si oui, précisez le motif (apprentissage du français, préparation à un examen…), la date et la durée.

11. Avez-vous effectué des stages de français ?

Oui □ Non □

Si oui, précisez le lieu et la durée du (des) stage(s).

12. Parlez-vous français régulièrement en dehors de l'université ?

Oui □ Non □

Si oui, dans quel cadre ?

Amis □

Famille □

Club / Associations □

Autre □

Si autre, précisez :

13. Avez-vous passé un examen français (DELF / DALF / TEF / TCF) ?

Oui □ Non □

Si oui, qu'avez-vous réussi ?

14. Comment estimez-vous votre niveau de français ?

□ Très mauvais

□ Mauvais

□ Moyen

□ Bon

□ Très bon

15. Quand vous allez sur Internet, est-ce qu'il vous arrive de communiquer en français ?

Non, jamais □

Oui, parfois □

Oui, souvent □

Si oui, avec qui ?

Amis □

Famille □

Forum □

Autre □

Si autre, précisez :

16. Ecoutez-vous de la musique dont les paroles sont en français ?

Non, jamais □

Oui, parfois □

Oui, souvent □

Si oui :

A/ Quels styles de musique et quels artistes ?

B/ Essayez-vous de comprendre les paroles pendant l'écoute ?

Oui □ Non □

C/ Portez-vous de l'intérêt aux paroles après l'écoute ?

Oui □ Non □

Si oui, de quelle manière?

Recherche du vocabulaire dans le dictionnaire / Traduction des paroles □

Demande à des amis □

Autre □

Si autre, précisez :

17. Lisez-vous des textes ou des livres en français (en dehors de ceux exigés par vos enseignants) ?

Non, jamais □

Oui, parfois □

Oui, souvent □

Si oui, quels types de textes ?

Romans □

Bandes dessinées □

Journaux / Magazines □

Présentations de produits français □

Paroles de chansons □

Textes trouvés sur Internet □

Autre □

Si autre, précisez :

Si oui, précisez la quantité annuelle :

18. Avez-vous d'autres contacts avec la langue française en dehors de l'université ?

Films en version originale française □

Chaînes TV françaises ou canadiennes □

Radios françaises ou canadiennes □

Autre □

Si autre, précisez :

MERCI BEAUCOUP

Consentement de participation

NOM :

Prénom :

L'enseignant Oreste FLOQUET de l'Università Sapienza et la doctorante Giulia BARRECA de l'Université Paris Ouest Nanterre La Défense et de l'Università Cattolica del Sacro Cuore vous proposent de participer à une enquête dans le cadre du projet universitaire : « InterPhonologie du Français Contemporain » (IPFC). Ce projet s'appuie sur des enregistrements effectués dans le monde entier auprès d'apprenants de français.

Ces enregistrements sont réalisés dans un but de recherche scientifique et d'applications pédagogiques. Les résultats peuvent en être publiés aussi bien dans des revues scientifiques que dans des ouvrages commercialisés (par exemple, manuels pédagogiques) et paraître sur l'Internet. Ils pourront éventuellement être intégrés à des systèmes de production et de perception de la parole.

Si vous acceptez de faire partie de l'échantillon d'études dans les conditions énoncées ci-dessus, veuillez signer ce document.

DATE et SIGNATURE

Consenso per il trattamento dei dati personali

COGNOME :

Nome:

Il docente Oreste FLOQUET dell'Università Sapienza e la dottoranda Giulia BARRECA dell'Université Paris Ouest Nanterre La Défense e dell'Università Cattolica del Sacro Cuore propongono di prendere parte a un'inchiesta linguistica nell'ambito del progetto universitario : «I nterPhonologie du Français Contemporain » (IPFC) (« Interfonologia del francese contemporaneo ») Tale progetto di ricerca prevede la realizzazione di registrazioni di apprendenti di francese lingua straniera, provenienti da diversi paesi del mondo.

I dati raccolti saranno utilizzati a scopi scientifici e didattici. I risultati potranno essere oggetto di pubblicazioni in riviste e opere di carattere scientifico e commerciale (per esempio, manuali didattici) . Inoltre, tali dati potranno apparire su Internet ed essere integrati a dei sistemi di sintesi e percezione della parola.

Vogliate firmare il presente documento per accettare le suddette condizioni di partecipazione.

DATA e FIRMA

Annexe 4

Protocole enquête IPFC-italien (Rome)

Étude longitudinale

Tâche de répétition des séquences de liaison isolées

Liste des stimuli d'entraînement

Type	Tokens
ADJ-NOM	Une grande femme ;
DET-NOM	Des bébés
PRP-NAM	En Grèce

Liste des stimuli

	Type	Tokens
Catégoriques	DET-NOM	Des amis ;
		Des ânons
		Des événements
		Des envoûtements
	PRO -VER	Vous êtes
		Vous encrez
	VER-PRO	Allons-y
		Donnez-en
	EXF-EXF	De plus en plus
Variables	ADJ-NOM	Un petit/grand[331] ami
		Un petit/grand ânon
		Un petit/grand événement
		Un petit/grand envoûtement
	AUX-PPA	Ils ont appris
		Ils ont acquis
	NOM-ADJ	Des efforts inutiles
		Des essais intéressants
	VER-ADV	Nous sommes ici
		Il sort aussi
	CON-PRO	Quand il neige

[331] « Petit » dans la première et deuxième séance (S1, S2) et « grand » dans la troisième et quatrième séance (S3, S4).

	PRP-NAM	En Inde
	PRP-NAM	En Aude
Erratique	EXF-EXF	Au fur et // à mesure
Distracteurs	Un petit/grand dessin	
	Des teintes	
	Des tantes	
	La tonte	
	Il enfonce	
	L'enfance	
	L'étain pur	
	De petites/grandes pinces	
	Il pense	
	Il dépense	
	Des linges	
	Des langes	
	Il longe	
	Il a décompté	
	Il a décanté	

Tâche de lecture de la liste de mots PFC

Liste des mots d'entraînement

- tasse
- moto
- caisse

Liste des mots cible selon l'ordre de soumission

1. roc ;
2. rat
3. jeune
4. mal
5. ras
6. fou à lier
7. des jeunets
8. intact
9. nous prendrions
10. fêtard
11. nièce
12. pâte
13. piquet
14. épée
15. compagnie
16. fête
17. islamique
18. agneau
19. pêcheur
20. médecin
21. paume
22. infect
23. dégeler
24. bêtement
25. épier
26. millionnaire
27. brun

28. scier
29. fêter
30. mouette
31. déjeuner
32. ex-femme
33. liège
34. baignoire
35. pécheur
36. socialisme
37. relier
38. aspect
39. niais
40. épais
41. des genêts
42. blond
43. creux
44. reliure
45. piqué
46. malle
47. gnôle
48. bouleverser
49. million
50. explosion
51. influence
52. mâle
53. ex-mari
54. pomme
55. étrier
56. chemise
57. brin
58. lierre
59. blanc
60. petit
61. jeûne

62. rhinocéros
63. miette
64. slip
65. compagne
66. peuple
67. rauque
68. cinquième
69. nier
70. extraordinaire
71. meurtre
72. vous prendriez
73. botté
74. patte
75. étriller
76. faites
77. feutre
78. quatrième
79. muette
80. piquais
81. trouer
82. piquer
83. creuse
84. beauté
85. patte
86. pâte
87. épais
88. épée
89. jeune
90. jeûne
91. beauté
92. botté
93. brun
94. brin

Tâche de lecture des séquences de liaison

Liste des séquences d'entraînement

Type	Tokens
ADJ-NOM	Une grande femme ;
DET-NOM	Des bébés
PRP-NAM	En Grèce

Liste des séquences cibles

	Type	Tokens
Catégoriques	DET-NOM	Des arbres
		Des âtres
		Des appartements
		Des appontements
	PRO -VER	Nous avons
		Nous amarrons
	VER-PRO	Allez-y
		Prenez-en
	EXF-EXF	De temps en temps
Variables	ADJ-NOM	Un petit/grand[332] arbre
		Un petit/grand âtre
		Un petit/grand appartement
		Un petit/grand appontement
	AUX-PPA	Ils ont acheté
		Ils ont achevé
	NOM-ADJ	Des enfants intelligents
		Des conseils idiots
	VER-ADV	Ils sont ici
		Il vient aussi
	CON-PRO	Quand il pleut
	PRP-NAM	En Inde
	PRP-NAM	En Aude
Erratique	EXF-EXF	Au fur et // à mesure
Distracteurs		Un petit/grand dessin
		Des teintes

[332] « Petit » dans la première et deuxième séances (S1, S2) et « grand » dans la troisième et quatrième séance (S3, S4).

	Des tantes
	La tonte
	Il enfonce
	L'enfance
	L'étain pur
	De petites/grandes pinces
	Il pense
	Il dépense
	Des linges
	Des langes
	Il longe
	Il a décompté
	Il a décanté
	Elle vient demain
	Quand vous arrivez
	Ces voitures
	Sans déclaration
	Premier novembre
	Un grand cheval

Tâche lecture texte PFC

Le Premier Ministre ira-t-il à Beaulieu ?

Le village de Beaulieu est en grand émoi. Le Premier Ministre a en effet décidé de faire étape dans cette commune au cours de sa tournée de la région en fin d'année. Jusqu'ici les seuls titres de gloire de Beaulieu étaient son vin blanc sec, ses chemises en soie, un champion local de course à pied (Louis Garret), quatrième aux jeux olympiques de Berlin en 1936, et plus récemment, son usine de pâtes italiennes. Qu'est-ce qui a donc valu à Beaulieu ce grand honneur ? Le hasard, tout bêtement, car le Premier Ministre, lassé des circuits habituels qui tournaient toujours autour des mêmes villes, veut découvrir ce qu'il appelle "la campagne profonde".

Le maire de Beaulieu - Marc Blanc - est en revanche très inquiet. La cote du Premier Ministre ne cesse de baisser depuis les élections. Comment, en plus, éviter les manifestations qui ont eu tendance à se multiplier lors des visites officielles ? La côte escarpée du Mont Saint-Pierre qui mène au village connaît des barrages chaque fois que les opposants de tous les bords manifestent leur colère. D'un autre côté, à chaque voyage du Premier Ministre, le gouvernement prend contact avec la préfecture la plus proche et s'assure que tout est fait pour le protéger. Or, un gros détachement de police, comme on en a vu à Jonquière, et des vérifications d'identité risquent de provoquer une explosion. Un jeune membre de l'opposition aurait déclaré : "Dans le coin, on est jaloux de notre liberté. S'il faut montrer patte blanche pour circuler, nous ne répondons pas de la réaction des gens du pays. Nous avons le soutien du village entier." De plus, quelques articles parus dans La Dépêche du Centre, L'Express, Ouest Liberté et Le Nouvel Observateur indiqueraient que des activistes des communes voisines préparent une journée chaude au Premier Ministre. Quelques fanatiques auraient même entamé un jeûne prolongé dans l'église de Saint Martinville.

Le sympathique maire de Beaulieu ne sait plus à quel saint se vouer. Il a le sentiment de se trouver dans une impasse stupide. Il s'est, en désespoir de cause, décidé à écrire au Premier Ministre pour vérifier si son village était vraiment une étape nécessaire dans la tournée prévue. Beaulieu préfère être inconnue et tranquille plutôt que de se trouver au centre d'une bataille politique dont, par la télévision, seraient témoins des millions d'électeurs.

Tâche lecture texte journalistique

Ghettos, sweet ghettos

Disneyland est à dix kilomètres, mais on s'y croit. Bienvenue à Bussy-Saint-Georges, en Seine-et Marne, un monde enchanté où toutes les rues se ressemblent.

Les deux cents acquéreurs qui habitent ici vivent retranchés derrière des caméras de vidéosurveillance. Les résidences sécurisées comme celle-là, fleurissent en France.

Un tiers des constructions sont aujourd'hui sécurisées. On sonne. Le joli Interphone doré est doté d'une caméra incorporée. La porte s'ouvre... sur un autre parlophone. Le sas d'observation. Souriez, vous êtes encore filmé. Deux étages plus haut, un couple de résidents allume son téléviseur, chaîne neuf. Les images sont en noir et blanc, mais la vue est imprenable sur le hall d'entrée. La vraie télé-réalité. « Cela décourage les intrus, confie Eric, trente-huit ans, mais ce n'est pas un bunker ». Il est impossible d'y accéder sans autorisation. Les appartements sont équipés d'une alarme, et la présence rassurante d'un régisseur complète le dispositif. Au centre de la résidence trône la piscine, elle aussi grillagée. Il faut une clef pour y accéder. Mais pas de caméra en vue autour du bassin. « Au départ, nous en installions, mais nous avons retiré les appareils par crainte du voyeurisme », explique Robert Monné, du groupe Monné-Decroix, l'un des deux exposants du Salon immobilier de Paris, qui se tient à l'Espace Champerret du vingt-neuf septembre au premier octobre. Il est le premier à avoir introduit le concept en France, en mille neuf cent quatre-vingt-onze. Un ghetto de riches paranoïaques? Non. Ces petites forteresses s'adressent majoritairement aux classes moyennes. « Ce sont-elles qui subissent le sentiment d'insécurité, observe le psychiatre Serge Hefez. Les plus aisés, eux, vivent déjà dans des lieux privilégiés et protégés ». À Bussy, un trois-pièces se loue moins de huit cents euros par mois. « Nous voulons offrir le confort au plus grand nombre, affirme Robert Monné. La sécurité en fait partie ».

Une stratégie payante pour le groupe, qui a construit deux cent cinquante résidences dans l'Hexagone et dont le chiffre d'affaires a été multiplié par cinq en dix ans.

Les promoteurs ont aussi dû convaincre les collectivités locales. « Quand on a commencé à construire, elles ne voulaient pas de notre produit», reconnaît Robert Monné. Avant de retourner leur veste à partir de deux mille deux. Parce que les élus n'hésitent plus à parler de sécurité. Parce qu'ils y ont vu leurs intérêts économiques, aussi ». « Les résidences sécurisées sont synonymes de réduction des coûts de gestion de l'espace public », explique la sociologue Marie-Christine Jaillet, directrice de recherche au CNRS. « La prise en charge des espaces verts, des parkings et de la voirie est privatisée, puisqu'elle est l'affaire de la copropriété ».

Pour les deux cents habitants qui vivent ici, ces enclaves résidentielles sont surtout synonymes de cocons ultraprotégés. « La sécurité est une des préoccupations majeures de nos clients », soutient Pierre Auberger, directeur marketing chez Bouygues Immobilier-

Dans son grand appartement de Bussy-Saint-Georges, Marjorie, vingt-cinq ans, mère d'un petit enfant, ne semble pas s'inquiéter outre mesure. Elle n'allume jamais sa télé pour vérifier qui sonne à son interphone, et utilise de temps en temps son alarme. « Ici, dans ce petit abri on se sent à l'écart des problèmes », dit-elle. « Lors des émeutes urbaines, en novembre deux mille cinq, nous n'avons eu aucun souci ». Pourtant, ces arsenaux sécuritaires de la résidence ne résistent pas longtemps aux petits malins qui veulent à tout prix y entrer. La piscine aurait même tendance à attirer les jeunes des alentours. « Ils arrivent du RER et sautent par-dessus le grand enclos, ils abiment des voitures », soupire Eric, qui a eu la mauvaise idée de les chasser. « Quatre voitures de la résidence ont été rayées ».

« Ghettos, sweet ghettos », publié le 28/09/2006 par Solenne Durox et Anne Vidalie dans le journal « l'Express, et ensuite adapté.

.

Test de jugement

Liste des stimuli

1. Ces arsenaux sécuritaires de la résidence ne résistent pas.

   - [sesaʀsəno]
   - [sezaʀsəno]
   - [seaʀsəno]

2. Deux exposants du Salon immobilier de Paris.

   - [døɛkspozɑ̃]
   - [døzɛkspozɑ̃]
   - [døsɛkspozɑ̃]

3. Le chiffre d'affaires a été multiplié par 5 en dix ans.

   - [diɑ̃]
   - [dizɑ̃]
   - [disɑ̃]

4. Ils abiment des voitures.

   - [ilsabim]
   - [ilzabim]
   - [ilabim]

5. Elle utilise de temps en temps son alarme.

   - [dətɑ̃ɑ̃tɑ̃]
   - [dətɑ̃sɑ̃tɑ̃]
   - [dətɑ̃zɑ̃tɑ̃]

6. On se sent à l'écart des problèmes, dit-elle.

   - [ditɛl]
   - [diɛl]
   - [dizɛl]

7. Dans ce petit abri on se sent à l'écart des problèmes.

   - [pətitabʀi]
   - [pətabʀi]
   - [pətiabʀi]

8. Ils arrivent du RER et sautent par-dessus le grand enclos.

   - [ləgʀɑ̃.ɑ̃klõ]
   - [ləgʀɑ̃tɑ̃klo]
   - [ləgʀɑ̃dɑ̃klo]

9. Quand on a commencé à construire.

   - [kɑ̃dɔ̃]
   - [kɑ̃tɔ̃]
   - [kɑ̃ɔ̃]

10. Ils ont aussi dû convaincre les collectivités locales.

    - [ilzɔ̃osi]
    - [ilzɔnosi]
    - [ilzɔ̃tosi]

11. Ces enclaves résidentielles sont surtout synonymes de cocons ultraprotégés.

    - [dəkɔkɔ̃zyltraprɔteʒe]
    - [dəkɔkɔ̃syltraprɔteʒe]
    - [dəkɔkɔ̃yltraprɔteʒe]

12. La vue est imprenable sur le hall d'entrée.

    - [ɛtɛ̃pʀənabl]
    - [ɛɛ̃pʀənabl]
    - [ɛsɛ̃pʀənabl]

13. Les résidences sécurisées comme celle-là, fleurissent en France.

    - [flœʀisɑ̃tɑ̃]
    - [flœʀisɑ̃.ɑ̃]
    - [flœʀisɑ̃n.ɑ̃]

Distracteurs

14. La porte s'ouvre.

- [syvʀ]
- [suvʀ]
- [zyvʀ]

15. Au centre de la résidence trône la piscine.

- [pisin]
- [piʃin]
- [pisɛ̃]

16. Il est le premier à avoir introduit le concept en France.

- [kɔ̃sɛp]
- [kɔ̃sɛpt]
- [kɔ̃sɛt]

17. Les résidences sécurisées sont synonymes de réduction des coûts de gestion de l'espace public.

- [kut]
- [ky]
- [ku]

18. La résidence ne résiste pas longtemps aux petits malins qui veulent à tout prix y entrer.

- [pətimalɑ̃]
- [pətimalɛ̃]
- [pətimalɔ̃]

19. Quatre voitures de la résidence ont été rayées.

- [katʀəvwatyʀ]
- [katʀevwatyʀ]
- [katvwatyʀ]

20. La sécurité est une des préoccupations majeures de nos clients.

- [maʒœʀ]
- [maʒoʀ]
- [maʒøʀ]

21. Un couple de résidents allume son téléviseur.

- [œ̃kypl]

- [œ̃kupl]
- [œ̃kɔpl]

22. Cela <u>décourage</u> les intrus.

- [dekuʀaʒ]
- [dekyʀaʒ]
- [dəkuʀaʒ]

23. Pierre Auberger, <u>directeur</u> marketing chez Bouygues Immobilier.

- [diʀɛktoʀ]
- [diʀɛktœʀ]
- [diʀɛktøʀ]

24. Nous avons retiré les appareils par crainte du <u>voyeurisme.</u>

- [vwajœʀizm]
- [vwajoʀizm]
- [vwajøʀizm]

25. Elles subissent le sentiment d'<u>insécurité</u>, observe le psychiatre Serge Hefez.

- [ɑ̃sekyrite]
- [ɛ̃səkyrite]
- [ɛ̃sekyrite]

26. En novembre 2005, nous n'avons eu aucun <u>souci</u>.

- [syzi]
- [sysi]
- [susi]

27. les résidences ne résistent pas <u>longtemps.</u>

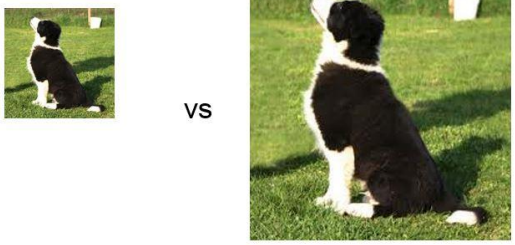
- [lɔ̃tɑ̃p]
- [lɔ̃gtɑ̃]
- [lɔ̃tɑ̃]

Étude transversale

Tâche de dénomination d'image

Image d'entraînement

Images (2 formats)	Séquences cibles	Input écrit (it)	Input oral (fr)
Chien	Un petit chien Un grand chien	*Un cane*	Un chien



Liste des images

Images (2 formats)	Séquences cibles	Input écrit (it)	Input oral (fr)
Arbre	Un petit arbre Un grand arbre	*Un albero*	Un arbre
Etang	Un petit étang Un grand étang	*Uno stagno*	Un étang
Appontement	Un petit appontement Un grand appontement	*Un pontile*	Un appontement
Âtre	Un petit âtre Un grand âtre	*Un focolare*	Un âtre
Appartement	Un petit appartement Un grand appartement	*Un appartamento*	Un appartement

Un albero



vs



Un pontile



vs



Uno stagno



vs



Un focolare



vs



Un appartamento



vs



Répétition et lecture d'une liste spécifique de mots (IPFC-italien)

Liste des mots d'entraînement

1. chat
2. éléphant
3. plante

Liste des mots cibles

Partie commune à toutes les équipes IPFC

1. Andes
2. anse
3. ballade
4. balle
5. bar
6. boule
7. bout
8. bu
9. bulle
10. eau
11. eu
12. eux
13. heure
14. hors
15. Inde
16. là
17. le pas
18. les pas
19. once
20. ou
21. pan
22. panse
23. parade
24. peau

25. peu
26. peur
27. ponce
28. pont
29. port
30. rat
31. tant
32. tante
33. teint
34. teinte

Liste spécifique à l'équipe IPFC-italien

1. bonheur
2. dehors
3. paire
4. flair
5. fleur
6. air
7. euros
8. heureux
9. oraux
10. domaine
11. demain
12. Ils sont
13. Ils ont
14. Hausse
15. ose
16. embrasser
17. embraser
18. thon
19. tonne
20. Louis
21. lui
22. inné

23. aîné
24. il devient
25. ils deviennent
26. exercice
27. exception
28. Jeanne
29. Jean
30. gens
31. gêne
32. entrevue
33. interview
34. héros
35. des
36. deux
37. doux
38. du

Tâche de reformulation

Liste des séquences d'entraînement

	Séquence source	Indice	Séquences cibles	Type de liaison
Première partie : reformulation du premier mot	Devant la table	Derrière	Derrière la table	PRP-DET
Deuxième partie : reformulation du dernier mot	Des filles	Garçons	Des garçons	DET-NOM

Liste des séquences cibles

	Séquence source	Indice	Séquences cibles	Type de liaison
Première partie : reformulation du premier mot	Sur un arbre	Sous	Sous un arbre	PRP-DET
	Sur une plante	Sous	Sous une plante	PRP-DET
	Derrière un arbre	Devant	Devant un arbre	PRP-DET
	Derrière une plante	Devant	Devant une plante	PRP-DET
	Nouvel ami	Ancien	Ancien ami	ADJ-NOM
	Mal écrit	Bien	Bien écrit	ADV-PPA
Deuxième partie : reformulation du dernier mot	Des femmes	Hommes	Des hommes	DET-NOM
	Cette femme	Homme	Cet homme	DET-NOM
	Très malheureux	Heureux	Très heureux	ADV-ADJ
	Moins malheureux	Heureux	Moins heureux	ADV-ADJ
	Plus malheureux	Heureux	Plus heureux	ADV-ADJ
	Toujours malheureux	Heureux	Toujours heureux	DET-NOM
	Il est malheureux	Heureux	Il est heureux	VER-ADJ
	C'est correct	Incorrect	C'est incorrect	VER-ADJ
	Je suis malheureux	Heureux	Je suis heureux	VER-ADJ