Université Paris Nanterre

200 av. de la République
92001 Nanterre Cedex
www.parisnanterre.fr

Membre de l'université Paris Lumières

École Doctorale :
Connaissance, langage, modélisation (ED 139)

Unité de recherche :
Modèles Dynamiques, Corpus (MoDyCo, UMR 7114)

Aline ETIENNE

Analyse automatique des émotions dans les textes :

Contributions théoriques et applicatives dans le cadre de l'étude de la complexité des textes pour enfants

Thèse présentée et soutenue publiquement le 21 juin 2023
en vue de l'obtention du doctorat de Sciences du langage – Traitement automatique des langues de l'Université Paris Nanterre
sous la direction de Mme Delphine Battistelli (Université Paris Nanterre)
et de M. Gwénolé Lecorvé (Orange, Lannion)

Jury :

Rapporteure :	Mme Núria Gala	Professeure des universités, Université Aix-Marseille
Rapporteur :	M. Dominique Legallois	Professeur des universités, Université Sorbonne Nouvelle
Membre du jury :	Mme Delphine Battistelli	Professeure des universités, Université Paris Nanterre
Membre du jury :	Mme Nathalie Blanc	Professeure des universités, Université Paul-Valéry Montpellier 3
Membre du jury :	Mme Anne Lacheret-Dujour	Professeure de universités, Université Paris Nanterre
Membre du jury :	M. Gwénolé Lecorvé	Chercheur, HDR Orange
Membre du jury :	Mme Iva Novakova	Professeure des universités, Université Grenoble Alpes
Membre du jury et présidente du jury :	Mme Sophie Rosset	Directrice de recherche, Université Paris Saclay

Financement et liens utiles

Le présent travail de thèse a été effectué au sein du **projet ANR TextToKids (ANR-19-CE38-0014)** qui en a permis le financement.

Le site Internet du projet est consultable ici : https://texttokids.irisa.fr/.

La chaîne d'extraction automatique de descripteurs de complexité développée dans le cadre de ce projet est testable en ligne à cette adresse : http://vheborto-corliapi.inist.fr:8984/

Le processeur EMOTYC de cette chaîne a été développé dans le cadre de cette thèse pour analyser les émotions.

Université Paris Nanterre
École doctorale 139 – Connaissance, Langage, Modélisation
Laboratoire Modèles dynamiques, Corpus (MoDyCo, UMR 7114)

Thèse de Traitement automatique des langues

présentée par

Aline ETIENNE

Analyse automatique des émotions dans les textes :

Contributions théoriques et applicatives dans le cadre de l'étude de la complexité des textes pour enfants

Thèse dirigée par :

Delphine Battistelli	Professeure des universités, Univ. Paris Nanterre
Gwénolé Lecorvé	Orange, Lannion

Rapporteurs :

Núria Gala	Professeure des universités, Univ. Aix Marseille
Dominique Legallois	Professeur des universités, Univ. Sorbonne Nouvelle

Examinateurs :

Nathalie Blanc	Professeure des universités, Univ. Paul-Valéry Montpellier 3
Anne Lacheret-Dujour	Professeure des universités, Univ. Paris Nanterre
Iva Novakova	Professeure des universités, Univ. Grenoble Alpes
Sophie Rosset	Directrice de recherche, Univ. Paris Saclay

Travail financé par le projet ANR TextToKids (ANR-19-CE38-0014).

Résumé

Notamment de par la diversité des moyens linguistiques employés pour les dénoter, les émotions exprimées dans un texte constituent un objet difficile à circonscrire. Leur étude, abordée ici dans le contexte de l'analyse de la complexité linguistique de textes jeunesse, pose alors de nombreux défis en linguistique comme en traitement automatique des langues (TAL). Cette thèse vise à déterminer comment explorer la dimension émotionnelle d'un texte de sorte à opérer une analyse qui rende compte de la diversité des marqueurs linguistiques des émotions (c'est-à-dire ne se limitant pas au lexique émotionnel) ; qui soit automatisable ; et qui puisse contribuer à mettre au jour des éléments de complexité des textes. L'objectif est donc de proposer des outils théoriques opératoires pour l'analyse linguistique des émotions, mobilisables pour évaluer le caractère plus ou moins accessible – c'est-à-dire compréhensible – d'un texte pour un enfant. La méthodologie mise en œuvre pour cela repose sur la définition d'un schéma d'annotation des émotions, intégrant des notions pertinentes à la fois sur le plan linguistique, TAL et psycholinguistique pour caractériser la notion d'émotion. Son application manuelle sur un corpus de plus de 1 500 textes a permis l'élaboration d'un corpus annoté en émotions, à partir duquel un outil d'analyse automatique des émotions dans les textes a été développé grâce aux techniques d'apprentissage automatique profond (modèle transformeur CamemBERT). Ce corpus annoté a aussi donné lieu à de nombreuses observations linguistiques aidant à mieux cerner le fonctionnement de l'expression des émotions.

Abstract

Because of the diversity of linguistic means used to denote them, emotions expressed in a text are a difficult object to grasp. The study of emotions, tackled here in the context of analyzing the linguistic complexity of texts for children, poses a number of challenges for both linguistics and natural language processing (NLP). The aim of this thesis is to determine how to explore the emotional dimension of a text in a way that enables an analysis that takes into account the diversity of linguistic markers of emotion (i.e. that is not limited to emotional lexicon); that can be automated; and that can contribute to uncovering elements of text complexity. The aim is therefore to propose operational theoretical tools for the linguistic analysis of emotions, which can be mobilized to evaluate the more or less accessible - i.e. comprehensible - nature of a text for a child. The methodology implemented for this purpose is based on the definition of an emotion annotation scheme, integrating notions relevant from linguistic, NLP and psycholinguistic perspectives to characterize the notion of emotion. Its manual application on a corpus of more than 1,500 texts enabled the creation of an emotion-annotated corpus, from which a tool for the automatic analysis of emotions in texts was developed using deep machine learning techniques (CamemBERT transformer model). This annotated corpus has also given rise to numerous linguistic observations, helping to better understand how emotions are expressed.

Remerciements

Ces quelques lignes sont destinées à toutes les personnes sans lesquelles cette thèse n'aurait pu aboutir et qui ont contribué, chacune à leur manière, au bon déroulement de ces presque quatre années de travail.

Je remercie en premier lieu Núria Gala et Dominique Legallois pour avoir, en leur qualité de rapporteurs, permis la soutenance de ma thèse ainsi que Nathalie Blanc, Anne Lacheret-Dujour, Iva Novakova et Sophie Rosset pour avoir participé à mon jury de thèse. Je suis particulièrement reconnaissante de l'intérêt vif que tous ont manifesté à l'égard de mon travail. Merci également à Sylvain Kahane et Mathieu Valette pour la pertinence de leur regard extérieur lors de mes comités de suivi de thèse.

Une très grande part de ma gratitude va bien entendu vers mes directeurs de thèse, Delphine Battistelli et Gwénolé Lecorvé, qui ont rendu tout ce travail possible en construisant, par la finesse et la justesse de leurs retours, leurs encouragements et leur soutien sans faille, un cadre solide dans lequel j'ai pu évoluer en confiance tout au long de ma thèse. Au-delà du plan professionnel, je les remercie pour cette incroyable aventure que j'ai vécue avec eux.

Je tiens aussi à remercier les membres du laboratoire MoDyCo et ceux du projet TextToKids pour m'avoir permis d'explorer le monde de la recherche dans toute sa dimension professionnelle mais aussi humaine. J'adresse un merci tout particulier à Alice, Camilla, Jade, Laura et Philomène pour leurs conseils et leur aide sur mon travail de thèse, mais avant tout pour tous ces précieux moments d'échange et de camaraderie qui ont ensoleillé mes années de doctorat.

Je n'oublie pas non plus les « petites mains », stagiaires – Daria, Katharine, Menel – et étudiantes – Aurore, Maud, Yolène – qui ont su faire preuve du sérieux nécessaire pour cette tâche minutieuse qu'est l'annotation de corpus. Merci aussi à Cyril et Christine pour s'être prêtés au jeu des évaluations perceptuelles.

Mes remerciements vont enfin à mon entourage et en particulier à mes parents et à Jérémy qui ont toujours été là pour m'écouter, même s'ils ne comprenaient pas toujours ce sur quoi je travaillais. À Laure qui n'a jamais hésité à mettre au service de ma thèse ses compétences de journaliste pour garantir des relectures de qualité. À Alain, mon pilier, qui a su tantôt me réconforter, tantôt se réjouir à mes côtés et qui a patiemment été le premier testeur de nombreux moments clés de ma thèse, du schéma d'annotation à l'oral de soutenance.

Table des matières

Résumé....7
Abstract....9
Remerciements....11
Introduction générale....19
1) Contexte sociétal et scientifique....19
2) Problématique et méthodologie....20
3) Organisation du manuscrit....21
Chapitre 1 Contexte applicatif....25
1) Analyse de la complexité linguistique d'un texte destiné aux enfants....26
1.1) Apports de la lisibilité et de la simplification de textes dans l'analyse des la complexité des textes....26
1.2) Étapes développementales et complexité des textes....28
2) Extraire automatiquement des descripteurs de complexité pour l'analyse de textes destinés aux enfants....33
2.1) Constitution d'un corpus de textes classés en âges éditeurs....34
2.2) Vers des descripteurs de la complexité des textes....37
2.3) Étude de corrélations entre âges éditeurs et descripteurs linguistiques de complexité des textes....43
3) Exploration expérimentale de la notion de complexité de textes....46
3.1) Un protocole expérimental pour évaluer l'impact du genre de texte et des dimensions émotionnelle et temporelle sur la compréhension....47
3.2) Résultats préliminaires sur l'évaluation des compétences de compréhension d'enfants de CE2, de CM1 et de CM2....50
Conclusion....53
Chapitre 2 Perspectives psycho-linguistiques, linguistiques et TAL dans l'analyse des émotions....55
1) Les émotions en psycholinguistique....55
2) Les émotions en linguistique....56
2.1) Les concepts clés de la description linguistique des émotions....57
2.2) Interactions entre émotions et d'autres champs sémantiques : le cas de la temporalité linguistique....58
3) Les émotions en TAL....63
3.1) Corpus annotés en émotions....63
3.2) Lexiques émotionnels....69
3.3) Systèmes d'analyse automatique des émotions....73
4) Propositions méthodologiques....76
Conclusion....77
Chapitre 3 Schéma d'annotation manuelle des émotions....79

1) Quatre unités linguistiques pour caractériser l'expression linguistique des émotions....79
1.1) L'unité linguistique exprimant l'émotion....79
1.2) L'unité linguistique désignant l'expérienceur....83
1.3) Les unités linguistiques désignant la cause et la conséquence de l'émotion 84
2) L'annotation des unités linguistiques....85
2.1) Glozz : un outil flexible pour l'annotation manuelle de textes....86
2.2) Les unités SitEmo et Expérienceur....87
2.3) Les unités SitCause et SitConséquence....91
3) Une ébauche de schéma pour annoter la temporalité linguistique des textes....94
Conclusion....98
Chapitre 4 Constitution d'un corpus annoté manuellement en émotions....99
1) Choix méthodologiques pour la délimitation d'unités annotées....100
2) Constitution d'un corpus de textes pour l'analyse de la dimension émotionnelle des textes....105
3) Processus d'annotation : équilibre entre quantité et qualité des annotations. .107
3.1) Annotation en émotions d'un corpus de textes hétérogènes....107
3.2) Mise en place d'un calcul d'accord inter-annotateurs....109
4) Caractérisation chiffrée du corpus annoté....115
Conclusion....123
Chapitre 5 Entraînement de modèles d'analyse automatique des émotions....125
1) Vue d'ensemble....125
1.1) Décomposition du problème....126
1.2) Développer des classifieurs....126
2) Définition des tâches de classification....130
2.1) Sélection de notions....130
2.2) Cinq tâches de classification....132
3) De corpus annoté à corpus d'apprentissage....133
3.1) De marqueurs linguistiques annotés à phrases annotées....133
3.2) Partitionnement du corpus....135
4) Choix méthodologiques....138
4.1) Outils d'apprentissage....139
4.2) Mesures d'évaluation....139
4.3) Facteurs de variation des performances....140
4.4) Mise en œuvre des expérimentations....142
4.5) Baselines....144
5) Résultat des expérimentations....147
5.1) Résultats de la tâche A....148
5.2) Résultats de la tâche B....149
5.3) Résultats de la tâche C....151

5.4) Résultats tâche D....152
5.5) Résultats tâche E....153
5.6) Synthèse....154
Conclusion....155
Chapitre 6 Analyse détaillée des résultats du modèle E....157
1) Performances chiffrées par classes pour le modèle E....158
2) Différences de performances entre le *dev* et le *test*....162
2.1) Piste 1 : composition du *test*....162
2.2) Piste 2 : distance entre les distributions de classes selon les couples (*c*,*m*).
....163
3) Matrices de confusion....164
4) Interactions entre catégorie émotionnelle et mode....167
4.1) Observation 1 : Impact du mode sur la prédiction de la catégorie émotionnelle....167
4.2) Observation 2 : Impact de la catégorie émotionnelle sur la prédiction du mode....169
5) Confrontation des annotations humaines et des prédictions machine....170
6) Évaluation perceptuelle par des experts....173
6.1) Protocole expérimental....173
6.2) Résultats....176
7) Comparaison avec le processeur ÉMOTIONS de la chaîne TextToKids....178
7.1) Principes des comparaisons....179
7.2) Résultats des comparaisons....181
Conclusion....183
Chapitre 7 Quelques observations menées à partir du corpus annoté en émotions..185
1) Analyse des usages des modes d'expression des émotions....185
2) La fonction structurante des émotions observée à partir d'annotations....192
3) Interactions entre deux dimensions sémantiques des textes : la temporalité linguistique et les émotions....203
3.1) Adverbiaux temporels et émotions....203
3.2) Modalité et émotions....208
Conclusion....211
Conclusion générale....215
1) Bilan....215
2) Perspectives....218
2.1) Approfondissement de l'étude de l'expression des émotions....218
2.2) Approfondissement de l'étude de la notion de complexité....219
2.3) Application à d'autres domaines....220
Annexe I Livret individuel....223
Annexe II Arbre de décision pour l'identification des modes d'expression....247

Annexe III Tableaux complémentaires pour le calcul d'AIA....249
Annexe IV Tableaux d'entités des unités Expérienceurs....255
Annexe V Répartition des phrases du corpus global au sein du corpus d'apprentissage
....257
Annexe VI Procédure de transposition des annotations à l'échelle de la phrase....259
Annexe VII Performances détaillées des *baselines*....261
1) Tâche A....261
2) Tâche B....261
3) Tâche C....262
4) Tâche D....262
5) Tâche E....262
6) Performances par classe des modèles de *baseline*....263
Annexe VIII Détails de la comparaison du *dev* et du *test*....265
Annexe IX Compléments aux matrices de confusion....269
1) Tables de croisement des classes....269
2) Matrices de confusion : *test vus* et *test non vus*....270
Annexe X Procédure d'analyse des interactions entre mode et catégorie émotionnelle
(prédictions du modèle E)....271
Annexe XI Tableaux de résultats complémentaires de l'évaluation perceptuelle....273
Annexe XII Consignes du protocole d'évaluation perceptuelle....277
Annexe XIII Compléments pour la comparaison du modèle E' à ÉMOTIONS....285
1) Comparaison avec le module TextBlob....285
2) Comparaison au module EmotaixRemanié....288
Annexe XIV Processus de révision du lexique EMOTAIX....295
Publications personnelles....299
Bibliographie....301

Introduction générale

1) Contexte sociétal et scientifique

Les émotions constituent un champ d'étude difficile à circonscrire, comme l'atteste la grande pluralité de définitions proposées dans le domaine de la psychologie (Piolat & Bannour, 2009). Un certain consensus émerge malgré tout, selon lequel les émotions correspondent à « une réalité psychologique complexe » (*ibid.* : p. 665) impliquant une évaluation cognitive de stimuli internes ou externes à l'individu, pouvant être accompagnée de réactions physiologiques ou comportementales.

Loin de rendre compte de cette complexité définitoire, les travaux de TAL n'abordent le plus souvent l'analyse automatique des émotions dans les textes que par la notion de catégorie émotionnelle (par ex. joie, culpabilité), en se focalisant sur le lexique émotionnel. Les émotions se distinguent alors usuellement d'autres classes sémantiques, nommées « opinions » et « sentiments ». Cette tripartition est par exemple illustrée dans l'édition 2015 du défi fouille de texte (DEFT) qui proposait trois tâches : une d'analyse d'opinions, à travers des catégories comme *Accord* ou *Valorisation* ; une d'analyse de sentiments, à travers les catégories *Satisfaction* et *Insatisfaction*[1] ; et une d'analyse d'émotions à travers des catégories comme *Plaisir, Amour, Mépris* ou *Tristesse* (Hamon et al., 2015).

Du point du vue linguistique, les émotions sont cette fois pensées comme une dimension sémantique des textes. Souvent considérées comme jouant un rôle clé dans la communication, elles sont perçues comme « des entités actives dans le sens où elles agissent sur et participent à la co-construction des actes dans la parole » (Chatar-Moumni, 2013, p. 6). Leur caractérisation fine et précise représente alors un véritable enjeu linguistique. Elle s'avère être de plus un réel défi à cause de la diversité des marqueurs intervenant pour exprimer les émotions, relevant de tous les niveaux de la langue (lexical, syntaxique, discursif, *etc.*). Une typologie fine de ces marqueurs, fondée sur trois modes d'expression des émotions, est proposée dans (Micheli, 2014).

Mes travaux de thèse portent sur **l'analyse des émotions dans le contexte de l'étude de la complexité des textes jeunesse**. Ils prennent en effet place dans le cadre du projet ANR TextToKids (2019-2023) qui vient répondre à un enjeu sociétal

1 Dans d'autres travaux, cette tâche est souvent réduite à l'étude des polarités positive, négative et neutre.

concernant une demande d'outils pour appréhender la complexité des textes à destination des enfants jeunes lecteurs. Avec l'essor des nouvelles technologies, en particulier d'Internet, et leur maîtrise souvent précoce par les enfants, ces derniers sont de fait confrontés à de plus en plus de documents textuels. Mais de par leurs compétences langagières encore en développement, les jeunes lecteurs n'ont parfois qu'un accès partiel au contenu informationnel de ces textes. Se pose donc la question, pour les rédacteurs et éditeurs de contenus jeunesse entre autres, de pouvoir cerner automatiquement et à l'aide de critères objectifs le niveau de difficulté d'un texte pour enfants.

Afin de répondre à ce besoin, le projet TextToKids adopte une démarche pluridisciplinaire, mêlant psycholinguistique, linguistique et TAL pour sonder automatiquement la complexité des textes destinés aux enfants. Cette démarche a pour ambition d'objectiver la notion de complexité et à proposer des pistes pour son évaluation, notamment en la mettant en regard avec des étapes développementales de la compréhension dégagées dans la littérature psycholinguistique. L'accent est en particulier mis sur une dimension sémantique jusqu'ici non explorée dans les travaux portant sur la simplification de textes ou sur la mesure de leur complexité malgré son rôle jugé primordial dans certains travaux en psycholinguistique (Blanc, 2010; Davidson, 2006) : les émotions des personnages.

Une caractérisation fine de l'expression de cette dimension sémantique constitue cependant en linguistique une tâche épineuse dont la difficulté est également attestée en TAL dans les travaux de détection des émotions (voir *supra*). Je défends ici une approche qui considère la notion d'émotion comme un objet d'étude à la croisée de la linguistique, du TAL et de la psycholinguistique. Elle vise en cela à proposer une modélisation des émotions compatible avec une analyse automatique des textes – exploitable pour l'évaluation de leur complexité – et traduisant des étapes développementales identifiées en psycholinguistique, tout en tenant compte de la multitude des moyens linguistiques d'expression des émotions.

2) Problématique et méthodologie

Pour étudier les émotions, perçues comme un facteur de complexité sémantique, la question qui a guidé mon travail est de déterminer comment explorer la dimension émotionnelle d'un texte de sorte à opérer une analyse qui rende compte de la diversité des marqueurs linguistiques des émotions, c'est-à-dire ne se limitant pas au lexique

émotionnel ; qui soit automatisable (du moins en partie) ; et qui puisse contribuer à mettre au jour des éléments de complexité des textes.

L'objectif de cette thèse est donc de proposer des outils théoriques opératoires pour l'analyse linguistique des émotions, mobilisables pour évaluer le caractère plus ou moins accessible – c'est-à-dire compréhensible – d'un texte pour un enfant. Il s'agit donc plus précisément de proposer une modélisation des émotions qui puisse être appliquée de manière systématique sur un grand ensemble de données textuelles, afin de pouvoir d'une part développer un système d'analyse automatique des émotions (méthodes d'apprentissage automatique) et d'autre part opérer des analyses linguistiques rigoureuses de la dimension émotionnelle des textes.

La méthodologie mise en œuvre pour répondre à cela repose sur une modélisation de la notion d'émotion intégrant de nombreux concepts (par ex. catégorie émotionnelle, mode d'expression) pertinents sur le plan linguistique – c'est-à-dire permettant une exploration approfondie de l'expression des émotions –, TAL – c'est-à-dire exploitables pour la détection automatique des émotions – et psycholinguistique – c'est-à-dire rendant compte d'étapes développementales de la compréhension. Cette modélisation prend la forme d'un schéma d'annotation, garantissant ainsi son application systématique à un corpus de plus de 1 500 textes. Ce corpus annoté a servi de fondement pour le développement de modèles informatiques capables d'identifier et de catégoriser automatiquement les émotions présentes dans un texte à l'aide de certaines des notions du schéma. Le corpus a aussi donné lieu à de nombreuses observations linguistiques permettant de mieux cerner le fonctionnement de l'expression des émotions. Elles soulignent notamment des interactions à ce jour jamais sondées entre les émotions et une autre dimension sémantique indispensable à la compréhension de textes : la temporalité ; mais également entre émotions et genre des textes.

3) Organisation du manuscrit

Le **chapitre 1** du manuscrit pose les grandes lignes du contexte applicatif de ma thèse en décrivant sommairement la façon dont la notion de complexité est abordée dans le projet TextToKids et en présentant différents travaux liés à l'étude de cette notion au sein du projet. Ce chapitre rend ainsi brièvement compte de mon implica-

tion tout au long du projet et permet de comprendre l'origine de choix opérés dans mon travail de thèse sur l'analyse des émotions.

Le reste du manuscrit est dédié à ma problématique de thèse, l'étude des émotions des personnages en tant que facteur de complexité des textes. Il présente donc tour à tour les travaux déployés pour répondre aux différents objectifs de ma thèse, à savoir la constitution d'une modélisation de la notion d'émotion, son application méthodique sur un corpus de textes afin de constituer un corpus annoté, puis l'utilisation de ce corpus pour développer des systèmes d'analyse automatique des émotions et effectuer des observations linguistiques plus générales sur l'expression des émotions qui pourraient être utiles à l'analyse de la complexité d'un texte.

Le **chapitre 2** offre tout d'abord un panorama des approches employées en (psycho-)linguistique et en TAL pour analyser les émotions. De ce panorama est dégagé un ensemble de notions à partir duquel est construite ma modélisation pluridisciplinaire des émotions.

Cette modélisation définit quatre unités linguistiques : l'unité émotionnelle à proprement parler, l'expérienceur de l'émotion, sa cause et sa conséquence. Elle se réalise concrètement sous la forme d'un schéma d'annotation des émotions. La modélisation et le schéma qui la concrétise sont présentés au **chapitre 3**.

Le schéma a été appliqué manuellement par six expertes en linguistique à un corpus de plus de 1 500 textes jeunesse de trois genres différents, constituant ainsi un corpus annoté décrit dans le **chapitre 4**.

Le **chapitre 5** est consacré à la présentation d'une série d'expériences informatiques mises en œuvre à partir du corpus annoté afin de tester la pertinence de ma modélisation pour la détection automatique des émotions. Plusieurs classifieurs ont ainsi été appris sur le corpus pour identifier automatiquement les catégories émotionnelles, les types d'émotion et les modes d'expression représentés dans une phrase donnée.

Un examen qualitatif approfondi des résultats obtenus par le plus performant de ces classifieurs est détaillé **chapitre 6**. Les prédictions du classifieur y sont notamment comparées avec des analyses effectuées par des humains et avec celles opérées par un autre système repérant les informations émotionnelles uniquement sur la base d'un lexique émotionnel.

Enfin, au-delà des expériences d'apprentissage automatique, le corpus annoté a donné lieu à des observations linguistiques visant à mieux cerner l'expression des émotions

dans les textes, qui sont exposées au sein du **chapitre 7**. Si les observations menées permettent de souligner les relations entre émotions et genre de textes ainsi qu'à propos de plusieurs notions du schéma entre elles (par ex. entre mode d'expression et catégorie émotionnelle), elles ont également fait émerger un phénomène linguistique jusqu'ici non identifié : le rôle des émotions dans la structuration textuelle. Des analyses ont par ailleurs été consacrées à l'identification des liens unissant les émotions à certaines notions temporelles.

Chapitre 1 Contexte applicatif

L'analyse des émotions – vues comme un facteur de complexité sémantique des textes – dont traite cette thèse s'insère dans le cadre plus large de l'étude de la complexité des textes destinés aux enfants, problématique générale du projet TextToKids. Ce premier chapitre présente ainsi le contexte dans lequel mon travail de thèse a été effectué. Il permet d'autre part de rendre brièvement compte de mon implication dans une série de travaux collaboratifs menés au sein du projet TextToKids. Ce projet, consacré à la compréhension de textes par les enfants sans troubles du développement, approximativement âgés de 6 à 14 ans, a pour objectif d'élaborer des outils, théoriques et pratiques, afin d'identifier automatiquement des portions de textes trop difficiles à comprendre pour un âge cible, tout en proposant des justifications voire des pistes de reformulation. L'étude de la complexité des textes y est abordée avec une approche éminemment pluridisciplinaire, à la croisée de la linguistique, de l'informatique et de la psycholinguistique. Cette étude se réalise concrètement par l'attribution à un texte d'un âge (ou d'une tranche d'âge) en dessous duquel (de laquelle) le texte posera potentiellement des difficultés de compréhension. Elle repose donc sur l'identification de seuils de complexité correspondant à des étapes développementales (et donc à des classes d'âges).

Cette identification s'appuie sur des marqueurs linguistiques de complexité (lexicaux, syntaxiques, *etc.*) pointés dans des travaux de psycholinguistique, de lisibilité et de simplification de textes, auxquels des niveaux de difficulté sont associés (section 1).

Les marqueurs ainsi dégagés sont traduits en descripteurs extractibles automatiquement au sein d'une chaîne de traitement des textes dont le but est d'objectiver l'analyse de la complexité d'un texte pour enfants. Appliquée à un corpus constitué spécifiquement pour le projet TextToKids, composé de textes classés selon les recommandations d'âges émises par les éditeurs, la chaîne met en regard ces âges éditeurs avec les normes développementales issues de travaux psycholinguistiques (section 2).

Le corpus de textes classé en âges éditeurs et les marqueurs de complexité ont étayé la conception d'un protocole expérimental visant à sonder plus en avant les processus de compréhension de textes chez les enfants. Ce protocole, particulièrement ciblé sur l'évaluation de l'impact des émotions et de notions temporelles dans la compréhension

de différents genres textuels (encyclopédique, fictionnel et journalistique), a d'ores et déjà pu être testé auprès d'enfants de CE2, de CM1 et de CM2 (section 3).

Le projet TextToKids est structuré en différents axes, appelés « *Work Package* » (WP), chacun impliquant plusieurs membres. L'un de ces axes est consacré à l'exploration expérimentale de la compréhension de textes : j'y ferai référence à plusieurs reprise dans ce chapitre et dans la suite du manuscrit par l'acronyme WP7.

1) Analyse de la complexité linguistique d'un texte destiné aux enfants

Les travaux menés dans le projet TextToKids ne se revendiquent pas directement des disciplines de la lisibilité et de la simplification de textes. Ces deux domaines proposent malgré tout des perspectives d'analyse riches, approchantes de celles du projet (section 1.1). L'apport méthodologique majeur de TextToKids consiste cependant à explorer la complexité des textes à l'aide de marqueurs linguistiques de complexité issus de la littérature psycho-linguistique rendant compte d'étapes développementales et ainsi associables à des âges seuil de compréhension (section 1.2).

Volontairement succinct et non exhaustif, le tableau des différentes recherches en lisibilité, en simplification de texte et en psycholinguistique dressé dans cette section a pour seule ambition de présenter dans les grandes lignes quelques travaux clés afin de positionner l'approche proposée dans le projet TextToKids.

1.1) Apports de la lisibilité et de la simplification de textes dans l'analyse des la complexité des textes

En TAL, l'analyse de la complexité d'un texte a été principalement abordée par les travaux en lisibilité et en simplification de textes. La lisibilité est un champ d'étude ancien, fondé sur la définition de *formules de lisibilité*, correspondant à des modèles statistiques visant à prédire la complexité d'un texte (Wilkens et al., 2022). Deux formules historiques sont par exemple celle de R. Flesh (1948) qui s'appuie sur le nombre de syllabes ou la longueur des phrases ou celle de E. Dale et J. S. Chall (Dale & Chall, 1948), qui utilise également la longueur des phrases mais aussi la proportions de mots simples. De manière générale, les travaux de lisibilité s'intéressent à catégoriser des textes en niveaux de complexité à l'aide de critères linguistiques (voir

par ex. (Forti et al., 2019)) et à proposer éventuellement des reformulations (voir par ex. (Vajjala & Meurers, 2014)). Il y est ainsi courant de recourir à des caractéristiques aisément calculables telles que la longueur moyenne d'une phrase (vue comme une approximation pour la complexité syntaxique), la longueur moyenne des mots en syllabes ou en caractères (vue comme une approximation de la complexité au niveau lexical) ou encore la présence de mots rares.

Les travaux de simplification de textes ont quant à eux pour objectif de proposer des textes adaptés à un public cible selon deux types principaux de démarche soulignés dans (Gala et al., 2018). La première est la création d'un texte respectant un ensemble de critères destinés à faciliter la compréhension par le public visé. La seconde est l'adaptation d'un texte de sorte à le rendre plus compréhensible pour le public cible, en modifiant sa forme (par ex. suppression de passages moins informatifs) ou ses caractéristiques linguistiques (lexique, syntaxe), tout en conservant son contenu sémantique. Ce sont des travaux relevant de cette deuxième démarche, tel que par exemple (Gala et al., 2020), que le projet TextToKids se rapproche le plus. S. Elguendouze (2020) signale par ailleurs que les travaux de simplification de textes s'appuient bien souvent sur les études de lisibilité en usant de leurs formules pour évaluer les productions simplifiées ou pour identifier les zones de complexité à simplifier. Les facteurs simplifiés correspondent de plus à ceux souvent inclus dans les calculs de lisibilité, avec sur le plan lexical la longueur des mots ou leur niveau de difficulté (évalué souvent d'après leur fréquence), sur le plan syntaxique certaines constructions jugées complexes et sur le plan discursif des anaphores jugées difficiles.

Il est à noter qu'en simplification ou en lisibilité, les cas d'usage sont souvent liés à l'apprentissage d'une langue seconde (François & Fairon, 2012), à la compréhension de textes en domaine spécialisé, comme par exemple le domaine médical (Cardon & Grabar, 2021) ou encore au contexte pédagogique de choix de textes adaptés à des enfants du niveau de l'école primaire, par ex. (Gala et al., 2020; Imperial & Ong, 2021).

Marqueurs de complexité issus de la lisibilité et de la simplification. S'ils ne peuvent être associés à des âges frontières, les critères de complexité issus de travaux de lisibilité ou de simplification représentent cependant des pistes précieuses pour l'analyse de la complexité des textes pour enfants. Des marqueurs lexicaux récurrents sont par exemple la diversité du lexique d'un texte (François et al., 2018), la fréquence des mots ou leur longueur (Elguendouze, 2020; Gala et al., 2018). De nom-

breux marqueurs d'ordre morphosyntaxique sont également identifiés, tels que la fréquence des flexions verbales (Elguendouze, 2020; Gala et al., 2018), la proportion des différents modes temporels (indicatif, subjonctif, *etc.*), des différents temps verbaux, des différents nombres pour les verbes fléchis (singulier, pluriel), des différentes personnes pour les pronoms et verbes fléchis ou des classes grammaticales (François & Fairon, 2012). Sont parfois mobilisés des marqueurs phonétiques, tels que la longueur des mots en phonèmes (Elguendouze, 2020; Gala et al., 2018) ou le score d'ordinarité phonétique des mots (Gromer & Weiss, 1990). Des exemples de marqueurs syntaxiques sont la longueur des phrases avec prise en compte de pronoms relatifs ou de conjonctions de subordination ou de coordination (De Belder & Moens, 20100101; Elguendouze, 2020; Gala et al., 2018) ou le ratio entre signes de ponctuation et mots (François & Fairon, 2012). Les marqueurs sémantiques, tels que les types d'expressions référentielles (syntagme nominal ou pronom) (Elguendouze, 2020; Gala et al., 2018) ou la présence de lien logique non explicites (Gala et al., 2018), restent rares.

L'approche proposée dans le projet TextToKids s'appuie sur les résultats de la plupart des travaux cités ci-dessus en termes marqueurs linguistiques jugés pertinents ; elle s'en distingue par le choix méthodologique de faire intervenir des marqueurs linguistiques supplémentaires qui soient motivés sur le plan développemental selon certains travaux en psycholinguistique sur la tranche des enfants jeunes lecteurs. À notre connaissance aucune démarche de ce type n'a en effet été proposée dans les domaines de la lisibilité et de la simplification de textes à destination d'enfants. De plus, ces marqueurs sont pour certains de nature sémantique et jusqu'ici jamais exploités pour l'analyse automatique de la complexité.

1.2) Étapes développementales et complexité des textes

La description de la complexité d'un texte à destination des enfants est abordée dans le projet TextToKids par le prisme de l'évolution des compétences linguistiques de compréhension. Ces dernières sont tantôt évaluées sur des énoncés isolés ciblant des caractéristiques linguistiques précises (voir par ex. (Lecocq, 1998)), tantôt sur des tâches de compréhension de textes entiers (voir par ex. (Blanc, 2010)). La compréhension est alors souvent définie, à la suite de (Zwaan & Radvansky, 1998, p. 162), comme la capacité d'un individu à se construire, au fur et à mesure de la découverte d'un texte, une représentation mentale de la situation décrite par le texte. La construction de cette représentation nécessite l'intégration de plusieurs dimensions si-

tuationnelles telles que par exemple l'espace, le temps, la causalité ou les entités de façon à retisser les liens qui unissent les situations décrites dans un texte. Certains travaux comme (Blanc, 2010; Dijkstra et al., 1995; Dyer, 1983) proposent pour leur part de recourir en plus à la dimension situationnelle des émotions des personnages.

En évaluant l'évolution des compétences de compréhension, les études psycholinguistiques mettent en lumière des marqueurs linguistiques représentant des points de difficultés potentielles pour les enfants. Elles donnent de plus une indication globale des âges auxquels ces difficultés peuvent se présenter. Les grandes tendances développementales identifiées sont bien sûr à pondérer du fait de la quasi-impossibilité à effectuer des associations strictes entre un marqueur linguistique et un âge auquel il est compris. Cela s'explique en premier lieu par les différences inter-individuelles de développement des compétences linguistiques. En second lieu, des différences du point de vue des matériaux linguistiques (énoncés isolés, textes entiers, *etc.*) et des procédés d'évaluation (réponse orale à une question, choix d'une image, *etc.*) employés pour mesurer les capacités de compréhension chez les enfants engendrent parfois des estimations divergentes des âges auxquels certains marqueurs linguistiques sont maîtrisés comme le remarque Paul W. van den Broek (1997). Malgré ces réserves, il reste que les marqueurs psycholinguistiques sont utiles pour échelonner la complexité des textes à destination des enfants du point de vue d'âges frontières, et non plus du simple point de vue d'une opposition entre texte accessible et texte non-accessible comme cela est le cas dans les travaux de lisibilité et de simplification.

Marqueurs morphologiques et lexicaux de complexité. Sur le plan morphologique, (Lecocq, 1998) met en évidence la difficulté, pour les enfants de moins de 9 ans, de tenir compte des marques du pluriel lors de la lecture d'un texte. Sur le plan lexical, ce sont les adjectifs ordinaux qui posent problème, jusqu'à 10 ans en modalité écrite et 9 ans en modalité orale (*ibid.*)[2].

Marqueurs syntaxiques de complexité. Du point de vue syntaxique, les marqueurs de difficulté identifiés dans la littérature psycholinguistique sont nombreux. Jusqu'à 9 ans environ, la résolution des anaphores est ainsi difficile, en particulier lorsqu'elles sont ambiguës (par ex. *Le garçon ne voit pas le monsieur bien qu'il porte des lunettes.*) (Lecocq, 1998), lorsque le référent du pronom est situé en dehors de la

2 Chaque âge frontière présenté ici correspond à l'âge en dessous duquel moins de 75 % des enfants évalués réussissent l'épreuve de compréhension.

phrase ou lorsqu'il avait un rôle sémantique différent dans la proposition précédente (par ex. *La maîtresse a salué le directeur quand il est entré dans la classe.*) (Kail & Léveillé, 1977). Les phrases présentant un objet clivé (par ex. *C'est le corsaire que le cosmonaute embrasse.*) (Bronckart, 1979; Vion & Amy, 1984) ainsi que certaines structures passives (par ex. *Le cheval est poursuivi par l'homme.*) (Bronckart, 1979; Lecocq, 1998) posent également des soucis d'interprétation[3]. En modalité écrite, les phrases avec une relative en QUI (par ex. *Le chien poursuit le cheval qui se retourne.*) et les superlatifs d'infériorité sont parfois difficilement comprises (Lecocq, 1998). En dessous de 10 ans, ce sont cette fois les phrases contenant une proposition relative circonstancielle (par ex. *Le livre sur lequel est posé le crayon est rouge.* ou *Le chat dont les yeux sont verts regarde la dame.*) (Lecocq, 1998) et celles contenant une relative avec un participe présent (par ex. *La vache poursuivant le chat est marron.*) (Bronckart, 1979; Lecocq, 1998) qui peuvent représenter des nœuds de difficulté. Les relatives en QUE constituent parfois un problème jusqu'à 10 ans (par ex. *La voiture blanche que la voiture rouge pousse.*) (Segui & Léveillé, 1977) voire jusqu'à 13 ans (par ex. *La vache que le chien poursuit est marron.*) (Lecocq, 1998).

Marqueurs sémantiques de complexité. De nombreux facteurs de difficulté sémantiques sont aussi identifiés dans la littérature psycholinguistique. Jusqu'à 9 ans environ, certains termes polysémiques posent par exemple des difficultés de compréhension (par ex. *un cœur droit, un esprit tordu*) (Tamine & Bonnet, 1982). Il est intéressant de noter que le nombre de sens associés à un mot est aussi un critère de complexité des travaux de lisibilité (Elguendouze, 2020) et de simplification de textes (Gala et al., 2018). Différentes valeurs modales (par ex. *je crois que P, je sais si P*) ne sont pas prises en compte par les enfants de moins de 9 ans qui ne retiennent que le sens de la proposition principale (Bassano, 1985; Hickmann et al., 1993). V. Tartas (2001, 2010) indique quant à elle que les enfants, de moins de 9 ans à nouveau, rencontrent des difficultés à manipuler les notions calendaires comme les jours, les mois ou encore les saisons, que ce soit pour positionner ou ordonner des évènements, ou maîtriser le caractère cyclique de ces notions et les liens qui les unissent. Reconstituer l'ordre d'évènements présentés de manière non chronologique dans des énoncés du type *Ilkka a lu la lettre après être allée à l'école.* pourrait également présenter une

3 Ce résultat n'est pas reproduit dans (Segui & Léveillé, 1977; Vion & Amy, 1984). Ceci s'explique peut-être par des différence de méthodologie : dans (Lecocq, 1998) la passation est écrite, avec ensuite une tâche de désignation d'image pour vérifier la compréhension (la méthodologie employée n'est pas précisée dans (Bronckart, 1979)) tandis que (Segui & Léveillé, 1977; Vion & Amy, 1984) sollicitent tous deux une passation orale, avec une évaluation de la compréhension reposant respectivement sur des questions ou le fait de mimer la situation.

difficulté[4] et ce jusqu'à environ 12 ans pour des énoncés du type *Max s'est mis à pleurer, Zoé étant partie à la piscine.* (Pyykkönen & Järvikivi, 2012). Jusqu'à 11 ans environ, les enfants éprouve des difficultés à inférer la conclusion d'un énoncé argumentatif utilisant « même » (Bassano & Champaud, 1989). La plupart des métaphores restent difficiles à interpréter pour les enfants de moins de 12 ans (Tamine & Bonnet, 1982). Enfin, (van den Broek, 1997) fait état d'une difficulté à identifier le thème commun entre différents épisodes d'une histoire, soit à identifier les relations thématiques et les morale d'une histoire jusqu'à environ 14 ans.

À propos d'une dimension sémantique en particulier : les émotions. Une dimension sémantique des textes en particulier, la dimension émotionnelle, se trouve au cœur du projet TextToKids. L'intérêt porté à cette dimension s'ancre sur le rôle qu'elle joue dans la compréhension de textes. Chez les enfants en particulier, (Blanc & Quenette, 2017) souligne ainsi que l'incidence de la dimension émotionnelle d'un texte dans le processus de compréhension repose sur deux mécanismes principaux : d'une part, les émotions des personnages focalisent l'attention de l'enfant lors de l'activité de compréhension ; d'autre part, elles aident à maintenir la cohérence de la représentation élaborée du fait que « les ressentis d[es] personnage[s] incitent à établir des liens de nature causale avec les événements qui les ont suscités » (ibid. : 505). En établissant des liens de causalité entre une situation et une émotion, l'enfant serait par ailleurs plus à même d'identifier les évènements les plus importants de l'histoire. L'impact des émotions des personnages sur la construction et la mémorisation de la représentation mentale de la situation décrite par un texte se retrouve également chez les adultes (Dijkstra et al., 1995; Dyer, 1983). Du point de vue psycholinguistique, il apparaît ainsi que quand des émotions sont exprimées dans un texte, elles sont considérées comme un facteur important pour la compréhension du texte. Cette importance relève, d'une part, du rôle qu'elles jouent dans la saillance accordée aux situations auxquelles elles sont liées et, d'autre part, du rôle qu'elles jouent dans l'élaboration de liens de causalité entre différentes situations décrites dans un texte.

Si la présence d'émotions dans un texte facilite de manière générale le processus de compréhension du texte, des étapes développementales dans leur propre acquisition

4 Les travaux de synthèse présentés dans (Hickmann, 2012) indiquent une éventuelle influence de la langue maternelle sur l'acquisition de la temporalité : les résultats de (Pyykkönen & Järvikivi, 2012), obtenus sur des enfants finnophones, ne s'appliquent ainsi peut-être pas sur des enfants francophones. Ces résultats ne sont par ailleurs pas reproduits dans (Blything et al., 2015) (enfants anglophones). Cela s'explique peut-être par les méthodologie d'évaluation très différentes mises en œuvre. (Blything et al., 2015) emploie la modalité audio-visuelle et évalue la compréhension à l'aide d'exercices ludiques sur écran tactile tandis que (Pyykkönen & Järvikivi, 2012) s'appuie sur la modalité écrite et un questionnaire (écrit également).

peuvent être dégagées. Jusqu'à 10-11 ans environ, les catégories émotionnelles négatives (en particulier la peur) ne sont ainsi pas clairement identifiées par les enfants (Baron-Cohen et al., 2010; Simoës-Perlant & Lemercier, 2018). (Davidson, 2006) et (Blanc & Quenette, 2017) dégagent quant à eux une norme développementale concernant la maîtrise des types d'émotions. Plus spécifiquement, ils constatent une meilleure compréhension des émotions de base (joie, colère, *etc.*) que des émotions complexes (embarras, culpabilité, *etc.*) lorsque l'émotion est directement désignée par un terme du lexique émotionnel (joyeux, honteux, *etc.*), et ce jusqu'à 10-11 ans[5]. Ce constat est également effectué pour les émotions suggérées. De manière générale, les enfants de moins de 10-11 ans ont du mal à identifier l'émotion du personnage lorsqu'elle est simplement suggérée par un évènement inducteur d'émotion (par ex. *Elle retrouve un ami cher après un longue séparation.* suggère la joie) que lorsqu'elle est explicitement mentionnée (par ex. *heureux*) ou exprimée par un comportement émotionnel (par ex. *il sourit*) (Blanc, 2010; Blanc & Quenette, 2017; Creissen & Blanc, 2017). Les trois façons d'exprimer les émotions qui se dégagent des études psycholinguistiques – le fait de suggérer une émotion, de la désigner par un terme du lexique ou de décrire un comportement qui en résulte – sont l'un des piliers de la modélisation des émotions proposée dans le présent travail (*cf.* chapitres 2 & 3).

Marqueurs exploratoires de complexité. En s'inspirant des marqueurs sollicités dans les travaux de lisibilité et de simplification et de ceux issus des études psycholinguistiques, un petit nombre de marqueurs intuitivement jugés pertinents pour l'analyse de la complexité sont proposés dans le projet TextToKids pour permettre une analyse plus complète. Ces marqueurs, fondés sur l'intuition, sont nommés marqueurs « exploratoires ». L'échelle Dubois-Buyse (Ters et al., 1988) qui offre un vocabulaire de 4 000 mots courants, classés selon leur niveau de difficulté orthographique sur une échelle allant du CP au lycée, est donc sollicitée dans le projet. Deux marqueurs morphosyntaxiques, le nombre de temps verbaux différents ainsi que la proportion des différents nombres pour les pronoms ont été retenus. Trois marqueurs syntaxiques sont également avancés : les distances moyenne et maximale entre un mot et ses dépendants, le nombre de dépendances entrantes/sortantes pour chaque mot et la profondeur de l'arbre de dépendances. Du point de vue sémantique, la présence et le nombre de connecteurs logiques par différentes catégories (addition, but, cause, comparaison, concession, *etc.*) ainsi que les entités nommées sont pris en compte.

5 Résultat également confirmé dans (Baron-Cohen et al., 2010).

Un ensemble de marqueurs linguistiques auxquels des niveaux de complexité – idéalement sous la forme d'âges d'acquisition – peuvent être associés ont ainsi été retenus dans le projet TextToKids. Ces marqueurs que j'ai directement contribué à identifier sont au fondement d'une chaîne d'extraction automatique de descripteurs linguistiques de complexité (*cf.* section 2.2). Ce sont les suivants :

- 17 marqueurs de complexité issus de travaux de lisibilité ou de simplification de textes (4 lexicaux, 1 morphologique, 7 morphosyntaxiques, 2 phonétiques, 2 syntaxiques et 2 sémantiques). Parmi eux, deux – la fréquence des morphèmes composant les mots et la fréquence des flexions verbales dans la langue française – n'ont pu être exploités faute de ressources disponibles ;
- 39 marqueurs de complexité issus d'études psycholinguistiques (1 lexical, 1 morphologique, 12 syntaxiques et 25 sémantiques). Parmi eux, quatre – l'identification de la morale d'une histoire, le clivage de l'objet, l'utilisation de « même » pour inférer des conséquences contre-intuitives et les termes polysémiques – n'ont pas pu être intégrés dans la chaîne d'analyse de la complexité car leur repérage a été jugé trop difficile à automatiser ;
- 8 marqueurs de complexité exploratoires (1 lexical, 2 morphosyntaxiques, 3 syntaxiques et 2 sémantiques).

Un tableau résumant tous ces marqueurs et explicitant leur lien avec les descripteurs de la chaîne d'analyse sera mis en ligne au terme du projet (prévu pour fin 2023).

2) Extraire automatiquement des descripteurs de complexité pour l'analyse de textes destinés aux enfants

Telle qu'elle est abordée dans le projet TextToKids, l'analyse automatique de la complexité des textes pour enfants nécessite deux éléments clés. Le premier est la constitution d'un large corpus de textes destinés aux enfants (section 2.1). Ce corpus est au cœur d'expérimentations d'ordre linguistique, informatique et psycholinguistique. Le second est une chaîne de traitement des textes, chargée d'extraire automatiquement des descripteurs de complexité des textes (section 2.2). Ces descripteurs sont le fruit de la traduction des marqueurs de complexité linguistique relevés dans la section 1.

2.1) Constitution d'un corpus de textes classés en âges éditeurs

Afin de mener des analyses linguistiques et des expériences informatiques pour l'exploration de la complexité des textes, il est nécessaire de disposer d'un corpus de textes, échelonné en niveaux de difficulté. La notion de complexité étant abordée au sein du projet comme l'association d'un âge ou d'une tranche d'âge à un texte, en dessous desquels le texte posera des difficultés de compréhension, les niveaux de difficulté associés aux textes se devaient de correspondre à de tels âges ou tranches d'âge. Le travail préliminaire présenté dans (Blandin et al., 2020) soulignait déjà qu'aucun corpus de ce type n'existait. Une tâche centrale du projet dans laquelle j'ai été impliquée a donc été la constitution d'un tel corpus. Trois décisions ont alors dû être prises : celle concernant les âges servant à structurer le corpus, celle des types de textes inclus dans le corpus et celle des sources. De plus, il existe actuellement plusieurs versions du corpus, chacune répondant à des usages spécifiques.

Un corpus classé en âges éditeurs. Dans un corpus idéal, les âges associés aux textes sont issus d'évaluations psycholinguistiques de la compréhension, menées auprès d'enfants, afin de refléter la « vérité terrain ». Cependant, au vu du temps considérable nécessaire pour mener des expériences de ce type et étant donné le grand volume de données requis par certaines méthodes informatiques employées dans le projet (par ex. apprentissage automatique), il n'était pas envisageable de construire un tel corpus. Il a donc été décidé de partir des nombreuses recommandations d'âge associées par les éditeurs aux textes qu'ils diffusent (ci-après nommés « âges éditeurs »). Les ouvrages de la collection C*hien Pourri* de C. Gutman sont par exemple associés à la tranche 6-8 ans par la maison d'édition L'école des loisirs et le journal *Albert* à la tranche 9-14 ans par la maison d'édition La Poule Qui Pond. Ces recommandations sont le plus souvent fixées de manière subjective, sans qu'une grille de critères (linguistiques) explicites ne soit utilisée. Elles répondent ainsi le plus souvent à des normes commerciales et marketing et ne sont pas attribuées par des experts en développement du langage. Elles traduisent malgré tout un certain niveau d'expertise, acquis avec l'expérience et qui s'avère de fait cohérent avec les observables linguistiques présents dans les textes (*cf.* section 2.3).

L'un des usages prévus pour ce corpus classé en âges éditeurs est l'entraînement d'un modèle capable de prédire l'âge adéquat associé à un texte. En plus des textes destinés aux enfants, il était alors nécessaire d'inclure des textes représentant le niveau de

difficulté le plus élevé, soient des textes destinés aux adultes. Pour les besoins des expériences informatiques, cette catégorie « adultes » a été traduite par la tranche d'âge 14-18 ans. Dans une version du corpus encore en cours de développement, cette tranche adultes a finalement été fixée aux âges 19-25 ans. La tranche 12-18 ans a été ajoutée pour tenir compte des ouvrages destinés aux jeunes adultes[6].

Un corpus hétérogène. Pour assurer sa représentativité, trois genres de textes ont été inclus dans le corpus : l'encyclopédique, le fictionnel et le journalistique. Un effort a également été fourni pour inclure, pour chaque genre, plusieurs sources de textes. Les textes fictionnels pour adultes sont issus pour la plupart d'œuvres libres de droit disponibles sur le site du projet Gutenberg[7]. Ceux pour enfants sont principalement tirés de manuscrits d'ouvrages de littérature jeunesse, fournis par les auteurs eux-mêmes (par ex. S. Morgenstern, A. Desarthe, B. Friot ou G. Nail) ou la maison d'édition (L'école des loisirs pour les ouvrages de M.-A. Murail). Deux partenaires du projet, les journaux *P'tit Libé* et *Albert* ont mis à disposition l'ensemble de leurs numéros qui constituent ainsi les deux sources principales des textes journalistiques jeunesse. Des articles de journaux tels que *Le monde diplomatique* ou *L'humanité* ont été choisis manuellement pour composer l'ensemble des textes journalistiques pour adultes. Les textes encyclopédiques sont les moins diversifiés en sources. Ils proviennent ainsi majoritairement d'encyclopédies collaboratives en ligne, facilement traitables automatiquement et libres de droit, Wikipédia pour les adultes et Vikidia pour les enfants (8-13 ans).

Un corpus multiple. Il existe actuellement plusieurs versions de corpus, chacune ayant été constituée à des moments différents du projet, pour répondre à des objectifs spécifiques. À ce jour, le corpus général de référence est celui décrit dans (Rahman et al., 2020) – ci-après nommé corpus Coling2020. Ce corpus a permis l'entraînement d'un modèle de prédiction automatique de l'âge associé à un texte. Il est donc divisé en trois sous-ensembles : l'ensemble d'entraînement, celui de développement et celui de test. Les caractéristiques de ce corpus, comprenant au total de 1 487 textes (157 177 phrases) sont résumées par le tableau 1. La composition en terme de tranches d'âges éditeur est synthétisée tableau 2 et illustrée figure 1.

6 Cette tranche est celle proposée dans l'article Wikipédia consacré à la littérature « Young Adult » https://fr.wikipedia.org/wiki/Litt%C3%A9rature_Young_Adult.

7 https://www.gutenberg.org/

	Entraînement			**Développement**			**Test**		
	Nb textes	Nb phrases	Âge moyen	Nb textes	Nb phrases	Âge moyen	Nb textes	Nb phrases	Âge moyen
Enc.	254	40 000	14,63	57	10 473	14,06	47	7 958	14,73
Fic.	397	47 354	10,44	93	12 566	9,38	78	13 237	10,24
Jou.	391	18 247	11,64	74	3 224	11,58	96	4 118	11,67
Total	**1 042**	**105 601**	**12,25**	**224**	**26 263**	**11,53**	**221**	**25 313**	**11,90**

Tableau 1: Caractéristiques du corpus Coling2020, selon le genre de texte et les ensembles d'entraînement, de développement et de test

	Entraînement	**Développement**	**Test**
Enc.	3-9, 4-8, 8-12, 8,13, 14-18	4-8, 8-12, 8,13, 14-18	4-8, 8-12, 8,13, 14-18
Fic.	0-2, 0-3, 2-7, 3-5, 3-6, 3-7, 3-8, 3-9, 4-7, 4-8, 4-9, 5-12, 5-7, 5-8, 5-9, 6-8, 6-9, 7-11, 8-10, 8-11, 8-13, 10-12, 10-13, 10-14, 11-13, 12-14, 14-18	0-3, 3-5, 3-6, 3-7, 3-8, 4-6, 4-8, 4-9, 5-9, 6-8, 6-9, 8-11, 10-12, 14-18	0-3, 2-4, 3-5, 3-6, 3-7, 3-8, 3-9, 4-6, 4-7, 4-8, 4-9, 5-9, 6-12, 6-8, 6-9, 7-11, 8-11, 10-12, 10-14, 14-18
Jou.	3-9, 7-12, 12-14, 14-18	3-9, 7-12, 12-14, 14-18	3-9, 7-12, 12-14, 14-18

Tableau 2: Répartition des tranches d'âges éditeur selon le genre de texte, entre les ensembles d'entraînement, de développement et de test du corpus Coling2020

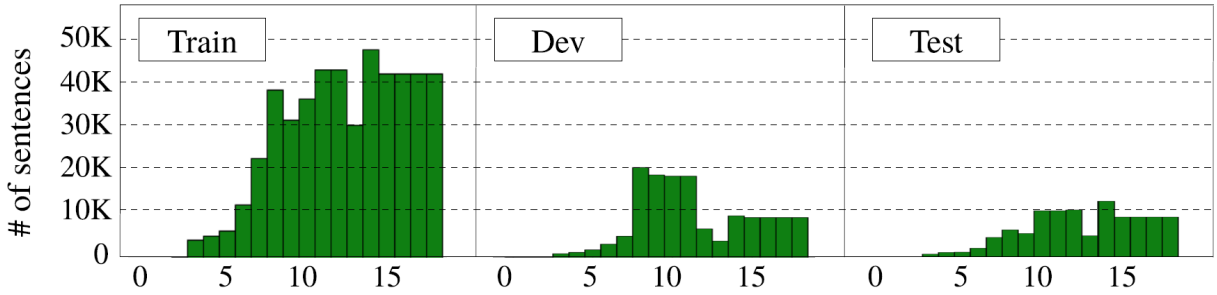


Figure 1: Distribution des phrases du corpus Coling2020 selon les différents âges dans les ensembles d'entraînement (train), de développement (dev) et de test – Figure reprise de (Rahman et al., 2020)

Seule une sous-partie de ce corpus, une portion de l'ensemble des textes fictionnels, a été sollicitée lors d'expériences de mise en corrélation de descripteurs linguistiques et d'âges éditeurs, présentées dans (Battistelli, Etienne, et al., 2022) (*cf.* section 2.3). Ce corpus est nommé TALN2022.

Présentant des recouvrements partiels avec le corpus Coling2020, un corpus dédié à l'analyse des émotions a été constitué. Il a été annoté en émotions (*cf.* chapitre 4) et a permis la mise en œuvre d'expériences de développement de systèmes d'analyse automatique des émotions (*cf.* chapitre 5). Ses caractéristiques sont présentées plus en détails chapitre 4.

Un corpus spécifique au WP7 du projet TextToKids, présentant des chevauchements partiels avec le corpus Coling2020 d'une part et le corpus annoté en émotions d'autre part, existe également. Ce corpus a été constitué pour permettre l'élaboration de protocoles expérimentaux d'évaluation de la compréhension chez les enfants plus spécifiquement axée sur les dimensions émotionnelle et temporelle des textes. Les caractéristiques de ce corpus sont exposées section 3.1.

Enfin, une version révisée du corpus Coling2020 à laquelle je collabore est en cours d'élaboration. Cette version, toujours divisée en ensembles d'entraînement, de développement et de test, constituera le livrable final du versant corpus du projet. L'une des modifications majeures porte sur la tranche d'âge de la catégorie « adultes », passant de 14-18 ans à 19-25 ans (*cf. supra*) afin de mieux échelonner les niveaux de difficultés des textes et de mieux rendre compte de la diversité des contenus de littérature jeunesse. De manière générale, un effort particulier sera concentré, dans cette version révisée, à l'augmentation de la diversité des sources de textes et ce en particulier pour le genre fictionnel. L'objectif de la version révisée est en outre de rendre plus comparables les travaux menés dans les différents axes du projet. L'ensemble des textes du corpus WP7 intégrera par exemple la partie test du corpus révisé, afin de permettre une comparaison plus efficace des résultats des expériences psycholinguistiques avec les résultats de modèles de prédiction de l'âge associé à un texte.

2.2) Vers des descripteurs de la complexité des textes

L'un des principaux apports du projet TextToKids est une chaîne de traitement automatique des textes proposant l'extraction de différentes caractéristiques linguistiques, utiles à l'analyse de la complexité des textes destinés aux enfants. Si le noyau de cette chaîne a été développé dans (Blandin et al., 2020), un travail conséquent de refonte a été effectué et présenté dans (Battistelli, Etienne, et al., 2022). Cette refonte s'est poursuivie et la dernière version à jour de la chaîne est disponible et testable en ligne à cette adresse : http://vheborto-corliapi.inist.fr:8984/chain/. Trois versions seront donc mentionnées ici : la chaîne dite « initiale », présentée dans (Blandin et al., 2020) ; une première amélioration de cette chaîne nommée « chaîne TALN2022 », présentée dans (Battistelli, Etienne, et al., 2022) ; et la dernière version de la chaîne, dite « révisée ». La structure de la chaîne initiale a été conservée au fur et à mesure des remaniements. La chaîne révisée est donc organisée en modules appelés « processeurs » as-

sociés à un niveau d'analyse de la langue (par ex. phonétique, sémantique), chacun gérant l'extraction d'un ensemble de caractéristiques linguistiques spécifiques.

Traduction des *marqueurs* de complexité en *descripteurs* de complexité. Le remaniement de la chaîne initiale a été effectué grâce aux marqueurs de complexité linguistique identifiés dans la littérature psycholinguistique, en lisibilité et en simplification de textes (*cf.* section 1.2). Ces marqueurs de complexité ont ainsi été traduits en éléments extractibles automatiquement et sont nommés *descripteurs* linguistiques de complexité. Le marqueur de complexité correspondant au caractère cyclique des notions temporelles (heures, jours, mois, *etc.*), difficile pour les enfants jusqu'à 9 ans environ (*cf.* section 1.2), repose par exemple sur l'identification d'adverbiaux temporels itératifs. Cela se traduit explicitement dans la chaîne révisée par neuf couples de descripteurs, où l'un des descripteur correspond à un nombre brut d'adverbiaux identifiés et l'autre à une proportion parmi l'ensemble des adverbiaux temporels du texte analysé par la chaîne[8]. L'extraction de ces descripteurs est gérée par le processeur sémantique « adverbiaux temporels » de la chaîne révisée.

À propos des marqueurs de complexité émotionnels. Un autre exemple de marqueur de complexité, la difficulté des enfants à comprendre les émotions suggérées et ce jusqu'à 10-11 ans (*cf.* section 1.2) s'ancre sur l'identification de phrases exprimant une émotion suggérée. Il se traduit ainsi par un couple de descripteur dans la chaîne révisée : l'un exprime le nombre de phrases comportant au moins une émotion suggérée et l'autre la proportion de ces phrases au sein du texte analysé. Dans la chaîne révisée, ces descripteurs sont extraits par le processeur sémantique « emotyc », ci-après noté EMOTYC.

Ce processeur que j'ai développé moi-même utilise un classifieur issu des expériences informatiques présentées chapitre 5. Les dix-neuf couples de descripteurs repérés par EMOTYC (à chaque fois, un nombre d'occurrences brut de phrase et une proportion de phrases au sein du texte) renvoient à des notions de la modélisation des émotions que je propose chapitre 3, plus précisément le mode d'expression, le type d'émotion et la catégorie émotionnelle. Les figures 2 et 3 montrent les résultats de l'application de EMOTYC, tels qu'ils sont visibles sur la plateforme en ligne[9], sur deux phrases ex-

8 Les neuf couples de descripteurs correspondent aux types d'adverbiaux suivants : 'adverbiaux_purement_iteratifs', 'adverbiaux_duratifs_iteratifs', 'adverbiaux_iterateur_calendaire', 'adverbiaux_iterateur_frequentiel', 'adverbiaux_iterateur_quantificationnel', 'adverbiaux_dur_iter_absolu', 'adverbiaux_dur_iter_anaphorique', 'adverbiaux_dur_iter_deictique', 'adverbiaux_dur_iter_relatif'.

9 http://vheborto-corliapi.inist.fr:8984/

traites de l'article[10] « Quelles sont les conséquences du mal-logement ? », *P'tit Libé*, n°91, 1 au 7 février 2019. Une phrase isolée étant traitée à chaque fois, seuls les descripteurs de type nombre brut sont reproduits ici. Sur ces deux figures, le descripteur « nombre_avec_emo » qui correspond au nombre de phrases émotionnelles du texte analysé a pour valeur 1 puisque l'unique phrase testée est à chaque fois émotionnelle. Pour la phrase de la figure 2, EMOTYC repère notamment la dénotation de la peur et l'usage de deux modes d'expression des émotions : le mode désigné (usage du lexique émotionnel) et le mode suggéré (description d'une situation conventionnellement associée à un ressenti émotionnel). Cela concorde avec la présence du terme « peur » et avec la description de l'état de la maison (« des grandes fissures dans les murs », « qu'elle s'écroule »).

Resultats:


```
{
  "emotyc_phrase": [
    {
      "sent_id": 0,
      "sentence": "Quand il y a des grandes fissures dans les murs d'une maison, ses habitants ont peur en permanence qu'elle s'écroule.",
      "nombre_avec_emo": 1,
      "nombre_avec_emo_comportementale": 0,
      "nombre_avec_emo_designee": 1,
      "nombre_avec_emo_montree": 0,
      "nombre_avec_emo_suggeree": 1,
      "nombre_avec_emo_base": 1,
      "nombre_avec_emo_complexe": 0,
      "nombre_avec_cat_autre": 0,
      "nombre_avec_cat_admiration": 0,
      "nombre_avec_cat_colere": 0,
      "nombre_avec_cat_culpabilite": 0,
      "nombre_avec_cat_degout": 0,
      "nombre_avec_cat_embarras": 0,
      "nombre_avec_cat_fierte": 0,
      "nombre_avec_cat_jalousie": 0,
      "nombre_avec_cat_joie": 0,
      "nombre_avec_cat_peur": 1,
      "nombre_avec_cat_surprise": 0,
      "nombre_avec_cat_tristesse": 0,
```


Figure 2: Descripteurs extraits par EMOTYC pour une phrase exprimant de la peur

Pour la phrase de la figure 3, EMOTYC trouve à nouveau les modes désigné et suggéré mais identifie cette fois de la tristesse. Cette analyse s'accorde avec la présence du verbe « souffrent » et avec la description de l'état des personnes impliquées dans la situation (« victimes du mal-logement », « graves problèmes psychologiques »).

10 Ce texte est bien entendu absent de l'ensemble ayant servi à l'entraînement du classifieur sur lequel repose EMOTYC. Notons qu'il est par ailleurs analysé au chapitre 7 pour mettre en avant le rôle des émotions dans la structuration textuelle.

Les sorties de EMOTYC prennent tout leur sens dans la suite du manuscrit, avec la présentation détaillée de la modélisation des émotions et du fonctionnement du classifieur sur lesquels reposent l'analyse faite par le processeur.

Resultats:


```
{
  "emotyc_phrase": [
    {
      "sent_id": 0,
      "sentence": "Les victimes du mal-logement souffrent aussi souvent de graves problèmes psychologiques.",
      "nombre_avec_emo": 1,
      "nombre_avec_emo_comportementale": 0,
      "nombre_avec_emo_designee": 1,
      "nombre_avec_emo_montree": 0,
      "nombre_avec_emo_suggeree": 1,
      "nombre_avec_emo_base": 1,
      "nombre_avec_emo_complexe": 0,
      "nombre_avec_cat_autre": 0,
      "nombre_avec_cat_admiration": 0,
      "nombre_avec_cat_colere": 0,
      "nombre_avec_cat_culpabilite": 0,
      "nombre_avec_cat_degout": 0,
      "nombre_avec_cat_embarras": 0,
      "nombre_avec_cat_fierte": 0,
      "nombre_avec_cat_jalousie": 0,
      "nombre_avec_cat_joie": 0,
      "nombre_avec_cat_peur": 0,
      "nombre_avec_cat_surprise": 0,
      "nombre_avec_cat_tristesse": 1,
```


Figure 3: Descripteurs extraits par EMOTYC pour une phrase exprimant de la tristesse

Il est à noter que la chaîne révisée (comme la chaîne TALN2022) comporte un second processeur sémantique d'analyse des émotions, déjà présent dans la chaîne initiale. Ce processeur nommé « émotions », ci-après noté ÉMOTIONS, s'appuie sur une version remaniée du lexique émotionnel EMOTAIX (Piolat & Bannour, 2009) pour identifier et quantifier les termes de 24 catégories émotionnelles et sur l'outil d'analyse de sentiments TextBlob pour attribuer un score de polarité au texte analysé, outil qui prend également appui sur un lexique. Le fonctionnement de ÉMOTIONS est présenté chapitre 6, lors d'une comparaison de ses performances avec le classifieur sous-jacent à EMOTYC. De cette comparaison émerge l'avantage majeur de EMOTYC : comme il n'utilise pas de lexique émotionnel figé, il retrouve beaucoup plus d'informations émotionnelles.

Principes de la refonte de la chaîne. L'enjeu de la refonte à laquelle j'ai contribué a été double. Il s'agissait d'une part d'identifier les descripteurs de la chaîne initiale devant être écartés car insuffisamment motivés du point de vue (psycho-)linguistique (par ex. la proportion de « stop words »). Les descripteurs conservés – soient ceux correspondant à des marqueurs linguistiques d'intérêt – ont parfois été réorganisés ou ont vu les ressources servant à leur repérage mises à jour et étoffées (par ex. connecteurs de type explicatifs et justificatifs).

D'autre part, il était nécessaire de déterminer quels nouveaux descripteurs intégrer parmi ceux issus de la traduction des marqueurs de complexité. Les nouveaux descripteurs extraits par la chaîne révisée relèvent notamment :

- de l'analyse lexicale (par ex. proportion de mots appartenant aux différents niveaux de l'échelle Dubois-Buyse (Ters et al., 1988)) ;
- de l'analyse de constructions syntaxiques particulières, réalisée à l'aide de patrons morphosyntaxiques et syntaxiques (par ex. proportion de tournures à la voix passive, proportion de subordonnées relatives typées selon la nature de leur introducteur) ;
- ou encore de l'analyse sémantique effectuée par exemple à partir de transducteurs (Teissedre, 2012) pour les adverbiaux temporels, de listes de métaphores (Blanchard et al., 2001) ou de mon modèle d'analyse fine des émotions (EMOTYC).

La chaîne révisée (comme la chaîne TALN2022) s'appuie sur la librairie open source STANZA, une chaîne neuronale permettant entre autres d'effectuer une analyse morphosyntaxique et syntaxique des textes (Qi et al., 2020). Le choix de cette librairie est motivé par deux principaux éléments : elle permet d'effectuer une analyse des textes dans une architecture de traitement homogène et également d'ajouter des attributs à chaque niveau d'analyse à différentes échelles (token, mot, phrase et document).

Les traitements opérés par la chaîne révisée s'appliquent à deux niveaux d'analyse : au niveau du document (par ex. la proportion de verbes au présent dans l'ensemble d'un texte) et au niveau de la phrase (par ex. la proportion de mots appartenant au vocabulaire de l'échelle Dubois-Buyse répartis par niveaux scolaires).

En l'état actuel, la chaîne révisée comporte 21 processeurs, répartis en six niveaux d'analyse linguistique et recouvrant 347 descripteurs. Parmi ces descripteurs, 214 sont d'ordre sémantique. 64,3 % traduisent un marqueur de complexité identifié dans la

littérature psycholinguistique, 46,2 % dans la littérature de lisibilité ou de simplification de textes[11] et 32 % un marqueur exploratoire, soit un marqueur intuitivement considéré comme pertinent pour l'analyse de la complexité mais non identifié dans la littérature psycholinguistique, en lisibilité ou en simplification (*cf.* section 1). Ces caractéristiques sont résumées dans le tableau 3. Le détail des descripteurs et de leur correspondance avec des marqueurs de complexité sera exposé dans la documentation qui accompagnera *in fine* la chaîne du projet TextToKids.

Au terme du projet, la chaîne comportera également un processeur visant à prédire directement le niveau de complexité d'un texte, en lui associant un âge cible. Cette prédiction sera assurée par un modèle appris automatiquement sur le corpus final de TextToKids, dans la lignée des modèles déjà proposés dans (Rahman et al., 2020) sur le corpus Coling2020 et de (Battistelli, Etienne, et al., 2022) sur le corpus TALN2022.

Niveau linguistique	**Nb de processeurs**	**Nb de descripteurs**	**Prop. (%) de descripteurs traduit d'un marqueur de complexité**		
			psycholing.	**de lisibi. / de simplifi.**	**exploratoire**
Graphie / Typographie	1	4	0	50	50
Lexique	3	17	11,8	17,6	70,6
Morphosyntaxe	4	66	3,0	89,4	7,6
Phonétique / Phonologie	1	7	0	100	0
Sémantique	8	214	90,2	6,5	3,3
Syntaxe	4	39	66,7	46,2	17,9
Total	***21***	***347***	***64,3***	***46,2***	***32,0***

Tableau 3: Répartition des processeurs et descripteurs selon le niveau linguistique

Les descripteurs de la chaîne révisée correspondent à des valeurs numériques (généralement nombres d'occurrences brutes ou proportions), comme cela était déjà le cas dans (Battistelli, Etienne, et al., 2022; Blandin et al., 2020). Il est à noter que depuis juin 2022, j'ai collaboré avec C. Parisse, R. Rahman et D. Battistelli au développement d'une chaîne parallèle, reprenant uniquement les processeurs sémantiques de la chaîne révisée. Cette chaîne parallèle, appelée « Sémantique » sur l'interface Web[12], extrait des descripteurs textuels et non plus numériques. Les descripteurs fondés sur l'identification d'adverbiaux temporels ne correspondent par exemple ainsi plus à de simples occurrences ou proportions mais à des listes de marqueurs repérés dans les

11 Certains marqueurs de complexité sont identifiés à la fois en psycholinguistique et en lisibilité ou simplification.
12 http://vheborto-corliapi.inist.fr:8984/

textes comme étant des adverbiaux. Cela permet un usage directement opérationnel pour des linguistes ou psycholinguistiques désirant connaître les marqueurs sémantiques pertinents identifiés par la chaîne (et non seulement leur nombre ou leur proportion dans un texte). La chaîne « Sémantique » permet l'extraction de descripteurs textuels pour les processeurs : « adverbiaux temporels », « émotions », « emotyc », « entités nommées », « modalité », « métaphores », « connecteurs organisateurs » et « propositions subordonnées ».

2.3) Étude de corrélations entre âges éditeurs et descripteurs linguistiques de complexité des textes

La chaîne TALN2022 a été appliquée sur le corpus TALN2022. À partir des analyses ainsi effectuées, j'ai participé à l'identification de corrélations entre âges éditeurs et descripteurs linguistiques.

Corpus. Les expérimentations présentées dans (Battistelli, Etienne, et al., 2022) ont été menées sur le corpus TALN2022, correspondant à une sous-partie du corpus Coling2020, dont seules les œuvres de fiction publiées par des éditeurs professionnels ont été retenues[13]. L'objectif était d'obtenir un corpus homogène en genre, afin de limiter les effets de genre sur la lisibilité soulignés dans (Sheehan et al., 2008) ou encore dans (Flor et al., 2013).

Ces expérimentations incluent la mesure de corrélations entre âges éditeurs et descripteurs linguistiques, ainsi que l'entraînement d'un modèle de prédiction de l'âge à associer à un texte. Afin de limiter le biais sur la longueur des textes pour l'apprentissage des modèles de prédiction de l'âge cible (par ex. association des textes longs à un âge élevé), les textes de plus de 10 000 caractères ont été découpés en segments d'environ 5 000 caractères, en respectant les frontières de paragraphes. Le corpus se compose ainsi de 1 130 textes de fiction (env. 70,8K phrases), correspondant soit à une portion d'ouvrage, soit à un ouvrage entier. Chaque texte (entier ou portion) est associé à une tranche d'âge, selon les recommandations des éditeurs. Dix-sept tranches d'âges, de 0-3 ans à 14-18 ans, sont représentées. Le corpus a été partitionné de manière aléatoire

13 Le corpus Coling2020 intègre, parmi ses sources de textes de littérature jeunesse, des livres numériques diffusés librement sur le site https://www.ebookids.com/. Ces ouvrages paraissaient dans un premier temps intéressants de part leur facilité d'accès et la diversité des auteur·e·s, thématiques et tranches d'âges qu'ils proposent. Il a finalement été décidé d'exclure ces textes du corpus TALN2022 et de la version révisée finale du corpus TextToKids car ils n'émanent pas d'auteur·e·s et de maisons d'édition professionnel·le·s.

en ensembles d'entraînement, de développement et de test, avec pour seule contrainte le fait que tous les textes issus d'un même ouvrage devaient se trouver dans le même ensemble. Les mesures de corrélation ont été effectuées sur l'ensemble de test uniquement. Les caractéristiques du corpus TALN2022 sont résumées au tableau 4.

	Nb textes	**Nb phrases**	**Tranches d'âges éditeur**
Train	746	41 672	0-2, 0-3, 2-7, 3-5, 5-7, 6-8, 7-11, 8-10, 8-11, 8-13, 10-12, 10-13, 10-14, 11-13, 12-14, 14-18
Dev	196	10 888	0-3, 3-5, 4-6, 6-8, 8-11, 10-12, 14-18
Test	188	11 096	0-3, 2-4, 3-5, 4-6, 6-8, 7-11, 8-11, 10-12, 10-14, 14-18

Tableau 4: Caractéristiques du corpus TALN2022 – Tableau repris de (Battistelli, Etienne, et al., 2022)

Expérimentation : mesures de corrélations entre descripteurs linguistiques et âges éditeurs. La chaîne TALN2022 a été appliquée sur le corpus TALN2022 afin de tester la pertinence des recommandations d'âge faites par les éditeurs. Plus précisément, les corrélations entre les descripteurs linguistiques extraits par la chaîne et les âges donnés par les éditeurs ont été étudiées. Les âges éditeurs ont ainsi été mis en regard avec les normes développementales issues de travaux psycholinguistiques. Il s'agit alors d'observer l'évolution des corrélations entre descripteurs linguistiques et âge éditeur au fil des âges. Des pics de corrélation, particulièrement saillants pour certains descripteurs, apparaissent alors autour de certains âges. Dans (Battistelli, Etienne, et al., 2022), ces pics interviennent pour identifier des « âges frontières » de la présence de certains descripteurs. L'hypothèse posée est que, si les âges éditeurs sont fixés de manière adéquate, ces âges frontières correspondent aux étapes développementales dégagées dans la littérature psycholinguistique sur la compréhension.

L'évolution des corrélations au fil de l'âge éditeur sont présentées figure 4 pour un descripteur syntaxique (nombre moyen de mots), un descripteur morphosyntaxique (conjonctions de subordination) et quatre descripteurs sémantiques (embarras, joie, explicatifs et justificatifs, pointage absolu).

Pour le descripteur syntaxique « nombre moyen de mots », la corrélation mesurée augmente fortement à partir de 10 ans. Ce descripteur, qui se rapporte au nombre moyen de mots par phrase dans le texte, est en lien avec un critère très fréquemment employé en lisibilité : la longueur de la phrase, sachant qu'une phrase plus longue est jugée plus complexe (Elguendouze, 2020). Le pic positif de corrélation constaté avec l'âge éditeur le plus élevé de notre corpus convient donc aux attentes pour ce descrip-

teur. Le descripteur morphosyntaxique « conjonctions de subordination » (pourcentage des conjonctions sur le nombre total de mots) est inspiré de travaux de lisibilité qui observent les proportions de différentes parties du discours (François & Fairon, 2012). Si ce descripteur n'est pas associé en tant que tel à une norme développementale, il présente dans le corpus TALN2022 un pic de corrélation positive aux alentours de 7 ans. Ce pic indique que ce critère permet de classer plus efficacement les textes autour des âges frontière 6-8 ans.

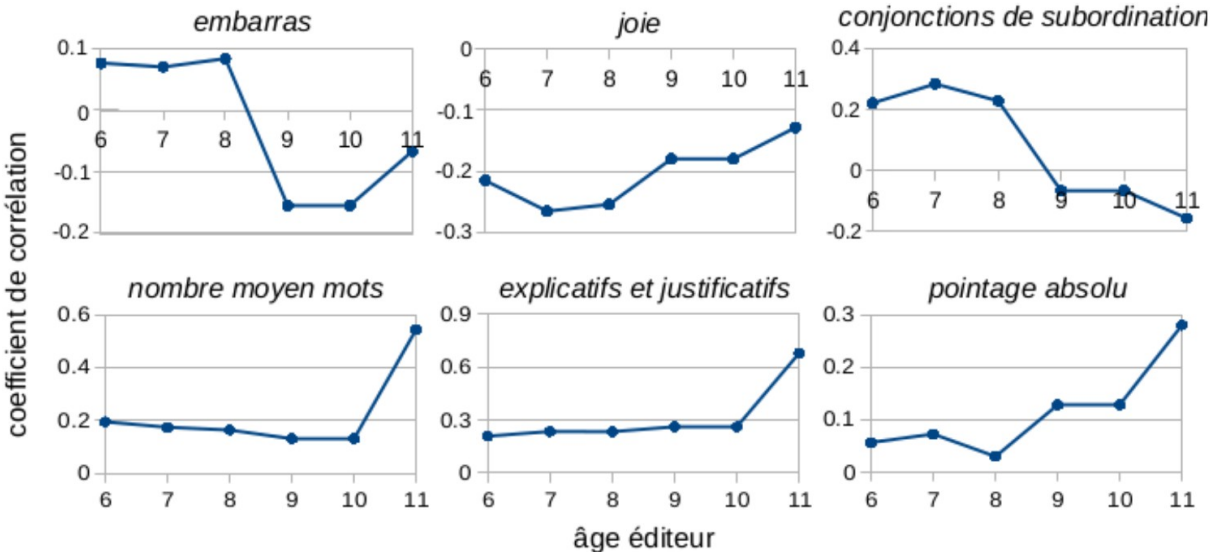


Figure 4: Évolution de la corrélation entre plusieurs descripteurs et l'âge éditeur – Figure reprise de (Battistelli, Etienne, et al., 2022)

Les descripteurs sémantiques « joie » et « embarras » issus de ÉMOTIONS renvoient à la proportion de termes appartenant au lexique de ces deux émotions (*cf.* section 2.2). Ces deux descripteurs rendent compte de la norme développementale psycholinguistique, selon laquelle les émotions de base comme la joie sont mieux comprises que les émotions complexes comme l'embarras, et ce jusqu'à 10-11 ans (*cf.* section 1.2). L'évolution de la corrélation du descripteur « joie » illustrée figure 4 montre une présence plus prononcée de cette émotion dans les textes pour les enfants plus jeunes (8 ans et moins), ce qui concorde avec les attentes de la norme développementale. Le descripteur embarras apparaît également plus corrélé avec les âges éditeurs inférieurs à 9 ans. Il y a donc cette fois une divergence entre les recommandations éditeurs et celles pouvant être dégagées de la littérature psycholinguistique.

L'évolution des corrélations pour le descripteur « explicatifs et justificatifs » (proportion de connecteurs de type *car*, *puisque*, *voilà*, *pourquoi*) comporte également un pic positif vers 10 ans. Des travaux en simplification recourent à ce descripteur pour traduire le fait qu'un lien logique non explicite est plus complexe qu'un lien logique explicite (Gala et al., 2018). Une association plus forte entre ce descripteur et les jeunes âges est donc attendue mais le contraire est observé, ce qui reflète une difficulté potentielle de compréhension sur ce critère dans le corpus TALN2022.

Enfin, le descripteur « pointage absolu », qui correspond dans cette version de la chaîne[14] à la proportion d'adverbiaux temporels calendaires absolus (par ex. *au cours du 19è siècle*) rapporté au nombre de mots du texte, présente quant à lui un pic positif à partir des âges éditeur 9-10 ans. Cela indique que ce descripteur est opératoire pour identifier les textes considérés comme adaptés aux enfants de 9 ans – et plus – par les éditeurs. Cette observation est en adéquation avec les savoirs psycholinguistiques concernant la compréhension des notions calendaires (*cf.* section 1.2)

Parmi les exemples de descripteurs analysés, les corrélations observées avec les âges éditeurs concordent globalement avec les attentes correspondant aux normes développementales issues de la littérature psycholinguistique. Malgré les quelques divergences constatées (descripteurs « embarras » et « explicatifs et justificatifs »), l'analyse de ces corrélations tend ainsi à valider la pertinence de la catégorisation en âges par les éditeurs. Cette analyse a donc permis de justifier *a posteriori*, la mobilisation d'un tel corpus pour entraîner à partir des âges éditeurs un modèle de prédiction de l'âge cible d'un texte comme cela a été proposé dans (Rahman et al., 2020) sur le corpus Coling2020 et dans (Battistelli, Etienne, et al., 2022) sur le corpus TALN2022.

3) Exploration expérimentale de la notion de complexité de textes

Si des étapes développementales de la compréhension de textes (ou tout au moins relativement à certains marqueurs linguistiques en particulier) se dégagent déjà de la littérature psycholinguistique (*cf.* section 1.2), l'étude de cette question reste encore largement à approfondir. Des explorations auprès d'enfants jeunes lecteurs ont ainsi été prévues dans le projet TextToKids, afin de confirmer certaines hypothèses sur les

14 Dans la version actuelle de la chaîne disponible en ligne, ce descripteur correspond désormais à la proportion d'adverbiaux temporels de pointage absolu par rapport au nombre d'adverbiaux temporels identifiés dans le texte.

processus de compréhension et de tenter d'en mettre au jour de nouveaux éléments. Ces explorations ont pris la forme d'expériences psycholinguistiques menées dans des écoles auprès d'enfants de primaire. Les expériences sont plus particulièrement axées sur trois facteurs influençant la compréhension d'un texte : le genre de texte, sa dimension émotionnelle et sa dimension temporelle. Ces deux dimensions sémantiques se sont imposées comme objets d'étude de choix au sein du projet de fait de leur rôle important déjà signalé par la littérature psycholinguistique dans le processus de compréhension de textes par les enfants (*cf.* section 1.2), et ce même si elles représentent chacune un véritable défi pour l'analyse linguistique et pour le TAL.

Un protocole expérimental visant à évaluer l'impact de ces trois facteurs sur la compréhension de textes par les enfants a donc été élaboré (section 3.1). Même si l'ensemble des expériences prévues dans le cadre du projet n'ont pas encore été conduites, quelques grandes tendances se dégagent cependant déjà d'une série d'expériences menée en juin 2021 sur des enfants de CE2, de CM1 et de CM2 (section 3.2).

3.1) Un protocole expérimental pour évaluer l'impact du genre de texte et des dimensions émotionnelle et temporelle sur la compréhension

Afin d'évaluer l'impact des dimensions émotionnelle et temporelle ainsi que du genre du texte sur la compréhension des textes par les enfants, plusieurs expériences impliquant le même protocole expérimental ont été imaginées.

Protocole expérimental. La méthodologie retenue consiste à sonder la compréhension à partir de textes réels[15], tel que cela a déjà pu être mis en œuvre dans des travaux comme (Blanc, 2010; Creissen & Blanc, 2017). Afin d'étudier l'impact du genre de texte, neufs documents sont soumis à chaque enfant évalué : trois textes encyclopédiques, trois fictionnels et trois journalistiques. Un ensemble d'items visant à tester la compréhension est élaboré pour chaque texte à partir d'une analyse linguistique fine des dimensions émotionnelle et temporelle. Les textes ainsi que les items qui leur sont associés sont présentés sous la forme d'un livret individuel (reproduit en Annexe I), remis à chaque enfant. La passation des expériences s'effectue en classe, lors de trois séances pendant lesquelles les enfants sont confrontés à trois textes (un de chaque

15 Par opposition à des textes construits spécifiquement pour répondre aux besoins de l'expérience.

genre). Pour limiter le biais des capacités de lecture, chaque texte est accompagné d'un fichier audio, correspondant à la lecture du texte sur un ton neutre. Les sujets devaient donc lire et écouter le texte, à deux reprises, avant de répondre aux items sans nouvel accès au texte.

Choix des textes. Pour faciliter la passation des expériences, chaque texte se devait d'être entièrement contenu sur une page de format A4, en police Arial 14 (police adaptée aux jeunes lecteurs) et donc comporter 300 à 400 mots. Le matériel expérimental a été choisi parmi la grande quantité de textes rassemblés lors la constitution des corpus du projet (*cf.* section 2.1). Après avoir découpé cet ensemble en portions de 300-400 mots, le prototype de la chaîne d'extraction de descripteurs de complexité présentée dans (Blandin et al., 2020) (la chaîne initiale) a été appliqué. Une sélection semi-automatique des textes a ainsi été effectuée sur deux critères : 1) l'âge cible prédit[16] et 2) le score de polarité du texte, obtenu grâce au processeur ÉMOTIONS de la chaîne TextToKids (*cf.* section 2.2).

Seules quelques dizaines de textes ont ainsi été conservées. Une analyse linguistique focalisée sur les dimensions émotionnelle et temporelle a été menée pour juger de la pertinence des textes par rapport aux objectifs des expérimentations. L'analyse en émotions a été effectuée grâce à la modélisation (c'est-à-dire le schéma d'annotation) que je présente chapitre 3, permettant ainsi de porter une attention toute particulière à la façon dont les émotions sont exprimées dans les textes, par exemple explicitement par le biais de termes du lexique émotionnel ou implicitement par la description de situations évoquant une émotion. Les différents adverbiaux temporels mobilisés dans les textes pour ancrer des évènements sur l'axe temporel ou les ordonnancer entre eux ont été relevés, ainsi que différents marqueurs de modalité. À l'issue de cette analyse, l'échantillon sélectionné s'est avéré insatisfaisant pour étudier l'impact des émotions et de la temporalité sur la compréhension de texte. De nouveaux textes ont donc été ajoutés manuellement, d'après leurs caractéristiques émotionnelles et temporelles.

Les analyses linguistiques ont porté sur un total de 31 textes, de sources variées :

- 11 textes encyclopédiques extraits de l'encyclopédie en ligne Vikidia ou de la revue *Histoires vraies* ;

16 Le prototype de (Blandin et al., 2020) intégrait un processeur fondé sur un modèle capable de prédire automatiquement l'âge cible d'un texte (modèle entraîné à partir d'un corpus de textes classés en âges éditeurs, *cf.* section 2). À terme, un processeur de ce type sera également proposé dans la chaîne disponible en ligne (http://vheborto-corliapi.inist.fr:8984/).

- 11 textes fictionnels extraits d'ouvrages de littérature jeunesse, écrits par F. Rabelais, A. Laroche, H. Bichonnier, J. K. Rowling, M. Aymé, C. Gutman, M.-A. Murail, D. Pennac ou M. Desplechin ;
- 9 textes journalistiques issus des journaux *Albert* ou *P'tit Libé*.

Le corpus final des expériences ne comporte que 9 textes et ses caractéristiques sont présentées dans le tableau 5.

Nom du sous-corpus	**Genre**	**Sources**	**Nb de textes**	**Nb de mots**	**Nb moyen de mots/texte**
WP7Encyclo	Encyclopédique	Vikidia	3	1 153	384,3
WP7Fiction	Fictionnel	*Dragonne de minuit*, A. Laroche ; *La bande à Tristan*, M.-A. Murail ; *Verte*, M. Desplechin	3	1 084	361,3
WP7Journa	Journalistique	*Albert*, *P'titLibé*	3	1 129	376,3
Total			**9**	**3 366**	**374,0**

Tableau 5: Caractéristiques du corpus WP7

Choix des items. L'élaboration du matériel expérimental repose sur la constitution, pour chaque texte sélectionné, d'items visant à évaluer la compréhension. Ces items prennent ici la forme d'affirmations, à propos desquelles les enfants doivent donner leur avis en répondant par vrai ou par faux. Pour chaque texte, dix items ont été construits en veillant à avoir des proportions similaires d'affirmations vraies et d'affirmations fausses. Sur ces dix items quatre portent sur la dimension temporelle, quatre sur la dimension émotionnelle et deux sur le contenu global du texte. Lorsque cela était possible, les affirmations temporelles ont été équilibrées pour sonder des informations de type ancrage temporel et de type ordonnancement. De même, les affirmations émotionnelles ont été équilibrées pour évaluer des informations émotionnelles implicites et explicites. Les affirmations sont formulées en reprenant autant que possible les mots du texte et en suivant des patrons d'énoncés typiques (par ex. « quand il se passe X, l'individu se sent Y », « X a eu lieu après Y », *etc.*). Un onzième item de type question à choix multiple a été ajouté afin de demander à l'enfant quel était, selon lui, le genre du texte (roman, journal d'information, recette de cuisine, *etc.*).

Choix de la population. Les expériences ont été menées dans des classes autonomes en lecture, soient des CE2, des CM1 et des CM2 pour la première expérimentation. La seconde s'est focalisée sur des CE2 et des CM2 pour mieux mettre en évidence les différences développementales entre classes d'âges.

Réplicabilité et comparabilité. Le principe de réplication des résultats se situe au cœur de la démarche expérimentale de l'évaluation des compétences de compréhension des enfants. Une première expérience a donc été mise en œuvre afin de confirmer et de préciser des résultats obtenus dans des études psycholinguistiques antérieures. Ces études soulignent l'importance des dimensions temporelles et émotionnelles des textes dans le processus de compréhension d'une part (voir par ex. (Blanc, 2010)) et l'impact du genre de texte d'autre part (voir par ex. (Aboud et al., 2019; Baretta et al., 2009)). Après analyse des résultats de cette première passation, l'expérience a été réitérée de façon quasi-identique auprès d'autres enfants, afin de s'assurer que les observations effectuées sont régulières, que le patron de résultat qui se dégage est suffisamment robuste pour tirer des généralisations sur l'évolution des compétences de compréhension. Une dernière expérience, ayant recours au même matériel expérimental que les deux premières, est prévue afin de mettre au jour des processus de simplification efficaces, c'est-à-dire permettant d'améliorer la compréhension des textes par les enfants.

3.2) Résultats préliminaires sur l'évaluation des compétences de compréhension d'enfants de CE2, de CM1 et de CM2

À partir du matériel expérimental constitué, concrétisé sous la forme d'un livret accompagné d'enregistrements audios, une première vague d'expérience a été menée en juin 2021 dans des écoles de Marseille et de Montpellier auprès de 103 enfants de CE2, 78 de CM1 et de 64 CM2. Après nettoyage des résultats (suppression des sujets n'ayant pas rempli l'intégralité du livret, de ceux dont les résultats diffèrent de plus de deux écarts-types des camarades de leur classe d'âge), seuls 71 sujets de CE2, 70 de CM1 et 54 de CM2 sont conservés. Les détails des résultats sont encore à paraître[17]. Il reste que je peux d'ores et déjà exposer ici de grandes tendances qui s'en dégagent. De manière générale, il apparaît ainsi que la compréhension des textes – tous genres confondus – s'améliore avec l'âge, comme le montre la distribution des pourcentages de réussite selon le niveau scolaire de la figure 5.

Les enfants répondent globalement mieux aux questions émotionnelles qu'aux questions temporelles (*cf.* figure 6) mais cet effet pourrait être pondéré par le genre de

17 Une communication intitulée « Implication des dimensions temporelles et émotionnelles dans la compréhension de textes narratifs, journalistiques et encyclopédiques chez l'enfant » a été acceptée au colloque international RIPSYDEVE qui se tiendra les 26 et 27 juin 2023.

texte qui, de manière générale, semble influencer le degré de compréhension. La distribution des pourcentages de bonnes réponses selon le genre de texte illustrée figure 7 indique ainsi des performances (légèrement) moindre sur le genre fictionnel.

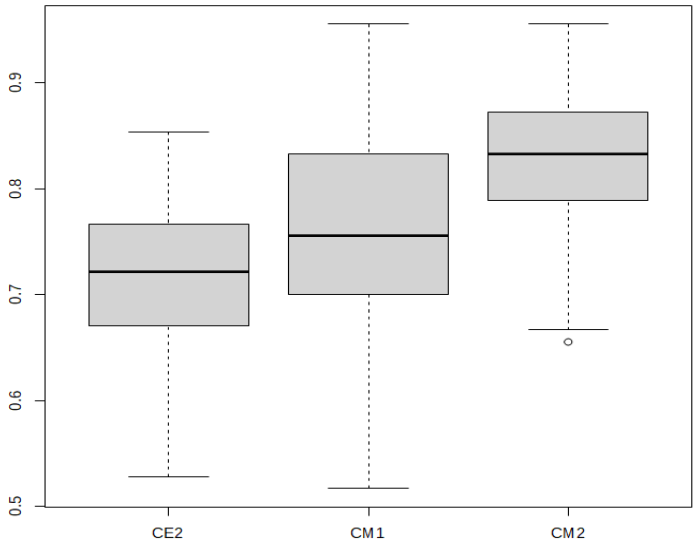


Figure 5: Distribution des pourcentages de bonnes réponses selon le niveau scolaire, tous types de questions et tous textes confondus – Expérience 1

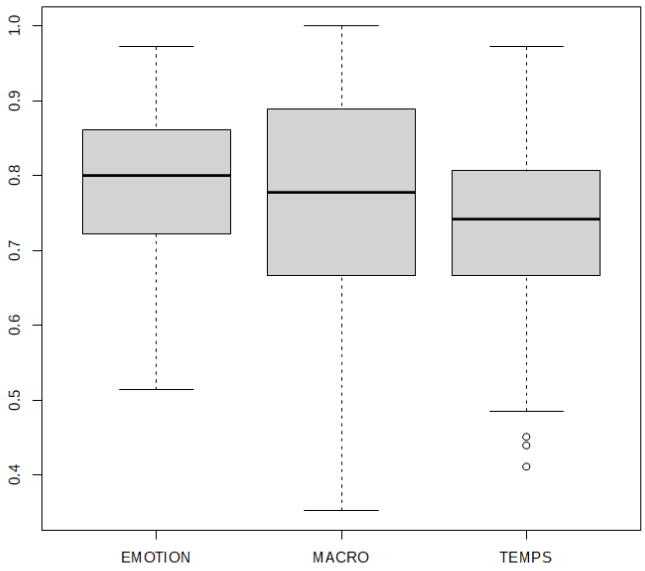


Figure 6: Distribution des proportions de bonnes réponses selon le type de question, tous âges et tous textes confondus – Expérience 1

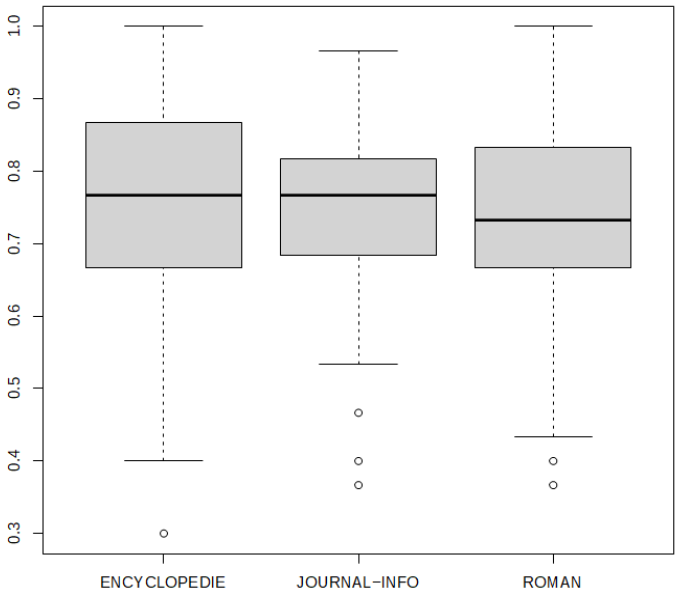


Figure 7: Distribution des pourcentages de bonnes réponses selon le genre de texte, tous âges et types de question confondus – Expérience 1

Une expérience de réplication a été menée courant janvier 2023 auprès de trois classes de CE2 et de trois classes de CM2 uniquement. Le matériel expérimental était quasiment identique à celui de la première expérience et ne présentait que quelques modifications pour certains items dont les résultats s'étaient avérés inexploitables (items massivement réussis ou au contraire échoués, toutes classes d'âges confondues). Les résultats de cette expérience de réplication sont toujours en cours de traitement.

Une troisième et dernière expérience est envisagée et devra avoir lieu d'ici la fin du mois de juin 2023. Elle a pour ambition de consolider les résultats des deux premières expériences en se focalisant uniquement sur les trois textes fictionnels du livret et en demandant aux enfants de CE2, CM1 et CM2 d'évaluer à chaque fois leur confiance en la certitude de leur réponse aux dix questions associées au texte. Parallèlement, le même type de protocole sera testé auprès d'adultes. Les résultats des expériences d'évaluation de la compréhension devraient par ailleurs être mis en regard avec l'analyse effectuée par la chaîne d'extraction de descripteurs de complexité, dans la perspective de dégager des corrélations entre descripteurs et résultats des expériences.

L'ultime objectif des expériences menées dans le cadre du WP7 est d'établir une hiérarchie entre les différents facteurs influençant la compréhension sondés et d'en tirer profit pour dégager des préconisations de rédaction (par ex. selon l'âge ciblé et le genre de texte traité, la dimension émotionnelle est-elle à privilégier ?). Ces préconisations pourraient être complétées par les éventuelles corrélations dégagées entre descripteurs de complexité et résultats des expériences.

Conclusion

Dans le projet TextToKids, la notion de complexité de textes pour le public des enfants jeunes lecteurs est étudiée en lien avec la notion de compréhension de textes.

De ce fait, de nombreuses études psycholinguistiques ont été passées en revue afin d'identifier des marqueurs linguistiques reflétant des étapes développementales de l'évolution des compétences de compréhension. Pour offrir une analyse plus complète de la complexité des textes, des marqueurs issus des travaux de lisibilité et de simplification ont également été retenus et quelques marqueurs exploratoires – c'est-à-dire intuitivement considérés comme pertinents pour l'analyse de la complexité – ont été ajoutés. 64 marqueurs de complexité ont ainsi été identifiés (17 issus de travaux de lisibilité ou de simplification, 39 de travaux psycholinguistiques et 8 exploratoires).

58 de ces marqueurs dont l'extraction automatique pouvait relativement aisément être mise en œuvre ont été traduits en *descripteurs* de complexité au sein d'une chaîne de traitement automatique des textes. Organisée en modules, les « processeurs », gérant chacun l'extraction d'un ensemble de descripteurs, cette chaîne vise à objectiver l'analyse de la complexité des textes. Elle intègre à ce jour 347 descripteurs, dont 223 sont motivés sur le plan psycholinguistique et reflètent des étapes développementales de la compréhension chez l'enfant. La spécificité de la chaîne par rapport à des travaux de lisibilité ou de simplification est d'une part de s'efforcer à motiver sur le plan psycholinguistique les descripteurs et d'autre part d'intégrer un grand nombre de descripteurs sémantiques (214 descripteurs sémantiques au total sont extraits par la chaîne). Appliquée à un corpus de textes fictionnels classés en âges éditeurs, elle a permis de mettre en regard ces âges avec les étapes développementales dont rendent compte les descripteurs de complexité. La pertinence des recommandations d'âges proposées par les éditeurs a ainsi été globalement confirmée.

En s'appuyant sur le corpus général de textes classés en âges éditeurs constitué pour le projet TextToKids (corpus composé de textes encyclopédiques, fictionnels et journalistiques) et sur les marqueurs de complexité émotionnels et temporels identifiés en psycholinguistique, un protocole expérimental a été élaboré. Testé auprès d'enfants de CE2, de CM1 et de CM2, ce protocole a permis de faire émerger l'impact du genre de texte sur les processus de compréhension. L'apport majeur de cette première expérience a cependant été de souligner l'intérêt de prendre en compte deux dimensions sémantiques, la temporalité et les émotions des personnages, pour appréhender la question de la compréhension (par ex. accès facilité au contenu émotionnel par rapport au contenu temporel). Afin de répliquer et ainsi de confirmer ces tendances, une seconde expérience, dont les résultats sont en cours de traitement par les membres du WP7, a été menée. Une troisième et dernière expérience viendra approfondir les résultats des deux premières en étudiant uniquement les textes fictionnels.

La suite du manuscrit se focalise désormais sur les émotions des personnages, avec pour objectif de déterminer comment aborder leur analyse pour concilier des visées à la fois psycholinguistique, linguistique et TAL. Des pistes pour l'exploration des interactions entre les émotions et la temporalité seront également proposées.

Dans le contexte générale de l'étude de la complexité des textes, telle qu'elle est abordée dans le projet TextToKids, la contribution majeure de mon travail de thèse a été le développement du processeur sémantique EMOTYC, testable en ligne (http://vheborto-corliapi.inist.fr:8984/chain/) et dont l'analyse se distingue clairement de celle du processeur ÉMOTIONS, présent dès la version initiale de la chaîne TextToKids. Les différences entre ces deux processeurs sont brièvement décrites à la section 2.2 mais une comparaison plus approfondie des analyses qu'ils opèrent est menée au chapitre 6.

Chapitre 2 Perspectives psycholinguistiques, linguistiques et TAL dans l'analyse des émotions

Pour les enjeux qu'elle pose, linguistiques d'abord, informatiques ensuite et pour son importance dans la compréhension de textes par les enfants, la notion d'émotion occupe une place importante au sein du projet TextToKids. Très difficile à circonscrire de par la diversité des moyens linguistiques employés pour l'exprimer et des catégories sémantiques qu'elle recoupe, la façon de la traiter varie en outre selon le domaine qui l'étudie (la linguistique, la psycholinguistique et le TAL, pour parler de ceux auxquels nous nous référons ici principalement). Le présent chapitre est consacré à la présentation des choix théoriques et méthodologiques que nous avons opérés pour aborder la notion d'émotion.

Après un bref rappel des notions clés mobilisées en psycholinguistique pour analyser les émotions (section 1), celles utilisées dans diverses études linguistiques seront ensuite exposées (section 2). La question de l'analyse automatique des émotions et des fondamentaux théoriques qui la sous-tendent seront également évoqués (section 3) avant de présenter de manière synthétique l'ensemble notionnel retenu pour caractériser les émotions dans mon approche (section 4).

1) Les émotions en psycholinguistique

Les études psycholinguistiques mêlant analyse des émotions et compréhension de textes mettent l'accent sur la façon dont les émotions sont *exprimées* dans les textes et pas seulement sur les *catégories* émotionnelles.

Elles sont alors abordées à travers leur relation avec la compréhension de textes, définie comme la capacité du lecteur à se construire une représentation mentale de la situation décrite par le texte (Zwaan & Radvansky, 1998). En intervenant aux côtés d'autres dimensions sémantiques telles que le temps ou l'espace pour construire cette représentation, les émotions des personnages – par opposition à celles du lecteur (Dijkstra et al., 1995) – représentent alors un élément clé qui doit être pris en compte pour étudier la compréhension de textes. Un état de l'art détaillé du rôle des émo-

tions dans la compréhension de textes chez les enfants et de l'évolution de la compréhension des émotions elles-mêmes est exposé chapitre 1.

Si l'analyse des émotions en lien avec la compréhension de textes a le plus souvent recours aux *catégories émotionnelles* (par ex. joie, peur, *etc.*), certains travaux comme (Blanc & Quenette, 2017; Davidson, 2006) emploient également la notion de *type d'émotion* (de base ou complexe).

Les segments textuels exprimant une émotion peuvent par ailleurs être reliés à d'autres segments textuels grâce à des relations sémantiques. De ce point de vue, (Blanc, 2010) suggère une relation forte entre la causalité et les émotions des personnages dans la construction de la représentation mentale de la situation décrite par un texte. Les concepts de *cause* et de *conséquence* de l'émotion prennent ainsi part à la description de l'expression des émotions.

Même si peu d'études psycholinguistiques considèrent la diversité linguistique des façons d'exprimer les émotions, elles représentent une contribution cruciale pour les travaux visant une analyse fine des émotions dans les textes. De ce point de vue, (Blanc, 2010) et (Creissen & Blanc, 2017) distinguent trois manières d'exprimer les émotions[18] : les labels émotionnels (par ex. apeuré, heureux), les descriptions de comportements émotionnels (par ex. pleurer, rire) et des descriptions de situations associées conventionnellement à un ressenti émotionnel (par ex. le décès de quelqu'un peut conventionnellement être associé à de la peur ou à de la tristesse). Ce travail ne propose cependant pas une caractérisation linguistique détaillée de ces différents modes d'expression des émotions.

Par ailleurs, plusieurs lexiques émotionnels issus du domaine de la psychologie seront présentés dans la section 3.2.

2) Les émotions en linguistique

Des différentes approches employées en linguistique pour analyser les émotions émergent trois concepts clés : la catégorie émotionnelle, le mode d'expression de l'émotion et l'expérienceur de l'émotion (section 2.1). Si elles ne s'y intéressent pas explicitement, ces approches laissent transparaître l'intérêt d'observer les interactions entre les émotions et d'autres dimensions sémantiques, en particulier la temporalité

18 Le chapitre 3 propose une définition plus détaillée de ces façons d'exprimer les émotions.

linguistique, pour mieux cerner le fonctionnement de l'expression des émotions (section 2.2).

2.1) Les concepts clés de la description linguistique des émotions

La prise en compte de tous les moyens linguistiques exprimant les émotions représente un véritable défi pour les linguistes. C. Kerbrat-Orecchioni souligne à cet égard la « fantastique diversité des moyens que peut investir le langage émotionnel, puisque tout mot, toute construction peuvent venir en contexte propice se charger d'une connotation affective » (Kerbrat-Orecchioni, 2000, p. 50). L'étude de l'expression des émotions nécessite de fait de prendre en compte des observables très nombreux et hétérogènes et ce même s'il devient alors « difficile d'associer [aux émotions] un (ou plusieurs) lieu(x) stable(s) dans le système linguistique » (Micheli et al., 2013, p. 9).

La plupart des travaux linguistiques tendent ainsi à se focaliser sur des aspects restreints de la dénotation des émotions. Lorsqu'un panel de catégories d'émotions (ex. joie, peur, colère, embarras, culpabilité) est traité, les travaux se concentrent généralement sur l'étude du lexique émotionnel (voir par exemple le lexique émotionnel constitué à l'aide de critères combinatoires dans (Diwersy et al., 2014)) et même parfois sur l'étude du lexique verbal seulement (voir par exemple l'étude des verbes dits de sentiment proposée dans (Mathieu, 2005)). Les analyses fines de marqueurs ne relevant pas du lexique émotionnel strictement se limitent quant à elles le plus souvent à l'examen de quelques structures ou de quelques catégories émotionnelles. Ce phénomène est illustré pour le français par exemple par l'étude de syntagmes ayant pour prédicat sémantique « peur » dans (Dobrovolski & Bresson, 1995), par l'analyse de la structure « être en colère » vs « *être en peur » dans (Leeman-Bouix, 1995), ou encore par l'étude de diverses structures (morpho-)syntaxiques exprimant la joie et la rage dans (Gross, 1995).

À ma connaissance, seul (Micheli, 2014) développe pour le français une approche des émotions incluant non seulement une grande hétérogénéité de marqueurs linguistiques (c'est-à-dire non strictement lexicaux) pouvant les exprimer mais aussi visant à proposer une typologie des différentes façons d'exprimer les émotions. À ce jour, (Micheli, 2014) constitue par ailleurs un des rares travaux proposant de dépasser le cadre strict de la phrase pour étudier l'expression des émotions. Il analyse par exemple l'expression de la peur dans un extrait du roman *Bel Ami*, de G. de Maupassant dans le

but d'observer les interactions entre des marqueurs des émotions pour certains situés dans des phrases différentes. Il relève à ce titre un ensemble de phrases comportant des termes du lexique émotionnel de la peur qui « disent » l'émotion (« inquiétait », « épouvantable », « peur », « terreur », *ibid.*, p. 137). La désignation directe de la peur est complétée par l'évocation d'éléments opérant ce que R. Micheli nomme une « schématisation » de la situation qui sert « d'étayage » (terme lui aussi introduit par R. Micheli) à la dénotation de l'émotion (par ex. évocation de la brièveté de la vie qui dure « Quelques aurores, voilà tout » et du caractère inévitable de la mort, *ibid.*, p. 141). Cet étayage met en mots « les raisons pour lesquelles il convient d'éprouver les émotions » désignées dans le texte (*ibid.,* p. 143). Enfin, un ensemble de marqueurs syntaxiques (ex. énoncé exclamatif, dislocation à droite, énoncé elliptique – *ibid.*, pp. 144-145) et textuels (organisation périodique des énoncés du texte, *ibid.*,pp. 146-147) interviennent à leur tour pour « montrer » (*ibid.*, p. 146) que le personnage ressent de la peur. L'analyse proposée dans (Micheli, 2014) illustre la convergence des différents marqueurs linguistiques situés dans différentes phrases pour l'expression d'une même émotion. Une définition plus détaillée des trois modes d'expression qu'il distingue – le dit, l'étayé et le montré – est introduite chapitre 3.

En plus de la *catégorie émotionnelle* et du *mode d'expression*, les travaux linguistiques caractérisent également souvent l'expression des émotions en étudiant la dénotation de l'*expérienceur* de l'émotion, soit l'entité (humain, animal, *etc.*) qui ressent l'émotion. Cette notion intervient par exemple dans l'analyse des verbes dits de sentiment de Y. Mathieu (Mathieu, 2005) et dans celle des expressions de peur par D. Dobrovolski et D. Bresson (1995). Elle intervient également dans la définition des modes d'expression par R. Micheli (Micheli, 2014), qui utilise la capacité d'une émotion à être plus ou moins facilement auto-attribuée (c'est-à-dire attribuée à l'énonciateur) ou allo-attribuée (c'est-à-dire attribuée à une entité autre que l'énonciateur) pour distinguer entre eux les différents modes d'expression des émotions.

2.2) Interactions entre émotions et d'autres champs sémantiques : le cas de la temporalité linguistique

Si les interactions entre les deux champs sémantiques que sont les émotions et la temporalité n'ont été que peu – voire pas – explicitement étudiées, divers travaux linguistiques fournissent des pistes de réflexion méritant d'être explorées plus en avant. De fait, ce sont des éléments d'interaction que j'ai été amenée à repérer dans l'analyse de

textes. La temporalité est entendue ici au sens large avec une approche telle que celle défendue par exemple par (Battistelli, 2009; Gosselin, 2005), c'est-à-dire comme regroupant – et soulignant l'articulation étroite entre – les dimensions sémantiques de temps (passé, présent, futur), d'aspect (accompli, inaccompli, *etc.*), de modalité (épistémique, axiologique, *etc.*) et de prise en charge énonciative (positionnement de l'énonciateur par rapport au contenu propositionnel), dorénavant notées TAME. Comme (Battistelli, 2009, p. 24) le note, « la temporalité linguistique, c'est [ainsi] cette manière dont la langue donne à « voir » à la lecture linéaire d'un texte des situations comme stables, comme évolutives, comme bornées, ... et saisissables à des moments, selon une certaines chronologie, reconstruite ou donnée. C'est aussi une manière de présenter des situations comme situées dans un registre de l'asserté, de l'hypothétique, du possible, ... et, à nouveau, de les relier entre elles (sur l'axe calendaire par exemple). Ou non. Du moins pas sur le plan strictement temporel (toujours au sens de relations du type succession ou concomitance) ».

Adverbiaux temporels et intensification de l'émotion. Les adverbiaux temporels, souvent appelés compléments circonstanciels de temps dans la grammaire traditionnelle, revêtent des formes très variées (adverbes, syntagmes prépositionnels, propositions subordonnées, *etc.*) et localisent dans le temps des contenus propositionnels. Si elle n'approfondit pas cette analyse, C. Roméro souligne le rôle des adverbes temporels tels que *toujours*, *jamais*, *souvent*, *tout le temps*, *parfois* ou encore *de temps en temps* dans le processus d'intensification d'une émotion exprimée (Romero, 2017). Les adverbiaux temporels indiquant une fréquence – voire une durée – apparaissent ainsi comme des marqueurs de temporalité en interaction avec l'expression des émotions.

Marqueurs de modalité, marqueurs d'émotion. La modalité linguistique est un phénomène complexe, souvent difficile à définir et ancré sur une grande hétérogénéité de marqueurs linguistiques. L. Gosselin la définit comme étant un « mode de validation » d'une prédication (2010, p. 309). Toute proposition énoncée se trouve ainsi associée à une modalité – plus ou moins marquée linguistiquement et facilement identifiable – qui (in)valide sous une forme ou une autre le contenu propositionnel. L. Gosselin (2005, 2010) distingue six catégories modales auxquelles sont associés des continuums de valeurs. Tout au long de ses ouvrages, il propose de plus de nombreux marqueurs et exemples analysés, constituant, à ma connaissance, le travail le plus exhaustif et le plus détaillé de ce point de vue. Ces catégories, continuums de valeurs et exemples de marqueurs sont illustrés par le tableau 6.

Avec les quelques exemples présentés tableau 6, il apparaît clairement comment certaines catégories modales recoupent l'expression des émotions, en particulier les modalités appréciatives (par ex. heureux, admirer) et axiologiques (par ex. honteux). Ce ne sont cependant pas les seules à jouer un rôle dans l'expression des émotions.

Catégorie modale	Continuum de valeurs	Exemples de marqueurs (Gosselin, 2010)
aléthique	impossible, contingent, possible, nécessaire	rectangulaire, « il neige », nécessité, chimiquement
épistémique	exclu, contestable, probable, certain	croire, probable, certitude, douteux, vraisemblable
déontique	interdit, facultatif, permis, obligatoire	interdire, obliger, permettre, impérativement
boulique	rejeté, indifférent, désiré	rejeter, souhaitable, désirer, vouloir
appréciative	indésirable, indifférent, désirable	beau, heureux, admirer, souffrance, chouette !
axiologique	blâmable, indifférent, louable	injuste, lâche, voleur, désapprouver, honteux

Tableau 6: Catégories modales, continuums de valeurs et exemples (Gosselin, 2005 ; 2010)

Émotions étayées, adverbiaux temporels, temps verbaux et modalité. R. Micheli (2014) définit les émotions étayées comme inférées par l'interlocuteur grâce à la schématisation, dans le discours, d'une situation propre à être associée à une émotion (*cf.* section 2.1). Il liste au total sept critères pouvant servir cette schématisation, chacun à même de donner des éléments clés pour inférer l'émotion. Parmi ces critères se trouve celui de la « distance (temps et espace) » (p. 115). Une situation présentée comme étant plus proche dans le temps (et dans l'espace) du locuteur ou de l'interlocuteur aura ainsi un « potentiel affectif » (p. 115) plus élevé (par ex. « au moment même où je vous parle » (p. 116)). Ce critère fait écho à une des observations menées en psychologie grâce au lexique LIWC (Tausczik & Pennebaker, 2010), selon laquelle un individu qui raconte un évènement émotionnel dans lequel il se sent encore impliqué aura tendance a utiliser plus de temps du présent (*cf.* section 3.2). Tous les marqueurs linguistiques permettant de situer un évènement comme étant plus ou moins proche du locuteur dans le temps, soient en général les temps verbaux et les adverbiaux temporels, seront ainsi susceptibles d'être impliqués dans l'étayage d'une émotion, dans son expression.

D'autres critères évoqués par R. Micheli sont ceux « des conséquences et de leur degré de probabilité » (p. 116) et du « potentiel de maîtrise » (p. 117). Décrire les conséquences d'une situation, qu'elles soient néfastes ou positives, peut ainsi influencer le potentiel affectif de la situation décrite. Le niveau de probabilité de ces conséquences est également un facteur très important pour l'étayage de l'émotion (par ex. usage du futur pour réduire la part d'incertitude des conséquences à venir (p. 124) ou présenter un risque comme étant « avéré » (p. 127)). Par ailleurs, si la situation est présentée comme pouvant être contrôlée ou non et par qui (locuteur, un autre agent, *etc.*), le potentiel affectif de la situation s'en verra à nouveau modifié (par ex. indiquer que les conséquences sont « inéluctables » (p. 124)). Les temps verbaux et adverbiaux temporels, de par leur capacité à indiquer l'arrivée imminente – ou non – de conséquences constituent à nouveau des marqueurs potentiels à la fois de temporalité et d'émotion. Les notions de probabilité et de potentiel de maîtrise renvoient quant à elles aux valeurs de la modalité épistémique, voire déontique (*cf.* tableau 6).

Le dernier critère introduit par R. Micheli est celui de la « significativité normative » (p. 118). Ainsi, si une situation est présentée comme contrevenant aux normes, aux valeurs d'un groupe social donné, elle pourra étayer des émotions négatives comme de la culpabilité, de la honte ou de l'indignation. Ce critère paraît fortement lié à des marqueurs exprimant, en plus d'une émotion étayée, de la modalité axiologique (*cf.* tableau 6).

Émotions et prise en charge énonciative. La prise en charge énonciative est une notion linguistique qui renvoie à la façon dont un énonciateur transforme un contenu propositionnel en un énoncé en le situant dans le temps, le dotant d'une valeur aspectuelle et d'une modalité. Ainsi intrinsèquement liée au temps, à l'aspect et à la modalité, la prise en charge renvoie plus spécifiquement au positionnement de l'énonciateur par rapport au contenu prédicatif, à la source de l'information transmise et à son degré de vérité. La notion de prise en charge se cristallise donc surtout autour des concepts de vérité, d'assertion ou encore de source du savoir. Le locuteur – l'individu qui parle ou écrit – doit alors être différencié de l'énonciateur – l'individu qui prend en charge l'énoncé – qui est la source du point de vue exprimé (Coltier et al., 2009).

Les liens entre prise en charge et émotions ont déjà été quelque peu explorés par A. Rabatel à travers l'étude de l'empathie linguistique, renvoyant au phénomène où « un locuteur prête sa voix à un autre [...] pour envisager un évènement, une situation, à sa place » (2013, p. 2). Ce procédé, représentant parfois un ressort rhétorique

pour l'argumentation, repose donc sur la « mise en spectacle » (2013, p. 6) par le locuteur/énonciateur premier de l'émotion ressentie par un énonciateur second.

Sans être forcément nommée telle qu'elle, la question de la prise en charge apparaît également dans les définitions des modes d'expression des émotions de R. Micheli (2014), en particulier celles des émotions montrées et étayées où l'attribution de l'émotion à un expérienceur passe par des jeux de prise en charge (une émotion montrée sera par exemple par défaut attribuée au locuteur, à moins qu'il ne cède la parole à un autre énonciateur par le biais d'un discours rapporté).

Si elle n'est également pas explicitement citée, l'importance de la prise en charge peut cependant être devinée dans (Leeman-Bouix, 1995). Ce travail, focalisé sur l'étude de la structure *être en N*, où *N* est un nom d'émotion (par ex. être en colère), présente ainsi le *N* comme correspondant à une sorte de « compte-rendu de perception » (*ibid.,* p. 67) exprimé par le locuteur et renvoyant à la façon dont l'entité sujet du prédicat *être en N* est perçue par le locuteur.

Les liens entre émotions et prise en charge sont enfin aisément illustrés par les exemples construits [1] à [5], où chaque énoncé réalise le même contenu prédicatif (être_triste(x), où x = Lou) mais avec différents niveaux de prise en charge.

[1] Pol aurait dit que Lou est triste.
[2] Pol dit que Lou est triste.
[3] Lou a l'air triste.
[4] Lou est triste.
[5] Que Lou est triste !

Dans les énoncés [1] et [2], le locuteur pose une autre entité, un autre énonciateur (Pol) comme la source de l'information « Lou est triste » et s'en distancie ainsi. Cette distance est accentuée en [1] par l'emploi du conditionnel « aurait » qui exprime l'incertitude du locuteur. Les énoncés [1] à [5] présentent une certitude croissante du locuteur sur l'information « Lou est triste » ainsi qu'une impression grandissante à la fois de l'intensité de l'émotion de Lou et de l'implication du locuteur (le locuteur paraît plus affecté par cette tristesse en [5]). Les marqueurs de prise en charge semblent ainsi pouvoir intervenir dans l'expression des émotions, au moins d'une part pour savoir à qui attribuer l'émotion, et d'autre part en jouant un rôle sur son intensité.

Sans avoir jamais été profondément explorés, les liens entre dimensions TAME et expression des émotions transparaissent ainsi dans diverses études linguistiques où les

marqueurs de temporalité constituent tantôt des marqueurs émotionnels à part entière (*adverbiaux temporels* et *marqueurs modaux*), tantôt des marqueurs qui intensifient l'émotion (*adverbiaux temporels* et *marqueurs de prise en charge*) et tantôt des marqueurs permettant d'attribuer l'émotion à un expérienceur (*marqueurs de prise en charge*). Étant donné la complexité du champ sémantique de la temporalité prise au sens large (c'est-à-dire comme rendant compte d'une interaction très forte entre les dimensions du temps, de l'aspect, de la modalité et de la prise en charge), j'ai choisi de me concentrer sur certains travaux de référence seulement. La notion d'*adverbial temporel* est ainsi abordée ici telle qu'elle est conceptualisée dans (Battistelli et al., 2008) puis dans (Battistelli, 2009; Teissedre, 2012), la notion de *modalité* dans (Gosselin, 2005, 2010) et la notion de *prise en charge* dans (Battistelli & Amardeilh, 2009).

3) Les émotions en TAL

Dans le domaine du TAL, l'analyse des émotions est généralement perçue comme une tâche de classification nécessitant l'élaboration de corpus annotés en émotions (section 3.1). Une autre ressource souvent utilisée dans ce domaine pour l'étude des émotions sont les lexiques émotionnels, qu'ils aient été constitués dans le cadre de travaux de TAL ou de psychologie (section 3.2). En s'appuyant sur un corpus annoté en émotions et en sollicitant parfois des lexiques émotionnels, de nombreux travaux développent ainsi des systèmes d'analyse automatique des émotions (section 3.3).

3.1) Corpus annotés en émotions

En TAL, l'annotation des émotions est usuellement considérée comme une tâche difficile, notamment à cause du manque de consensus sur le nombre et les types de catégories émotionnelles (par ex. colère, peur, *etc.*), le manque de frontière nette entre ces catégories ou encore la grande variabilité des moyens d'exprimer les émotions. Cela conduit à une grande hétérogénéité de corpus annotés en émotions, comme l'illustrent plusieurs revues de littérature.

Par exemple, L.-A.M. Bostan et R. Klinger (2018) comparent quatorze corpus et mettent en avant les différences qui incluent, mais ne se limitent pas à : l'objectif de l'annotation (par ex. la fouille de réseaux sociaux, le développement d'agents conversationnels), le genre de textes (par ex. titres d'articles de presse, contes), les schémas d'annotation (types et nombre de catégories émotionnelles, annotations d'autres élé-

ments telles que la polarité, l'intensité ou la cause de l'émotion), les procédures d'annotation (par ex. annotation manuelle par des experts, annotation manuelle par *crowd-sourcing*) et la granularité des annotations (par ex. phrases ou tweets entiers). Pour sa part, E. Öhman (2020) souligne les difficultés qui émergent dans la tâche d'annotation en émotions, notamment concernant le choix des catégories émotionnelles employées, le type de procédure d'annotation mise en œuvre ou encore la subjectivité inhérente à l'annotation en émotions, qui conduit le plus souvent à des accords inter-annotateurs bas. Elle se focalise sur ce qui peut influencer la qualité des annotations, d'après ce qui transparaît des mesures d'accords inter-annotateurs. Des facteurs influençant la qualité des annotations sont par exemple le genre et le sujet des textes annotés, le nombre et le type des catégories émotionnelles ou encore le nombre d'annotateurs et le type d'entraînement qu'ils ont reçu.

(Bostan & Klinger, 2018; Öhman, 2020) montrent ainsi clairement que la *catégorie émotionnelle* constitue la notion la plus communément employée pour analyser les émotions en TAL, avec certains travaux se concentrant également sur d'autres notions, telle que la *cause* de l'émotion. Cependant, ces deux revues de littérature n'offrent pas de vue d'ensemble de la façon dont plusieurs notions d'importance dans l'approche proposée dans ma thèse, à savoir les *modes d'expression*, l'*expérienceur* et les types d'unités annotées, sont prises en compte (ou non) dans les schémas d'annotation en TAL.

Langue étudiée et genre textuel. Deux remarques préliminaires, indépendantes de l'analyse des émotions en tant que telle, concernent la langue des corpus annotés en émotions et les genres des textes qui les composent. La plupart des travaux de TAL traitent ainsi des corpus anglophones – voire multilingues – avec, à ma connaissance, seulement deux corpus francophones annotés en émotions proposés dans (Fraisse & Paroubek, 2015; Paroubek et al., 2010). La majorité des corpus sont par ailleurs composés de textes issus des réseaux sociaux (tweets ou textes assimilés, posts de blogs ou commentaires Web) ou de textes de fiction. Peu semblent composés d'extraits de textes journalistiques ou être des corpus multi-genres. Les deux corpus francophones sont composés de textes issus des réseaux sociaux.

Les catégories émotionnelles. Quand des travaux de TAL s'attaquent à l'analyse des émotions, ils s'appuient généralement sur des typologies issues de la psychologie définissant un ensemble d'émotions de base. La plupart des travaux de TAL citent

ainsi (Ekman, 1992) qui distingue six émotions de base (colère, dégoût, joie, peur, surprise et tristesse). Ils reprennent alors partiellement (Bianchi et al., 2021; Cheng et al., 2017) ou totalement cette typologie (Gui et al., 2018; S. Mohammad, 2012; Strapparava & Mihalcea, 2007) ou ajoutent parfois une catégorie permettant d'annoter l'absence d'émotion ou des ressentis émotionnels mitigés (Alm, 2010; Alm et al., 2005; Aman & Szpakowicz, 2007; Buechel & Hahn, 2017; Russo et al., 2011). D'autres travaux citent (Plutchik, 1980) qui ajoute deux catégories supplémentaires (confiance et anticipation) à la liste d'émotions de base de P. Ekman. Ce sont par exemple les travaux décrits dans (Kim & Klinger, 2018, 2019b, 2019a; C. Liu et al., 2019; Öhman et al., 2020). Certains travaux enfin font appel à des listes de catégories, mélanges d'émotions de base et d'émotions complexes, issues d'autres références, d'intuition ou de retours empiriques sur corpus, comme c'est le cas dans (Bostan & Klinger, 2018; Demszky et al., 2020; Fraisse & Paroubek, 2015, 2015; S. Mohammad et al., 2018; S. Mohammad & Kiritchenko, 2018; Neviarouskaya & Aono, 2013; Schmidt et al., 2021).

L'expérienceur de l'émotion. En ce qui concerne l'*expérienceur* de l'émotion, il semble largement laissé pour compte dans les travaux de TAL. De nombreux travaux ne précisent ainsi pas explicitement *qui* ressent les émotions annotées (Bianchi et al., 2021; Ghazi et al., 2015; Gui et al., 2018; C. Liu et al., 2019; Russo et al., 2011), quand bien même l'expérienceur apparaît dans le schéma d'annotation (Neviarouskaya & Aono, 2013). Les travaux qui n'étiquettent pas l'expérienceur mais indiquent malgré tout clairement qui ressent l'émotion annotée se focalisent le plus souvent sur les émotions du scripteur (voir par ex. (Demszky et al., 2020; Fraisse & Paroubek, 2015; S. Mohammad, 2012; S. Mohammad et al., 2018; Öhman et al., 2020; Paroubek et al., 2010)). Plus rarement, ce sont les émotions du lecteur (Strapparava & Mihalcea, 2007) qui sont analysées, voire à la fois celles du scripteur et du lecteur (Buechel & Hahn, 2017). Les schémas qui prévoient l'annotation de l'expérienceur sont centrés principalement sur les émotions des personnages des textes (Alm, 2010; Alm et al., 2005; Cheng et al., 2017; Kim & Klinger, 2018, 2019a, 2019b; Schmidt et al., 2021), bien que (Cheng et al., 2017) annotent aussi les émotions du lecteur et du scripteur.

Les modes d'expression des émotions. Comme nous l'avons vu, une des notions essentielles intervenant pour caractériser la notion d'émotion en (psycho-)linguistique est la diversité des *modes d'expression* pouvant les dénoter (*cf.* sections 1 & 2). Bien que cette diversité ait déjà été mentionnée dans plusieurs travaux de TAL (voir par exemple (Alm et al., 2005; Bostan & Klinger, 2018; Demszky et al., 2020; Kim &

Klinger, 2018; S. Mohammad, 2012; S. Mohammad et al., 2018; S. Mohammad & Kiritchenko, 2018; Strapparava & Mihalcea, 2007)), elle n'a jusqu'ici jamais été conceptualisée.

De ce point de vue, (Alm, 2010) constitue le travail le plus détaillé. L'auteure liste les caractéristiques de 460 phrases émotionnelles en anglais extraites de contes pour lesquelles l'annotation produite suscite un fort accord inter-annotateurs. Bien qu'ils ne soient pas organisés en une typologie, les indices cités font écho aux marqueurs mentionnés pour les différents *modes d'expression* définis en linguistique (*cf.* section 2) et psycho-linguistique (*cf.* section 1). Par exemple, les « mots d'affect » (traduit de l'anglais, *ibid.*, p. 120) mentionnés correspondent aux émotions désignées. Les « items lexicaux ou syntagmes qui décrivent des actions, des propriétés, des comportements » (par ex. rire, sangloter, pleurer) (trad. de l'anglais, p. 120) renvoient aux émotions comportementales. Les « mots ou expressions d'une polarité négative ou positive » qui sont interprétés étant donné un « contexte [particulier] et des connaissances acquises » (par ex. un magnifique jardin de fleurs, un petit trésor sur cette terre) (trad. de l'anglais, pp. 120-121) sont similaires aux émotions suggérées. Enfin, le « discours direct » incluant des « exclamations ou des questions en WH, des énoncés courts, des interjections » s'apparente aux émotions montrées. Le schéma introduit chapitre 3 propose une façon de nommer et de catégoriser ces différents marqueurs des émotions.

Bien moins complets dans leur approche, deux autres travaux de TAL proposent une analyse des émotions ne s'appuyant pas (uniquement) sur le lexique émotionnel. A. Balahur, J.-M. Hermida et A. Montoyo (2012) s'attellent ainsi à la détection automatique des émotions « implicites » présentes dans de courts récits de situations émotionnelles vécues par des étudiants (par ex. « J'ai échoué aux examens. » exprime de la tristesse et « J'ai échoué aux examens parce que je n'ai pas assez travaillé. » de la culpabilité), similaires aux émotions suggérées mentionnées dans les études (psycho-)linguistiques. Ils emploient pour cela une ontologie, EmotiNet, dans laquelle sont stockées des situations susceptibles de déclencher des émotions. Pour leur part, E. Kim et R. Klinger (2019a) se concentrent sur l'analyse des « canaux » de communication non-verbal des émotions auxquels ont recours les personnages de nouvelles de fanfiction (*ibid.*, p. 2). Ils distinguent alors huit canaux : l'apparence physique, les expressions faciales, le regard, les gestes des mains et des bras, les mouvements de l'ensemble du corps, la voix, les relations spatiales et les sensations physiques. Cette étude approfondit ainsi les émotions comportementales citées dans les études psycho-linguistiques.

Processus d'annotation. L'annotation manuelle des émotions est une tâche complexe du fait de l'ensemble des critères que nous venons de voir. Les approches en TAL recourent ainsi majoritairement à des annotateurs humains, soit en petit nombre, généralement deux ou trois (Alm, 2010; Alm et al., 2005; Aman & Szpakowicz, 2007; Bianchi et al., 2021; Cheng et al., 2017; Demszky et al., 2020; Gui et al., 2018; Kim & Klinger, 2018, 2019b, 2019a; Neviarouskaya & Aono, 2013; Öhman et al., 2020; Paroubek et al., 2010; Russo et al., 2011; Schmidt et al., 2021; Strapparava & Mihalcea, 2007), soit en production participative (*crowdsourcing*) (Buechel & Hahn, 2017; C. Liu et al., 2019), soit les deux (S. Mohammad et al., 2018; S. Mohammad & Kiritchenko, 2018). Les annotations correspondent enfin parfois aux hashtags de tweets produits directement par le scripteur (S. Mohammad, 2012).

Des lexiques émotionnels peuvent être mobilisés pour sélectionner les phrases ou textes à inclure dans le corpus (Aman & Szpakowicz, 2007; Gui et al., 2018; Kim & Klinger, 2018; C. Liu et al., 2019; S. Mohammad et al., 2018; S. Mohammad & Kiritchenko, 2018; Neviarouskaya & Aono, 2013; Russo et al., 2011). Ces lexiques interviennent aussi parfois pour pré-annoter les émotions, avant une vérification humaine, par exemple (Ghazi et al., 2015). Il existe quelques corpus annotés automatiquement en émotions, comme (Bostan & Klinger, 2018) qui agrège automatiquement les étiquettes de corpus annotés en émotions existants afin de constituer un unique corpus annoté ; (Fraisse & Paroubek, 2015) qui recourt à des couples interjection-émotion pour récupérer des tweets et les annoter en émotions selon l'interjection qu'ils contiennent ; ou encore (Öhman et al., 2020) qui projette automatiquement des annotations sur des corpus parallèles, à partir d'un corpus annoté manuellement.

Malgré la prévalence de l'annotation manuelle, notons qu'il semble y avoir peu de guides d'annotation associés aux corpus. Seuls (Kim & Klinger, 2018) et (Schmidt et al., 2021) mentionnent explicitement la présence d'un guide.

Types d'unités annotées. Les unités annotées avec des informations émotionnelles sont généralement des textes entiers (Bianchi et al., 2021; Demszky et al., 2020; Fraisse & Paroubek, 2015; C. Liu et al., 2019; S. Mohammad, 2012; S. Mohammad et al., 2018) ou des phrases (Alm, 2010; Alm et al., 2005; Buechel & Hahn, 2017; Ghazi et al., 2015; Öhman et al., 2020; Strapparava & Mihalcea, 2007), bien que plusieurs échelles d'annotation soient parfois sollicitées, par exemple le texte et le paragraphe dans (Paroubek et al., 2010) et la phrase et le marqueur linguistique dans (Aman &

Szpakowicz, 2007; Neviarouskaya & Aono, 2013). Il est intéressant de noter que même s'ils annotent les émotions à l'échelle du texte, A. Fraisse et P. Paroubek (2015) observent comment des marqueurs linguistiques précis, les interjections, jouent un rôle dans l'expression des émotions. Les travaux qui effectuent l'annotation en émotions uniquement à l'échelle du marqueur linguistique repèrent également la *cause* de l'émotion (Gui et al., 2018; Russo et al., 2011) et son *expérienceur* (Cheng et al., 2017; Kim & Klinger, 2018, 2019b, 2019a; Neviarouskaya & Aono, 2013; Schmidt et al., 2021).

Accords inter-annotateurs dans les tâches d'annotation des émotions. Comme le souligne (Öhman, 2020), un témoignage fort de la difficulté de la tâche d'annotation en émotions est l'ensemble des mesures d'accord inter-annotateurs généralement basses obtenues pour différents travaux : avec une tâche d'étiquetage impliquant deux ou trois classes, l'accord ne dépasse jamais 70-80 % et il chute lorsque des classes sont ajoutées. Ces mesures sont habituellement limitées aux catégories émotionnelles, bien que certains travaux comme (Kim & Klinger, 2018) présentent des mesures d'accord pour chaque élément de leur schéma d'annotation. Par exemple, le Kappa de Cohen entre chaque paire d'annotateurs (trois annotateurs, soit trois paires en tout) s'échelonnent de 0,06 (catégorie « autres émotions ») à 0,40 (joie) selon la catégorie émotionnelle considérée. La variabilité de l'accord entre les paires d'annotateurs se reflète également dans les mesures de F1. Les valeurs de F1 *strict* s'échelonnent respectivement de 6 % à 40 % pour l'annotation des catégories émotionnelles (les valeurs de *fuzzy* F1 de 7 à 55 %), de 10 % à 63 % pour l'identification de l'expérienceur (*fuzzy* F1 de 18 à 68 %) et de 14 % à 65 % pour l'étiquetage de la cause (*fuzzy* F1 de 28 à 73 %). Toutes classes confondues, le Kappa de Cohen (trois annotateurs) pour l'identification de la catégorie émotionnelle est de 0,3476 dans (Schmidt et al., 2021).

Points de comparaison. Les travaux les plus approchants de la présente étude sont ceux annotant à l'échelle du marqueur textuel à la fois les unités émotionnelles – en les caractérisant par la *catégorie émotionnelle* qu'elles expriment – et leur *expérienceur* et leur *cause* (ou plus exactement de l'objet pour (Schmidt et al., 2021)). Parmi ces travaux, seuls ceux traitant de textes de fiction (Kim & Klinger, 2018, 2019b, 2019a; Schmidt et al., 2021) peuvent réellement être comparés à ce qui est présenté ici, au vu de l'impact du genre de texte sur les annotations produites (Öhman, 2020). Par ailleurs, (Kim & Klinger, 2019b, 2019a) ne prennent en compte les unités émo-

tionnelles que lorsqu'elles expriment une émotion qui implique deux personnages. Les deux travaux les plus semblables à ce qui est présenté ici sont donc seulement (Kim & Klinger, 2018; Schmidt et al., 2021). Il est cependant important de noter que, si (Schmidt et al., 2021) s'appuie sur une annotation en contexte de textes entiers, (Kim & Klinger, 2018) ne fait appel qu'à des triplets de phrases extraits d'ouvrages de fiction, en proposant néanmoins une annotation multi-étiquettes des catégories émotionnelles associées à une unité émotionnelle. À ma connaissance, (Kim & Klinger, 2018) est par ailleurs la seule étude qui mentionne brièvement le rôle des émotions dans la « compréhension de textes littéraires » (traduit de l'anglais, *ibid.*, p. 1346), bien qu'il ne s'appuie cependant pas sur des études psycho-*linguistiques* pour aborder l'analyse des émotions. À noter par ailleurs qu'aucun corpus n'a été annoté avec la notion de *conséquence* de l'émotion.

3.2) Lexiques émotionnels

L'étude de l'expression des émotions s'appuie bien souvent sur la constitution de lexiques émotionnels, dans lesquels un terme (ou un syntagme) se trouve associé à au moins une catégorie émotionnelle qu'il exprime. Si ces lexiques ne sont pas au cœur de l'approche revendiquée dans ma thèse, il paraissait incontournable de donner un aperçu des types de lexiques utiles à l'analyse des émotions, qu'ils soient élaborés dans le cadre strict d'études de TAL ou dans le cadre de travaux en psychologie. Ne sont mentionnés ici que les ressources anglophones apparaissant le plus fréquemment dans les travaux de TAL ainsi que quelques références majeures francophones.

Lexiques émotionnels issus du TAL. Dans le domaine du TAL, la constitution d'un lexique émotionnel est le plus souvent vue comme une tâche d'annotation de lexèmes. Elle représente cependant « un coût élevé et un considérable effort manuel requis par les annotateurs humains dans un cadre traditionnel où des experts choisis un à un sont engagés pour effectuer toute l'annotation » (traduit de l'anglais, S. M. Mohammad & Turney, 2013, p. 2). Certains auteurs se tournent ainsi vers les méthodes de *crowdsourcing*, moins coûteuses et permettant d'obtenir rapidement un grand nombre d'annotations. Le lexique EmoLex – parfois aussi nommé *National Research Council Canada (NRC) lexicon* – (S. M. Mohammad & Turney, 2013) regroupe ainsi environ 9 000 termes anglophones, annotés par *crowdsourcing* selon les huit catégories d'émotions de base de (Plutchik, 1980). Le *French Expanded Emotion Lexicon* (FEEL) (Abdaoui et al., 2017), comportant plus de 14 000 termes francophones, ré-

sulte de la traduction automatique de EmoLex, avec une étape de validation humaine de la traduction et de l'association des termes aux catégories émotionnelles, puis d'ajout de synonymes[19].

D'autres travaux s'appuient sur des corpus classés en émotions pour constituer automatiquement leur lexique émotionnel. C'est par exemple le cas du *Twitter Emotion Corpus (TEC) lexicon* (S. Mohammad, 2012), composé de plus de 11 000 termes anglophones extraits automatiquement selon la force de leur association avec une classe émotionnelle, depuis un corpus de tweets classés selon les six émotions de base de (Ekman, 1992). Une méthode similaire a été employée dans (Fraisse & Paroubek, 2015), pour constituer un lexique de plus de 1 500 termes francophones à partir d'un corpus de tweets classés en émotions, selon huit classes ('colère', 'plaisir', 'peur', 'apaisement', 'tristesse', 'mépris', 'surprise négative' et 'insatisfaction').

Le lexique WordNet Affect (Strapparava & Valitutti, 2004), regroupant une centaine de termes anglophones, correspond quant à lui à une ontologie où les termes ont été associés de manière semi-automatique à une émotion : quelques termes associés à une catégorie émotionnelle ont été identifiés manuellement puis la même catégorie a été attribuée à tous leurs synonymes d'après le réseau WordNet[20]. Si WordNet Affect emploie initialement neuf classes génériques pour répartir les termes ('émotionnel', 'affectif non-émotionnel', 'état mental non affectif', 'trait de personnalité', 'comportement', 'attitude', 'état physique' et 'sentiment'), un ensemble classé selon les émotions de base de (Ekman, 1992) est cependant proposé.

Du côté du TAL et de la linguistique de corpus français, des approches mieux fondées linguistiquement sont proposées pour l'élaboration de lexiques émotionnels. C'est par exemple le cas du lexique décrit dans (Mathieu, 2005), comportant plus de 900 verbes, répartis en douze classes ('amertume', 'amour', 'amusement', 'apaisement', 'consternation', 'déception', 'dédain', 'dégoût', 'déprime', 'dérangement', 'désapprobation', 'désir') ou de celui développé dans le cadre du projet ANR EMOLEX (Diwersy et al., 2014). Ce dernier s'intéresse tout particulièrement aux cooccurrents et collocatifs des termes d'émotion et propose pour le français, l'allemand, l'anglais, le russe et l'espagnol près de 800 lexies émotionnelles et 16 000 collocations. Les lexies, identifiées d'après des définitions lexicographiques et des tests de combinatoire, sont réparties en neufs classes ('respect', 'joie', 'admiration', 'colère', 'déception', 'jalousie', 'mépris', 'tristesse', 'surprise').

19 Bien qu'issu de la traduction de EmoLex, FEEL ne reprend que les six catégories d'émotions de (Ekman, 1992).
20 https://wordnet.princeton.edu/

Lexiques émotionnels issus de la psychologie. La constitution de lexiques pour l'étude des émotions est une tâche également présente dans les travaux de psychologie, afin de permettre une analyse automatique et plus objective des termes émotionnels présents dans les textes. Tous ces travaux s'appuient abondamment sur des études empiriques (par ex. récits autobiographiques de situations émotionnelles), des questionnaires, d'autres études de psychologie sur la définition de différents affects et leur catégorisation (émotions, sentiments, humeurs, *etc.*) et sur l'intuition pour déterminer à la fois les classes et les termes à inclure dans la ressource constituée. Le lexique *Geneva Affect Label Coder* (GALC) (Scherer, 2005) propose par exemple pour le français, l'anglais et l'allemand environ 280 radicaux choisis et classés intuitivement par l'auteur en 38 classes (par ex. 'admiration', 'amour', 'amusement', 'bonheur', 'colère' ou 'tension & stress').

Le *Linguistic Inquiry and Word Count* (LIWC) (Tausczik & Pennebaker, 2010) avait pour objectif initial de permettre le calcul des proportions de mots positifs et négatifs dans des textes. Il distingue finalement 80 catégories très diverses (par ex. 'travail', 'argent', 'mort'), dont plusieurs d'ordre morphosyntaxique (par ex. 'temps du passé', 'adverbes', 'clitiques', 'prépositions') et une classe sémantique contenant les termes émotionnels à proprement parler, la classe 'processus affectifs', décomposée en 'émotion positive' et 'émotion négative' (dont 'anxiété', 'colère' et 'tristesse'). Le LIWC comporte au total, plus de 100 000 termes anglophones (915 pour la classe 'processus affectifs') identifiés à partir de dictionnaires, thésaurus, questionnaires et de l'intuition puis validé par trois juges humains. Après avoir présenté le LIWC, Y. R. Tausczik et J. W. Pennebaker (2010) soulignent de nombreux phénomènes reliant caractéristiques linguistiques des textes et état émotionnel du scripteur ayant été mis en avant grâce à ce lexique. Un usage plus prononcé de pronoms de la première personne du singulier est par exemple constaté dans les récits d'individus souffrant de douleur physique ou morale. La proportion de verbes au passé est quant à elle plus élevée lorsque le scripteur raconte un évènement émotionnel déjà évoqué, ce qui indique une certaine distance psychologique par rapport à l'évènement, contrastant de fait avec les évènements précédemment non évoqués plus fréquemment relatés à l'aide de temps du présent. Enfin, dans des récits de femmes battues, un usage plus prononcé de mots d'émotions (par ex. *hurt, nasty, ugly, etc.*) montre une immersion plus profonde dans l'évènement traumatique, ce qui induit une expérience accrue de la douleur physique. Sur le plan linguistique, Y. R. Tausczik et J. W. Pennebaker (2010) rapportent des

corrélations significatives avec l'usage de mots émotionnels et d'une part celui d'articles, de prépositions et de mots de relativité (corrélations négatives) et d'autre part celui de pronoms, d'auxiliaires verbaux, et de la négation (corrélations positives). Toutes ces observations soulignent la nécessité de s'extraire de la simple étude du lexique émotionnel pour analyser *le caractère émotionnel des textes*.

Une version française du LIWC a été proposée dans (Piolat et al., 2011). Cette version contient plus de termes que le lexique d'origine à cause de phénomènes propres à la tâche de traduction. Elle a de plus nécessité quelques adaptations pour prendre en compte les spécificités du français par rapport à l'anglais, notamment l'élision des clitiques ou de la négation et le paradigme flexionnel des verbes en français, bien plus complexe que celui des verbes anglais. Cette version reste malgré tout très proche du LIWC anglais.

Consacré cette fois uniquement au lexique émotionnel, le lexique EMOTion AIX (EMOTAIX) (Piolat & Bannour, 2009) inclut plus de 2 000 « référents » francophones, soient des « terme[s] (en général un nom) qui rassemble[nt] de 1 à n mots comportant le même radical et partageant un sens proche » (Piolat & Bannour, 2009, p. 671). Le référent « horreur » rassemble par exemple les termes horreur, horrible, horriblement, horrifier et horrifié. Ce lexique est organisé de manière hiérarchique avec une symétrie basée sur la valence opposant les classes. Les classes du grain le plus gros sont ainsi 'malveillance' opposée à 'bienveillance', 'mal-être' opposé à 'bien-être' et 'anxiété' opposée à 'sang-froid'. Les classes les plus fines ou catégories sémantiques de base sont au nombre de 56, parmi lesquelles se trouvent par exemple les six émotions de base de (Ekman, 1992) ou les catégories 'humiliation', 'admiration', 'santé mentale', 'bon sens' et 'sérénité'. L'organisation du lexique et ses classes ont été retenues d'après de nombreux travaux de psychologie sur les notions d'émotion, d'affect, d'humeur et de sentiment. L'assignation de termes à une catégorie s'est quant à elle appuyée sur un certain nombre de critères linguistiques lexicographiques. Ces critères explicites ont été réemployés pour caractériser et aider l'annotation des émotions désignées dans le guide d'annotation (Étienne & Battistelli, 2021)(*cf.* chapitre 3). Ce sont également eux qui ont motivé l'utilisation du lexique EMOTAIX dans le projet TextToKids pour la chaîne d'extraction de descripteurs de la complexité des textes (*cf.* chapitres 1 & 6).

La diversité des lexiques émotionnels existants, des catégories émotionnelles qui les structurent et des méthodes pour les constituer illustrent une fois encore la difficulté

de la tâche d'analyse (automatique) des émotions. Même s'ils ne sont pas employés directement dans le travail présenté ici, ils constituent des ressources précieuses et offrent une meilleure vue d'ensemble des méthodologies d'approche de l'étude des émotions.

3.3) Systèmes d'analyse automatique des émotions

À partir d'un corpus annoté en émotions d'une taille et d'une qualité suffisantes, il est possible de développer un système d'analyse automatique des émotions. L'analyse des émotions[21] est alors considérée comme une tâche de classification : une classe parmi un ensemble donné (l'ensemble des catégories émotionnelles) doit être automatiquement associée à un item donné en entrée du système (phrase, texte, *etc.*). De manière générale, la grande hétérogénéité des corpus annotés se retrouve dans les travaux d'analyse automatique des émotions par la diversité des classes émotionnelles, des échelles de prédiction et des méthodes de développement de classifieurs en émotions. La façon de présenter les résultats varie également d'une étude à l'autre, rendant parfois difficile les comparaisons. Seules les méthodes principales intervenant pour l'élaboration de tels systèmes, reposant ou non sur l'utilisation de lexiques émotionnels, seront présentées ici. De même, seules de grandes tendances du point de vue des performances seront exposées.

Historiquement, les modèles de type *Support Vector Machine* (SVM) interviennent en classification automatique des émotions pour effectuer des prédictions à l'échelle de la phrase (Aman & Szpakowicz, 2007; S. Mohammad, 2012) ou du texte (Abdaoui et al., 2017; Balahur et al., 2012; Fraisse & Paroubek, 2015; S. Mohammad, 2012). Les entrées fournies à ces modèles sont le plus souvent de la forme « sacs de mots » mais incluent aussi parfois des traits issus de ressources émotionnelles comme par exemple WordNetAffect (Aman & Szpakowicz, 2007; Balahur et al., 2012; Strapparava & Mihalcea, 2007) ou le lexique FEEL (Abdaoui et al., 2017). Quelques modèles symboliques s'appuyant sur des lexiques (Strapparava & Mihalcea, 2007) ont également été proposés. Les modèles SVM obtiennent cependant globalement de meilleures performances. Les modèles symboliques proposés pour l'anglais dans (Strapparava & Mihalcea, 2007) reconnaissent par exemple la classe 'joie' à l'échelle de la phrase avec au

21 Les travaux d'extraction de la cause de l'émotion, bien que proposant des corpus annotés intéressants, visent seulement l'identification automatique de la cause d'une information émotionnelle *préalablement* fournie au modèle. Les systèmes développés dans le cadre de ces travaux ne cherchent donc pas à classer des items en émotions et ne seront ainsi pas évoqués dans cette section.

mieux une F-mesure de 0,12 et la classe 'colère' avec une F-mesure de 0,16. Le modèle SVM proposé pour cette langue dans (S. Mohammad, 2012) obtient sur le même corpus une F-mesure de 0,49 pour la 'joie' et de 0,35 pour la 'colère'.

Les meilleurs classifieurs en émotions actuels sont tous de type Transformeur. Le tableau 7 présente, pour plusieurs travaux, les F-mesures (F1) des meilleurs modèles obtenus. Seuls les résultats correspondant à l'une des catégories retenues dans le schéma d'annotation du chapitre 3, exceptées culpabilité et jalousie toujours absentes des travaux, sont reportés dans ce tableau[22], soit admiration, colère, dégoût, embarras, fierté, joie, peur, surprise et tristesse. La colonne « Toutes classes » contient les F-mesures moyennes du modèle (le type de moyenne « micro » ou « macro » est indiqué lorsqu'il était mentionné dans l'article). Le type de corpus, l'échelle de prédiction, la langue du corpus – identifiée en suivant la norme ISO-639-1[23]– et le nombre de classes émotionnelles sont également précisés. Si les performances atteintes par ces modèles fluctuent d'une classe à l'autre, elles dépassent rarement une F-mesure de 0,75-0,80. Ces performances sont globalement supérieures à celles des modèles historiques (SVM, symboliques). À ma connaissance, il n'existe cependant aucun classifieur en émotions de type Transformeur pour le français.

Réf.	**Corpus**	**Échelle**	**Lang.**	**Nb cla.**	**F1 du meilleur modèle**									
					Toutes classes	**ad.**	**co.**	**dé.**	**em.**	**fi.**	**jo.**	**pe.**	**su.**	**tr.**
(Liu *et al.*, 2019)	extraits de récits	texte	EN	6	0,60 (mi.)	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
(Demszky *et al.*, 2020)	posts reddit	texte	EN	28	0,46 (ma.)	0,65	0,47	0,45	0,43	0,36	0,51	0,60	0,50	0,49
				6	0,64 (ma.)	*n/a*	0,57	0,53	*n/a*	*n/a*	0,82	0,68	0,61	0,59
(Öhman *et al.*, 2020)	sous-titres	phrase	EN	8	0,54 (ma.)	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
			AR	8	0,46	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
			ZH	8	0,47	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
			NE	8	0,52	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
			FI	8	0,45	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
			DE	8	0,51	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
			TR	8	0,47	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
(Bianchi *et al.*, 2021)	tweets	texte	IT	4	0,71	*n/a*	0,88	*n/a*	*n/a*	*n/a*	0,87	0,51	*n/a*	0,6

Tableau 7: F1 des meilleurs classifieurs en émotions de type Transformeur

22 Les résultats obtenus pour d'autres catégories émotionnelles ne sont donc pas reportés dans le tableau.
23 https://fr.wikipedia.org/wiki/Liste_des_codes_ISO_639-1

Réf.	**Corpus**	**Échelle**	**Lexique**	**Nb cla.**	**F1 du meilleur modèle**				
					Toutes classes	**co.**	**pe.**	**su. nég.**	**tr.**
(Fraisse, Paroubek, 2015)	billets de blog	paragraphe	(Fraisse, Paroubek, 2015)	8	*n/a*	0,50	0,16	0,20	0,37
	tweets	texte	non	8	*n/a*	0,66	0,16	0,08	0,35
(Abdaoui *et al.*, 2017)	tweets	texte	FEEL	4	0,73 (mi.)	*n/a*	*n/a*	*n/a*	*n/a*

Tableau 8: F1 des meilleurs classifieurs en émotions pour le français

Les classifieurs en émotions qui existent pour le français sont tous de type SVM et recourent, ou non, à des lexiques émotionnels pour enrichir les entrées fournies aux modèles. Les F-mesures (F1) des meilleurs modèles SVM pour le français sont présentées tableau 8. Comme pour le tableau 7, seuls les résultats correspondant à une catégorie retenue dans le schéma d'annotation du chapitre 3, en l'occurrence colère, peur, surprise (négative) et tristesse sont reportés dans le tableau.

Des deux travaux ayant une approche similaire à celle que j'ai moi-même mise en œuvre (*cf.* section 3.1), seul E. Kim et R. Klinger (2018) proposent un système d'analyse automatique des émotions. Deux expériences sont menées à partir du corpus annoté qu'ils ont constitué. La première vise à développer un classifieur en émotions s'appliquant à des triplets de phrases. Cet objectif est proche de celui visé dans le présent travail (*cf.* chapitre 5). Deux modèles sont élaborés pour y répondre : un modèle symbolique qui s'appuie sur le *NRC lexicon* et un perceptron multi-couche (MPL) qui prend en entrée des sacs de mots. Les résultats obtenus par ces modèles sont présentés Tableau 9 en terme de précision, de rappel et de F-mesure, toutes classes confondues (les résultats par classe ne sont pas présentés dans (Kim & Klinger, 2018)). La seconde expérience consiste au développement d'un système pour retrouver les séquences exprimant une émotion. Le meilleur modèle obtient une F-mesure de 0,48 (modèle biLSTM-CRF avec des plongements de mots *fasttext* en entrée).

Modèle	**Lexique**	**Nb cla.**	**P**	**R**	**F**
symbolique	NRC lexicon	9	0,19	0,83	0,31
MLP	non	9	0,55	0,21	0,31

Tableau 9: Performances des classifieurs en émotions de (Kim & Klinger, 2018)

L'analyse automatique des émotions constitue ainsi une tâche de TAL complexe, que même les modèles de type Transformeur ne parviennent pas à résoudre totalement. Comme le souligne E. Öhman (2020), cela n'est pas nécessairement dû aux limites des techniques de TAL mais plutôt à la tâche elle-même : « ce type de tâche représente un défi pour les annotateurs humains. Si les annotateurs humains ne peuvent s'accorder sur les étiquettes, il n'est pas raisonnable de penser que les ordinateurs peuvent faire mieux » (2020, p. 9).

4) Propositions méthodologiques

La diversité des notions entrant dans la caractérisation des émotions en (psycho-)linguistique (*cf.* sections 1 & 2) et le fait que seules certaines aient été interviennent jusqu'ici en TAL pour l'analyse automatique des émotions (*cf.* section 3) soulignent la difficulté de cerner les émotions telles qu'elles sont effectivement réalisées dans les textes, lorsqu'on souhaite de plus ne pas se focaliser exclusivement sur le lexique émotionnel. L'approche mise en œuvre dans le présent travail s'appuie sur une modélisation qui vise à prendre en compte de manière explicite tous les concepts qui sous-tendent l'exploration des émotions et à organiser ces concepts de sorte à déterminer les liens qui les unissent. Cette modélisation intègre donc des notions pertinentes à la fois sur le plan psycholinguistique pour rendre compte des étapes développementales de compréhension chez les enfants, des notions pertinentes sur le plan linguistique pour permettre une caractérisation fine des émotions et des notions qui fassent sens dans la perspective du développement d'un outil d'analyse automatique des émotions.

La modélisation des émotions ainsi proposée prend la forme d'un schéma d'annotation permettant d'étudier les émotions des personnages des textes ainsi que celles des scripteurs – et non celles du lecteur – qu'elles soient explicitement ou implicitement dénotées par le texte. Des études linguistiques (section 2) et psycholinguistiques (section 1), quatre *modes d'expression des émotions* sont retenus : les labels émotionnels, les caractéristiques linguistiques d'un énoncé montrant l'émotion du locuteur, la description d'une situation qui suggère une émotion et la description d'un comportement émotionnel. Ces quatre modes d'expression sont nommés respectivement : désigné, montré, suggéré et comportemental. Les notions d'*expérienceur* et de *catégorie émotionnelle*, également mobilisées dans les études (psycho-)linguistiques interviennent aussi dans le schéma pour caractériser l'expression des émotions. Des études psycholinguistiques plus spécifiquement émergent enfin les notions de *cause* et de *consé-*

quence de l'émotion pour décrire une émotion. J'inclus donc ces deux notions dans le schéma d'annotation.

Étant donné l'intérêt que représente une étude conjointe des dimensions TAME et de l'expression des émotions pour mieux cerner comment ces dimensions sémantiques interagissent, j'ai en outre choisi d'inclure dans le schéma d'annotation une partie dédiée au repérage de diverses notions TAME (*adverbiaux temporels*, *marqueurs de modalité, marqueurs de prise en charge*).

Conclusion

De la linguistique et de la psycholinguistique se dégagent de nombreuses notions pertinentes pour caractériser l'expression des émotions dans toute sa complexité : *catégorie émotionnelle, type d'émotion mode d'expression, expérienceur, cause et conséquence de l'émotion*. Jusqu'à présent, aucun travail de TAL ne s'est cependant efforcé de considérer ensemble et de manière explicite toutes ces notions pour annoter les émotions dans les textes et proposer des classifieurs à même d'analyser automatiquement les émotions. Le schéma introduit chapitre 3 a été constitué pour remplir ce besoin en prenant en compte autant de concepts essentiels à l'expression des émotions que possible, ainsi qu'en précisant les relations qu'ils entretiennent. La priorité principale étant de constituer un schéma exhaustif, l'accent a été mis sur le fait de tester l'applicabilité du schéma, plutôt que sur la mise en œuvre directe d'un calcul d'accord inter-annotateurs. Le schéma a ainsi été employé par des annotateurs experts entraînés sur un corpus composé de plusieurs genres de textes. Le premier corpus francophone multi-genre annoté en émotions et comportant des informations sur l'expérienceur, la cause et la conséquence de l'émotion à l'échelle du marqueur linguistique a ainsi été constitué (*cf.* chapitre 4). Ce corpus a permis le développement de classifieurs pour l'analyse automatique de la dimension émotionnelle des textes (*cf.* chapitre 5) et de nombreuses observations linguistiques sur l'expression des émotions dans les textes (*cf.* chapitre 7).

Chapitre 3 Schéma d'annotation manuelle des émotions

Des perspectives d'analyse mise en œuvre dans des études linguistiques et psycholinguistiques portant sur l'expression des émotions, j'ai retenu six notions pour modéliser les émotions : la *catégorie émotionnelle*, le *type d'émotion*, le *mode d'expression*, l'*expérienceur de l'émotion*, sa *cause de l'émotion* et sa *conséquence* (*cf.* chapitre 4). Cette modélisation, qui se réalise sous la forme d'un schéma d'annotation, est exposée dans le présent chapitre.

Dans un premier temps, quatre unités linguistiques distinctes sur lesquelles s'appuie la modélisation pour articuler les six notions retenues dans mon approche seront définies (section 1). Le schéma d'annotation qui réalise concrètement la modélisation des émotions est ensuite décrit (section 2). J'exposerai enfin une seconde partie du schéma qui vise à annoter plusieurs notions en lien avec l'analyse de la temporalité linguistique (*adverbiaux temporels, marqueurs de modalité, marqueurs de prise en charge*) en interaction avec l'expression des émotions (section 3).

1) Quatre unités linguistiques pour caractériser l'expression linguistique des émotions

Quatre types d'unités linguistiques sont distingués dans la modélisation : celle exprimant une émotion, celle désignant l'entité qui ressent l'émotion, celle décrivant une situation qui cause l'émotion et celle dénotant une situation causée par l'émotion.

1.1) L'unité linguistique exprimant l'émotion

L'unité linguistique exprimant l'émotion est l'unité fondamentale de la modélisation de la notion d'émotion. Elle articule à la fois les notions de *mode d'expression*, de *catégorie émotionnelle* et de *type d'émotion*.

Mode d'expression. Il est question de s'intéresser ici à la manière dont une émotion est marquée linguistiquement dans un texte, en prenant en compte toute la diversité des manières d'exprimer l'émotion, allant du « simple » lexique émotionnel à la des-

cription d'une situation conventionnellement associée à une émotion. L'approche développée dans le schéma poursuit en cela celle proposée en linguistique par R. Micheli (Micheli, 2014) ou en psycholinguistique par N. Blanc (Blanc, 2010; Creissen & Blanc, 2017) (*cf.* chapitre 2). Des exemples de marqueurs que le schéma vise à capter sont soulignés dans les énoncés [6] à [9]. Ils concourent tous à exprimer une émotion (ici, la joie) d'un protagoniste (ici, Paul). Ils vont du strictement lexical isolé (l'adjectif « heureux » dans [6], le verbe « sourit » dans [7], l'interjection « youpi » dans [9]) ou non isolé (les noms « victoire » et « médaille » en [9] qui font partie de syntagmes plus larges), à des marques typographiques (la marque de ponctuation « ! » de [9]) en passant par des syntagmes verbaux (« a gagné la course » dans [8]) et des constructions syntaxiques (énoncé averbal « Enfin une victoire » et structure clivée « C'est la médaille que j'attendais depuis longtemps » en [9]).

[6] Paul est heureux.
[7] Paul sourit.
[8] Paul a gagné la course.
[9] « Youpi ! Enfin une victoire ! C'est la médaille que j'attendais depuis longtemps ! », dit Paul.

L'emploi de ces divers marqueurs linguistiques[24] – soulignés dans l'ensemble des exemples qui seront présentés dans ce chapitre – correspond aux divers *modes d'expression* des émotions pointés d'une part dans (Micheli, 2014) et d'autre part dans (Blanc, 2010). (Micheli, 2014, p. 17) définit trois modes d'expression d'une émotion :

- l'émotion « dite », désignée par des termes du lexique émotionnel. Par exemple, la colère est désignée par le segment « Une rage » dans l'énoncé [10] ;
- l'émotion « étayée », « inférée à partir de la représentation, dans le discours, d'un type de situation qui lui est conventionnellement associé sur le plan socio-culturel et qui est donc supposé lui servir de fondement ». La situation décrite en [11] fait par exemple état d'actes (marqués par successivement par « massacre », « incendie », « pille », « extermine », « égorge », « vend les petites filles et les petits garçons ») associables de manière socio-culturelle à une réaction émotionnelle négative ;
- l'émotion « montrée », cette fois « inférée à partir d'un ensemble de caractéristiques de l'énoncé » qui fonctionnent comme « des indices du fait que l'énonciation est cooccurrente avec le ressenti d'une émotion par le locuteur ». Ces

24 Le terme de marqueur linguistique est à entendre comme toute combinaison simple ou complexe de formes linguistiques.

caractéristiques correspondent aussi bien à des termes du lexique (par ex. interjections) qu'à des structures syntaxiques (par ex. énoncés averbaux) voire des marques typographiques (par ex. point d'exclamation marquant un énoncé exclamatif), comme l'illustre l'exemple [12] où les différents segments qui montrent la peur de l'énonciateur sont soulignés. L'association des marqueurs qui montrent l'émotion à une catégorie émotionnelle précise (joie, peur, *etc.*) ne s'effectue qu'au moyen d'un ensemble de caractéristiques qui convergent vers une même émotion. L'interprétation d'une émotion montrée, nécessite donc d'autres indices émotionnels (par ex. une émotion désignée par du lexique ou suggérée par la description d'une situation).

[10] Une rage m'a pris contre ce peuple.[25]

[11] C'est qu'à l'heure qu'il est, tout près de nous, là, sous nos yeux, on massacre, on incendie, on pille, on extermine, on égorge les pères et les mères, on vend les petites filles et les petits garçons.[26]

[12] Hélas ! qu'est-ce que la mort fait à notre âme ? quelle nature lui laisse-t-elle ? qu'a-t-elle à lui prendre ou lui donner ? où la met-elle ? (...) Ah ! un prêtre ! un prêtre qui sache cela ! Je veux un prêtre, et un crucifix à baiser ![27]

(Blanc, 2010, p. 258) distingue également trois façons d'exprimer les émotions :

- l'émotion « désignée », nommée de manière explicite, comme dans l'exemple [13] où le segment « furieux » nomme la colère de Eudes ;
- « l'expression comportementale de l'émotion », correspondant à la description d'un comportement qui traduit l'émotion ressentie par le personnage, tel que « éclate en sanglots » traduit la tristesse de « Elle » en [14] ;
- l'émotion « inférée » reposant sur la présence d'un « évènement inducteur d'émotion », comme l'illustre l'exemple [15] où la situation décrite (remise de récompense et attribution d'une fonction) permet d'inférer la joie du lapin Mudubina.

[13] Nous avons tous applaudi, sauf Eudes qui était furieux [...].[28]

[14] Elle éclate en sanglots.[29]

25 Extrait du *Dernier jour d'un condamné* de V. Hugo, cité par (Micheli, 2014 : 34).
26 Extrait de « Pour la Serbie » de V. Hugo, cité par (Micheli, 2014 : 73).
27 Extrait du *Dernier jour d'un condamné* de V. Hugo, cité par (Micheli, 2014 : 63).
28 Extrait du *Petit Nicolas*, de R. Goscinny cité par (Creissen, Blanc, 2017 : 275).
29 Exemple repris de (Blanc, 2010 : 258).

[15] Le puma remit la récompense promise au lapin Mudubina et le nomma Premier Ministre.[30]

Les correspondances entre les typologies de (Micheli, 2014) et (Blanc, 2010) sont représentées tableau 10. Ces deux typologies ont constitué le socle de celle finalement retenue au sein du schéma d'annotation des émotions présenté dans (Étienne & Battistelli, 2021). Dans ce schéma d'annotation sont ainsi distingués les quatre modes d'expression suivants auxquels sont associés divers types de marqueurs linguistiques :

- les émotions *désignées* par un terme du lexique émotionnel (marqueurs linguistiques de niveau lexical) ;
- les émotions *comportementales*, véhiculées par la description d'un comportement (marqueurs linguistiques de niveau lexical ou propositionnel) ;
- les émotions *suggérées* qui doivent être inférées de la description d'une situation conventionnellement associée à une émotion (marqueurs linguistiques de niveau lexical ou propositionnel) ;
- et les émotions *montrées* par des caractéristiques de l'énoncé relevant de plusieurs niveaux d'analyse (marqueurs linguistiques de niveaux lexical (par ex. interjections), syntaxique (par ex. énoncés averbaux) et typographique (par ex. « ... »)).

	Typologie des modes d'expression d'une émotion			
d'après (Blanc, 2010)	Émotion désignée *ex.* [8]	Expression comportementale de l'émotion *ex.* [9]		Émotion suggérée *ex.* [10]
d'après (Micheli, 2014)	Émotion dite *ex.* [5]		Émotion montrée *ex.* [7]	Émotion étayée *ex.* [6]
d'après (Etienne, Battistelli, 2021)	Émotion désignée *ex.* [1]	Émotion comportementale *ex.* [2]	Émotion montrée *ex.* [4]	Émotion suggérée *ex.* [3]

Tableau 10: Typologies des modes d'expression des émotions

La prise en compte de plusieurs modes d'expression des émotions représente une étape importante dans l'approche défendue dans le schéma pour plusieurs raisons. En premier lieu, comme nous venons de le voir, elle est nécessaire pour l'investigation de l'hétérogénéité des marqueurs des émotions dans un texte. En second lieu, elle aug-

30 Extrait des *Contes des Amériques*, cité par (Blanc, 2010 : 258).

mente le nombre de segments textuels potentiellement porteurs d'une information émotionnelle dans un texte, permettant un champ d'analyse plus vaste et approfondi.

En plus du mode d'expression de l'émotion, le schéma d'annotation caractérise une unité émotionnelle par la *catégorie émotionnelle* et le *type d'émotion* qu'elle exprime. Le choix des catégories émotionnelles lors de l'élaboration d'un schéma d'annotation, puis leur identification lors du processus d'annotation, représentent un nœud de difficulté car les frontières entre émotions sont souvent floues (cf. chapitre 2). Le schéma retient onze catégories émotionnelles, réparties entre émotions de base et émotions complexes. Par convention, ces catégories sont nommées en petites majuscules pour faire références aux catégories conceptuelles du schéma. Les six émotions de base sont celles introduites dans (Ekman, 1992) mais également fréquemment reprise à la fois dans les études psycholinguistiques (*cf.* chapitre 1) et TAL (*cf.* chapitre 2). Quatre émotions complexes (CULPABILITÉ, EMBARRAS, FIERTÉ, JALOUSIE) sont reprises de (Davidson, 2006) et (Blanc & Quenette, 2017). La catégorie ADMIRATION a été ajoutée comme cinquième émotion complexe, afin de mieux équilibrer les émotions de base et émotions complexes dans le schéma. Chacune de ces onze catégories correspond à des émotions plus spécifiques. Par exemple, la catégorie COLÈRE regroupe les émotions d'agacement, d'énervement, de rage ou encore de fureur.

1.2) L'unité linguistique désignant l'expérienceur

Conjointement à l'unité exprimant l'émotion, le schéma repère l'unité linguistique correspondant à l'*expérienceur*, c'est-à-dire l'entité (être humain, groupe d'êtres humains, animal, *etc.*) qui ressent l'émotion. Ces unités sont repérées par un surlignage bleu clair dans les énoncés [16] à [18] et dans le reste du chapitre. La notion d'expérienceur fait ainsi partie intégrante de la description d'une émotion dans le schéma. Ceci représente un parti pris théorique important. Il résulte de la volonté de prendre en compte plusieurs travaux linguistiques consacrés à l'analyse sémantique des émotions qui caractérisent l'expression des émotions à l'aide de cette notion (*cf.* chapitre 2). Ces travaux mettent l'accent sur les rôles sémantiques impliqués dans la description d'une unité lexicale émotionnelle (en réponse à des questions du type : qui ressent et par quoi est causée une émotion). Selon cette approche, une situation émotionnelle de PEUR telle que celle de l'exemple [18] est alors analysée en distinguant bien les deux entités impliquées, le chat de Paul et Paul, qui occupent effectivement deux rôles sémantiques distincts : le chat ressent la peur (il est l'*expérienceur*) tandis que Paul la

cause. Exceptés quelques travaux, par exemple (Kim & Klinger, 2018), l'usage en TAL est pourtant de repérer exclusivement l'unité émotionnelle « est effrayé » (*cf.* chapitre 2). Or ce repérage exclusif pourrait aboutir à un contre-sens du point de vue de la compréhension, avec, dans le cas de cet exemple, une interprétation selon laquelle l'état émotionnel des deux personnages serait le même. Une analyse de l'expression des émotions visant à tenir compte de facteurs de compréhension, telle que nous ambitionnons de le faire au sein du projet TextToKids, nécessite ainsi l'étude de l'expression de l'expérienceur.

[16] La famille de Paul est heureuse.
[17] La nouvelle a surpris Camille.
[18] Le chat de Paul est effrayé parce que Paul a sorti l'aspirateur.

1.3) Les unités linguistiques désignant la cause et la conséquence de l'émotion

En s'appuyant sur des études psycholinguistiques (*cf.* chapitre 2), le schéma s'intéresse aux unités linguistiques exprimant la cause ou la conséquence d'une émotion afin de mettre en évidence le fait que, au sein d'un texte, des segments émotionnels entrent en relation sémantique avec d'autres segments exprimant des causes de l'émotion ou ses conséquences.

Prenons l'exemple [19]. Le ressenti émotionnel du personnage de Julie est directement désigné par le terme du lexique émotionnel « jalousie », qui apparaît souligné dans [19]. La situation décrite dans la première phrase de [19] explicite la cause de cette émotion (notée en gras). La conséquence de cette émotion correspond quant à elle à la situation exposée dans la dernière phrase de [19] (notée en italique). Seule l'émotion éprouvée par le personnage assure ici un lien de causalité entre les trois phrases de l'exemple.

[19] **Marc a eu une plus grosse prime de fin d'année que Julie**. Pendant des mois, elle a ruminé sa jalousie. Hier, *elle a crevé les pneus de la voiture de Marc.*

Le guide (Étienne & Battistelli, 2021) qui accompagne le schéma indique un lien intéressant entre la notion de cause et celle d'émotion suggérée : lors de l'annotation de textes à l'aide du schéma, les segments étiquetés comme cause d'une émotion apparaissent régulièrement comme étiquetés par ailleurs comme relevant d'une émotion dite suggérée. Ce phénomène n'est en réalité pas très surprenant et se déduit aisément

des définitions des émotions suggérées et étayées (*cf.* section 1.1). Ainsi, du point de vue psycholinguistique, un segment qui suggère une émotion constitue une base sur laquelle le lecteur s'appuie pour effectuer une « inférence émotionnelle » (Creissen & Blanc, 2017, p. 266), c'est-à-dire pour déduire que le personnage ressent une émotion. Du point de vue linguistique, un segment qui suggère une émotion est explicitement posé comme exprimant le « fondement » (Micheli, 2014, p. 17) de l'émotion, ou encore les « causes » (*ibid.*, p. 120) à partir desquelles l'interlocuteur infère une certaine émotion. Par exemple, dans la première phrase de [20], le segment souligné « a gagné la course » suggère l'émotion de JOIE ressentie par l'entité « Paul »[31]. Cette première phrase exprime également la cause de cette émotion : le fait d'avoir gagné la course (proposition en gras « Paul a gagné la course »).

Le guide indique un second recoupement, cette fois entre la notion de conséquence et celle de mode comportemental. Ce lien s'ancre une fois encore dans la définition du mode d'expression. En psycholinguistique, un comportement émotionnel est vu comme « tradui[sant] » l'émotion ressentie par un personnage (Creissen & Blanc, 2017, p. 267). Autrement dit, le comportement du personnage découle de l'émotion ressentie. Le lien de causalité unissant ressenti émotionnel et comportement émotionnel est également perceptible dans l'approche linguistique de (Micheli, 2014). Il y est précisé en effet que l'interprétation d'un segment renvoyant à des expressions physiologiques ou comportementales de l'émotion nécessite une inférence de la part de l'allocutaire. La dernière phrase de [20] décrit par exemple le comportement émotionnel « sourit », qui peut être considéré comme la conséquence de la joie éprouvée par Paul (proposition « Il sourit », en italique dans l'exemple).

[20] **Paul** **a gagné la course**. Il est heureux. *Il* *sourit*.

2) L'annotation des unités linguistiques

Les quatre types d'unités linguistiques introduits en section 1 font l'objet d'un étiquetage informationnel encodé dans le schéma d'annotation. Cet étiquetage a été formalisé selon les principes de la plateforme d'annotation manuelle Glozz (Widlöcher & Mathet, 2012), retenue pour la flexibilité des structures d'annotation qu'elle permet de constituer (section 2.1). Chaque unité linguistique a ainsi été traduite en une unité d'annotation spécifique, pouvant être reliée à d'autres unités par un ensemble de relations défini dans le schéma. L'unité exprimant l'émotion est ainsi désormais notée Si-

31 Cette émotion est désignée par l'adjectif « heureux » souligné dans la deuxième phrase de [15].

tEmo et celle exprimant l'expérienceur est noté Expérienceur (section 2.2). Celles exprimant la cause et la conséquence d'une émotion sont respectivement notées SitCause et SitConséquence (section 2.3).

Les exemples présentés dans cette section reprennent le même type de marquage que celui de la section 1. Les unités SitEmo seront repérées par un soulignage simple et les unités Expérienceur par un surlignage bleu clair. Les marqueurs linguistiques correspondant à une unité SitCause apparaîtront en gras et ceux correspondant à une unité SitConseq en italique. Aucune mise en forme ne permet de rendre compte de manière satisfaisante des relations unissant plusieurs unités, elles ne seront donc pas représentées visuellement dans les exemples.

La figure 8 donne une vue synthétique des unités et relations qui forment le schéma d'annotation.

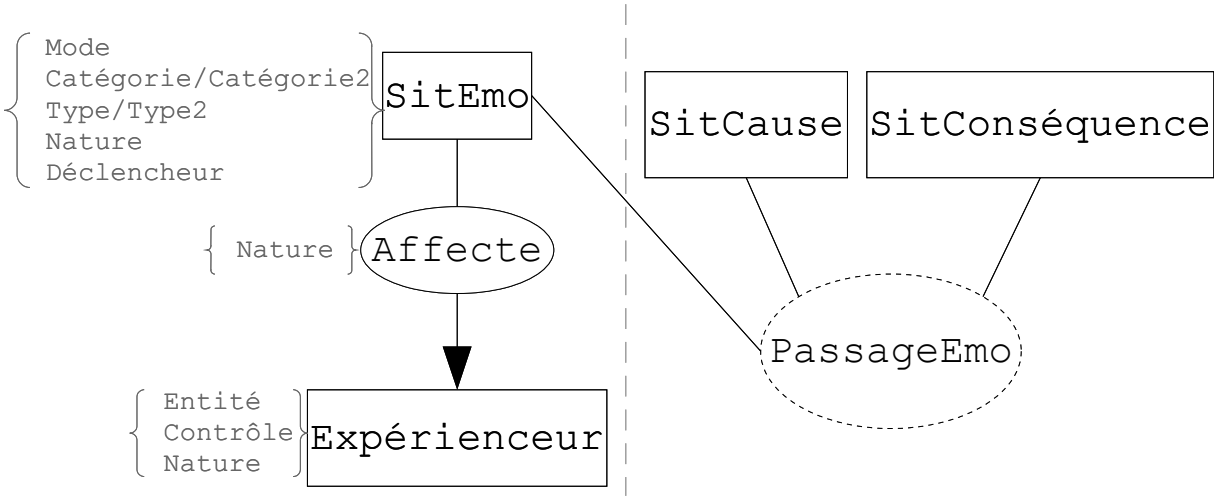


Figure 8: Unités et relations du schéma d'annotation

2.1) Glozz : un outil flexible pour l'annotation manuelle de textes

La plateforme d'annotation manuelle Glozz (Widlöcher & Mathet, 2012) repose sur la définition de trois structures d'annotation : les unités qui captent les segments textuels ; les relations qui lient deux structures d'annotation entre elles ; et les schémas qui lient un nombre illimité de structures d'annotation. Des informations peuvent être associées à ces trois structures grâce à des systèmes de traits. En plus de cette possibilité de lier des unités entre elles et de typer finement les structures d'annotation produites, Glozz offre une très grande flexibilité sur la taille des unités pouvant être

délimitées (caractères, mots, phrases, groupes de phrases, paragraphes, *etc.*), tout en permettant des chevauchements et emboîtements d'unités. La plateforme Glozz se prête donc particulièrement bien à une tâche d'annotation telle que celle visée ici, nécessitant l'analyse de marqueurs de taille très variable (allant de la simple marque de ponctuation à la proposition) et la mise en relation de différentes unités (par ex. unité émotionnelle et unité exprimant l'expérienceur). Elle a donc été retenue pour mettre en œuvre la modélisation de la notion d'émotion introduite section 1. Les principes régissant Glozz ont guidé la traduction de la modélisation en un véritable schéma d'annotation. Sa mise en application est ensuite menée *via* un processus d'annotation manuelle par des linguistes. Ce processus s'appuie sur un ensemble de règles et de critères pour l'aide à la décision qui garantissent une application du schéma la plus homogène possible. Les grandes lignes de cet ensemble, par ailleurs détaillé dans le guide qui accompagne le schéma (Étienne & Battistelli, 2021), seront présentées au chapitre 4.

2.2) Les unités SitEmo et Expérienceur

Les unités SitEmo, qui captent les unités émotionnelles présentes dans le texte, sont spécifiées par sept traits dont l'un correspond à la nature du segment textuel annoté (syntagme nominal, proposition, *etc.*), comme illustré par le tableau 11.

Trait	**Valeurs**
Mode	comportemental, désigné, montré, suggéré
Type & Type2	Base, Complexe, Aucun
Catégorie & Catégorie2	ADMIRATION, COLÈRE, *etc.*, Aucune
Déclencheur	Cœur de l'expression émotionnelle
Nature	Syntagme nominal, syntagme prépositionnel, *etc.*

Tableau 11: Traits des unités SitEmo

Trait *Mode* des unités SitEmo. Le trait *Mode* intervient pour renseigner le mode d'expression employé pour exprimer l'émotion de l'unité linguistique annotée. Une unité SitEmo ne peut avoir qu'une seule valeur de mode, parmi comportemental, désigné, montré et suggéré.

Traits *Catégorie, Catégorie2, Type* et *Type2* des unités SitEmo. Les traits *Catégorie* et *Type* indiquent respectivement la catégorie émotionnelle et le type d'émotion exprimés par les marqueurs annotés. Le schéma prend en compte onze catégories émotionnelles, départies entre émotions de base (COLÈRE, DÉGOÛT, JOIE, PEUR, SURPRISE, TRISTESSE) et émotions complexes (ADMIRATION, CULPABILITÉ, EMBARRAS, FIERTÉ, JALOUSIE). L'annotateur choisit obligatoirement une de ces catégories pour remplir la valeur du trait *Catégorie* d'une unité SitEmo. Lorsqu'une unité linguistique exprime un ressenti émotionnel ambigu, le trait *Catégorie2* permet à l'annotateur d'associer une deuxième catégorie à l'unité SitEmo. Il en va de même pour les traits *Type* et *Type2*. Les valeurs possibles pour les trait *Catégorie2* et *Type2* sont respectivement les mêmes que pour les traits *Catégorie* et *Type*, avec l'ajout des valeurs « Aucune »/« Aucun » qui interviennent par défaut lorsque l'unité émotionnelle ne dénote qu'une seule catégorie émotionnelle selon l'annotateur. Lorsque deux catégories émotionnelles (et deux types) sont renseignées, aucune hiérarchie n'est établie entre les deux : dans les analyses, les deux catégories émotionnelles et types d'émotions sont donc considérés à égalité.

Bien que les onze catégories du schéma soient à entendre comme des macro-catégories, regroupant plusieurs émotions spécifiques (par ex. plaisir, bonheur ou euphorie regroupés sous JOIE), cela ne suffit pas à rendre compte de toute la diversité des ressentis émotionnels présents dans les textes. Une unité appelée Autre a été définie. Elle est utilisée pour annoter des marqueurs exprimant d'autres émotions, telles que le MÉPRIS, l'AMOUR ou la HAINE. Cette unité n'est spécifiée que par un seul trait, *Remarque*, afin de laisser l'opportunité à l'annotateur de renseigner l'émotion exprimée par l'unité. Les expérienceurs, cause et conséquences de ces unités ne doivent pas être annotés.

Trait *Déclencheur* d'une unité SitEmo. Le trait *Déclencheur* d'une unité SitEmo spécifie le ou les éléments qui, au sein des marqueurs annotés, sont les plus saillants pour exprimer l'émotion. Autrement dit, quel(s) élément(s) constitue(nt) le cœur de l'expression émotionnelle. Par exemple, le déclencheur d'une émotion désignée par le marqueur « très heureux » sera « heureux ». Pour une émotion suggérée par le marqueur « a reçu beaucoup de cadeaux », le déclencheur sera « cadeaux » voire « beaucoup de cadeaux ». Ce trait, renseigné en saisie libre par l'annotateur, a été conçu pour permettre une analyse linguistique plus fine des expressions émotionnelles repérées dans les textes. S'il n'a pas à ce jour eu cet usage, il pourrait cependant s'avérer

précieux pour constituer des ressources utiles à la détection des émotions, par exemple des lexiques.

L'unité Expérienceur. Les unités Expérienceur sont spécifiées par trois traits, dont un renvoie à la nature du segment textuel annoté, à un grain de typage plus petit que pour les unités SitEmo (nom propre, pronom personnel, pronom relatif, *etc.*), comme illustré tableau 12. La majorité des expérienceurs étant désignés par des syntagmes nominaux, il semblait en effet pertinent d'opérer ici un typage plus fin.

Trait	**Valeurs**
Entité	Entité qui ressent l'émotion
Contrôle	Humain, Animé-Non Humain, Inanimé
Nature	Nom propre, pronom personnel, *etc.*

Tableau 12: Traits de l'unité Expérienceur

Trait *Entité* des unités Expérienceur. Le trait *Entité* d'une unité Expérienceur encode l'entité, soit la personne, l'animal, le groupe, *etc.* qui ressent l'émotion. Ce trait est renseigné en saisie libre par l'annotateur. Grâce au remplissage de ce trait, il devient possible pour l'annotateur de rendre compte des liens de coréférence qu'il a repérés entre différentes unités Expérienceur. Les entités annotées correspondent à des personnages des textes (et non du lecteur), puisqu'il s'agit là du focus du schéma d'annotation. Deux entités un peu particulières sont également prises en compte par le schéma : le scripteur et la *doxa*. Cette dernière renvoie à une entité abstraite qui représente les croyances, les opinions populaires. Cette entité se retrouve typiquement dans les formules impersonnelles.

Trait *Contrôle* des unités Expérienceur. Le trait *Contrôle* d'une unité Expérienceur indique le type de l'entité qui ressent l'émotion, parmi trois valeurs possibles : Humain, Animé – Non Humain et Inanimé. Le scripteur et la *doxa* sont systématiquement considérés comme Humain.

Lier les unités SitEmo à leur Expérienceur : la relation Affecte. Les unités SitEmo et leurs unités Expérienceur associées sont liées par une relation nommée Affecte. Cette relation est spécifiée par un seul trait indiquant la nature de la relation syntaxique qui unit les deux segments textuels captés par les unités SitEmo et Expé-

rienceur. À travers l'usage des unités Expérienceur et de la relation Affecte, l'objectif est d'atteindre une compréhension plus fine de l'expression linguistique des émotions dans les textes. Dans les exemples ci-dessous, la relation Affecte est représentée par une flèche unissant une unité SitEmo à une unité Experienceur dans les informations d'annotation détaillées sous le texte d'exemple.

L'énoncé [21] comporte une unité émotionnelle ambiguë. L'unité « a réussi cet exercice difficile tout seul » suggère en effet à la fois la JOIE et la FIERTÉ de l'entité « Paul ». Elle sera ainsi caractérisée par un trait *Mode* de valeur « Suggéré », un trait *Catégorie* de valeur « Joie » associé au trait *Type* de valeur « Base » et un trait *Catégorie2* de valeur « Fierté » associé au trait *Type2* de valeur « Complexe ». L'unité Expérienceur est quant à elle caractérisée par le trait *Entité* de valeur « Paul », le trait *Contrôle* de valeur « Humain » et le trait *Nature* de valeur « NP » pour Nom Propre.

[21] Paul a réussi cet exercice difficile tout seul.

Experienceur<Paul>{Entité : Paul, Contrôle : Humain, Nature : NP}

SitEmo<a réussi cet exercice difficile tout seul>{Mode : Suggérée, Type : Base, Catégorie : Joie, Type2 : Complexe, Catégorie2 : Fierté, Nature : Proposition}

Relations Affecte :

<Paul> <---- <a réussi cet exercice difficile tout seul>

L'exemple [22] montre plusieurs unités SitEmo qui captent la PEUR du même personnage, « Inès », exprimée à travers plusieurs modes d'expression. Il illustre également les chaînes de coréférence pouvant être retrouvées lors de l'annotation des unités Expérienceur.

[22] Inès se souvient des murs qui n'arrêtaient pas de faire « crac ». Ça lui faisait très peur. « La maison allait tomber sur nous ... » murmure-t-elle en se blottissant sur le canapé.[32]

Experienceur<Inès>{Entité : Inès, Contrôle : Humain, Nature : NP}

SitEmo<des murs qui n'arrêtaient pas de faire « crac »>{Mode : Suggéré, Type : Base, Catégorie : Peur, Type2 : Aucune, Catégorie2 : Aucune, Nature : SP, Déclencheur : faire « crac »}

Experienceur<lui>{Entité : Inès, Contrôle : Humain, Nature : ProPers}

32 Exemple extrait de « L'immeuble d'Inès et d'Adam menaçait de s'effondrer », *P'tit Libé*, n°91, 1er au 7 février 2019.

SitEmo<très peur>{Mode : Désigné, Type : Base, Catégorie : Peur, Type2 : Aucune, Catégorie2 : Aucune, Nature : SN, Déclencheur : peur}

SitEmo<La maison allait tomber sur nous>{Mode : Suggéré, Type : Base, Catégorie : Peur, Type2 : Aucune, Catégorie2 : Aucune, Nature : Proposition, Déclencheur : tomber sur nous}

SitEmo<…>{Mode : Montré, Type : Base, Catégorie : Peur, Type2 : Aucune, Catégorie2 : Aucune, Nature : Points de suspension, Déclencheur : …}

SitEmo<murmure>{Mode : Comportemental, Type : Base, Catégorie : Peur, Type2 : Aucune, Catégorie2 : Aucune, Nature : Proposition, Déclencheur : murmure}

Experienceur<elle>{Entité : Inès, Contrôle : Humain, Nature : ProPers}

SitEmo<en se blottissant sur le canapé>{Mode : Comportemental, Type : Base, Catégorie : Peur, Type2 : Aucune, Catégorie2 : Aucune, Nature : Proposition, Déclencheur : se blottissant}

Relations Affecte :

<Inès> <--- <des murs qui n'arrêtaient pas de faire "crac">

<lui> <--- <très peur>

<La maison allait tomber sur nous> ---> <elle>

<...> ---> <elle>

<murmure> ---> <elle>

<elle> ---> <en se blottissant sur le canapé>

2.3) Les unités SitCause et SitConséquence

L'annotation de la *cause* et de la *conséquence* des émotions n'a jusqu'ici été formalisée dans le guide (Étienne & Battistelli, 2021) que de manière succincte. Cette formalisation repose sur deux unités, les unités SitCause et SitConséquence, identifiant respectivement les marqueurs qui expriment une situation qui cause une émotion et ceux dénotant une situation qui résulte d'une émotion. Lors de l'annotation de textes, ces unités n'ont aujourd'hui été appliquées que dans leur version préliminaire pour annoter la cause et la conséquence à l'échelle de la phrase graphique. La formalisation repose de plus sur une structure d'annotation de Glozz de type schéma appelée PassageEmo qui intervient pour lier entre elles toutes les unités SitEmo participant de l'expression d'une même émotion (c'est-à-dire une même catégorie émotionnelle ressentie par un même expérienceur) ainsi que les unités SitCause et SitConséquence dénotant les causes et conséquences de cette émotion. Pensés pour un usage exploratoire, les unités SitCause, SitConséquence et le schéma PassageEmo ne sont spécifiés que par

un seul trait, *Remarque*, permettant à l'annotateur de renseigner, en saisie libre, toute remarque éventuelle liée à l'annotation produite.

Dans l'exemple [23], l'EMBARRAS ressenti par l'expérienceur « diable » est exprimé à travers six unités linguistiques. Deux unités SitEmo verbalisent la catégorie émotionnelle EMBARRAS (une la suggère tandis que l'autre la désigne). Chaque SitEmo est associée à une unité Expérienceur référant au diable. De plus, la première phrase de l'exemple décrit la cause de l'émotion (le fermier a vendu de beaux navets au marché, alors que le diable n'avait que des feuilles fanées à proposer) et est ainsi capturée par une unité SitCause. Enfin, la seconde phrase de l'exemple, marquée par une unité SitConséquence, dépeint la conséquence de l'émotion (le diable est parti, sans qu'on ne le revit jamais). L'ensemble de ces unités SitEmo, SitCause et SitConséquence est lié par une structure PassageEmo, puisqu'elles participent toutes de l'expression de l'EMBARRAS du diable.

[23] **Au marché, le paysan les vendit facilement, tandis que le diable ne récolta que moqueries pour ses feuilles fanées.** *Il en fut tellement vexé qu'on ne le revit jamais plus au pays*[33].

Experienceur<le diable>{Entité : le diable, Contrôle : Animé-Non Humain, Nature : SN}

SitEmo<moqueries>{Mode : Suggéré, Type : Complexe, Catégorie : Embarras, Type2 : Aucun, Catégorie2 : Aucune, Déclencheur : moqueries, Nature : SN}

Experienceur<Il>{Entité : le diable, Contrôle : Animé-Non Humain, Nature : ProPers}

SitEmo<tellement vexé qu'on ne le revit jamais plus au pays>{Mode : Désigné, Type : Complexe, Catégorie : Embarras, Type2 :Base, Catgorie2 : Colere, Déclencheur : vexé, Nature : SAdj}

SitCause<Au marché, le paysan les vendit facilement, tandis que le diable ne récolta que moqueries pour ses feuilles fanées.>

SitConsequence<Il en fut tellement vexé qu'on ne le revit jamais plus au pays.>

Relations Affecte :

<le diable> <--- <moqueries>

<Il> <--- <tellement vexé qu'on ne le revit jamais plus au pays>

Structure PassageEmo :

<moqueries> ; <tellement vexé qu'on ne le revit jamais plus au pays>; <Au marché, le paysan les vendit facilement, tandis que le diable ne récolta que moqueries pour ses feuilles

33 Exemple tiré du *Diable et le paysan*, de F. Rabelais.

fanées.>; <Il en fut tellement vexé qu'on ne le revit jamais plus au pays.>

Afin de mieux mettre au jour le rôle structurant des émotions dans les textes (*cf.* chapitre 7), (Battistelli, Étienne, et al., 2022) introduit un raffinement du schéma d'annotation pour l'annotation des causes et des conséquences de l'émotion. Le nouveau formalisme proposé s'affranchit dans un premier temps du principe d'annotation à l'échelle de la phrase graphique et emploie les unités SitCause et SitConseq pour annoter des unités linguistiques à l'échelle du marqueur linguistique. Les nouvelles unités SitCause et SitConseq ne sont spécifiées par aucun trait. Deux nouvelles relations sont introduites, en remplacement de la structure PassageEmo : la relation CauseDe, qui lie une unité SitCause à une unité SitEmo qui lui est associée ; et la relation ConseqDe, qui lie une unité SitConseq à une unité SitEmo qui lui est associée. Ces raffinements sont implémentables sous Glozz mais ne l'ont pas encore été à ce jour.

Les exemples [24] et [25] illustrent ce nouveau formalisme doté d'une délimitation plus fine des unités SitCause et SitConseq. En [24], l'émotion de JOIE de l'expérienceur « Paul » est captée par deux unités SitEmo (l'une désigne l'émotion tandis que l'autre la suggère). La cause de cette émotion est exprimée par la proposition « il a gagné la course » (unité SitCause), introduite par la conjonction de subordination « parce qu ». La relation CauseDe relie ensuite l'unité SitCause aux deux unités SitEmo annotant l'expression de la JOIE de Paul.

[24] Paul est heureux parce qu'**il a gagné la course**.

Experienceur<Paul>{Entité : Paul, Contrôle : Humain, Nature : NP}

SitEmo<heureux>{Mode : Désignée, Type : Base, Catégorie : Joie, Type2 : Aucune, Catégorie2 : Aucune, Déclencheur : heureux, Nature : SAdj}

Expérienceur<il>{Entité : Paul, Contrôle : Humain, Nature : ProPers}

SitEmo<a gagné la course>{Mode : Suggérée, Type : Base, Catégorie : Joie, Type2 : Aucune, Catégorie2 : Aucune, Déclencheur : gagné, Nature : Proposition}

SitCause<il a gagné la course>

Relations :

<Paul> <--Affecte-- <heureux>

<il> <--Affecte-- <a gagné la course>

<il a gagné la course> --CauseDe--> <heureux>

En [20], l'émotion de TRISTESSE de l'entité « Pierre » est désignée par une seule SitEmo. La conséquence de cette émotion est exprimée par « Marie lui a acheté une

glace » et captée par une unité SitConseq. La relation ConseqDe relie ensuite l'unité SitConseq à l'unité SitEmo dénotant la tristesse de Pierre.

[25] Comme Pierre est triste, *Marie lui a acheté une glace.*

Experienceur<Pierre>{Entité : Pierre, Contrôle : Humain, Nature : NP}

SitEmo<triste>{Mode : Désignée, Type : Base, Catégorie : Tristesse, Type2 : Aucune, Catégorie2 : Aucune, Déclencheur : heureux, Nature : SAdj}

SitConseq<Marie lui a acheté une glace>

Relations :

<Pierre> <--Affecte-- <triste>

<Marie lui a acheté une glace> --ConseqDe--> <triste>

3) Une ébauche de schéma pour annoter la temporalité linguistique des textes

La partie « TAME » du schéma d'annotation repose sur deux unités et trois relations, chacune visant à mieux cerner les interactions entre temporalité linguistique (les catégories TAME) et expression des émotions. Ces unités et relations sont représentées de manière synthétique par la figure 9.

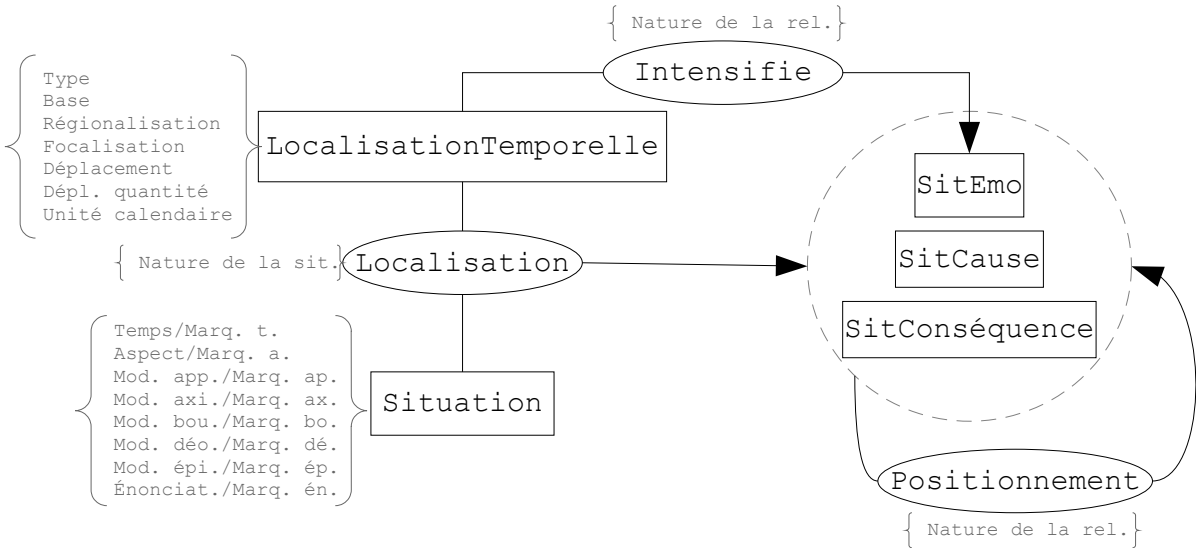


Figure 9: Unités et relations de la partie « TAME » du schéma d'annotation

À ce jour, aucune de ces structures n'a servi à annoter de manière systématique les informations temporelles présentes dans un texte[34]. Cependant, j'ai développé un script permettant de transformer les annotations d'adverbiaux temporels issues de (Teissedre, 2012) en unités LocalisationTemporelle exploitables sous Glozz.

Unité Situation. L'unité Situation permet l'annotation simultanée des informations relatives aux catégories TAME. Il n'est pas rare qu'un même marqueur linguistique articule des valeurs de plusieurs de ces catégories (par ex. une flexion verbale qui articule bien souvent des valeurs à la fois de temps, d'aspect et de modalité épistémique) ou que plusieurs marqueurs œuvrent de concert pour exprimer une catégorie TAME. Afin de ne pas démultiplier les unités et de faciliter le procédé d'annotation, l'unité Situation a été pensée pour s'appliquer à une proposition – voire une phrase graphique – entière. Cette unité est spécifiée par seize traits, présentés dans le tableau 13. Ces traits fonctionnent par paire : l'un indique la valeur associée à la catégorie TAME exprimée et l'autre renseigne, grâce à de la saisie libre, le ou les marqueurs dénotant cette valeur.

Les valeurs choisies pour le trait *Temps* offre la possibilité de distinguer si les notions de passé / présent / futur sont à entendre du point de vue de l'énonciateur principal ou d'un énonciateur second. Pour le trait *Aspect*, seules trois valeurs génériques ont été retenues car jugées suffisantes pour rendre compte des phénomènes que ce schéma ambitionne d'observer.

Les catégories modales ainsi que leurs valeurs associées sont issues de (Gosselin, 2005, 2010). Excepté pour la modalité épistémique, considérée comme nécessairement exprimée dans une proposition, la valeur « Non applicable » (N/A) a été ajoutée pour les traits de catégories modales et intervient lorsque la catégorie modale n'est pas présente dans la proposition annotée.

Les valeurs du traits *Énonciation* – renvoyant à la notion de prise en charge énonciative – sont quant à elles issues de (Battistelli & Amardeilh, 2009). Comme pour le trait *Modalité épistémique*, le trait *Énonciation* est considéré comme étant nécessairement réalisé au sein d'une proposition.

34 La partie « TAME » du schéma est cependant en cours d'utilisation par C. Bruneau, dans le cadre de sa thèse encadrée par D. Battistelli et J. Schuh.

Trait	Valeurs
Temps / MarqueurTemps	PasséE_principal, PrésentE_principal, FuturE_principal, PasséE_second, FuturE_second, Antérieur, Concomitant, Postérieur / *Saisie libre*
Aspect / MarqueurAspect	État, Processus, Évènement / *Saisie libre*
Modalité appréciative / Marqueur appréciatif	N/A, Indésirable, Désirable / *Saisie libre*
Modalité axiologique / Marqueur axiologique	N/A, Blâmable, Louable / *Saisie libre*
Modalité boulique / Marqueur boulique	N/A, Rejeté, Désiré / *Saisie libre*
Modalité déontique / Marqueur déontique	N/A, Interdit, Facultatif, Permis, Obligatoire / *Saisie libre*
Modalité épistémique / Marqueur épistémique	Hypothétique, Avéré / *Saisie libre*
Énonciation / Marqueur énonciation	Prise en charge totale, Prise en charge partielle, Prise en charge limitée à un cadre temporel, Prise en charge limité à un cadre spatial, Prise en charge limitée à un cadre notionnel, Prise en charge limitée à un cadre autre / *Saisie libre*

Tableau 13: Traits de l'unité Situation

Unité LocalisationTemporelle. De par les nombreux liens qu'elle semble entretenir avec l'expression des émotions et pour faciliter le processus d'annotation, une unité est spécifiquement dédiée à l'annotation des adverbiaux temporels : l'unité LocalisationTemporelle. Cette unité est spécifiée par sept traits présentés tableau 14. Les traits et les valeurs qui leurs sont associées sont retenus de (Battistelli, 2009; Teissedre, 2012). Ces travaux proposent une représentation formelle des adverbiaux correspondant à une succession d'opérations sémantiques s'appliquant sur un repère temporel initial, la *base*. Le trait *Base* indique ainsi quelle base est située au fondement de l'adverbial annoté et le trait *Unité calendaire* de quelle unité temporelle elle relève. Le trait *Type* intervient pour renseigner le type d'adverbial annoté. Un adverbial unaire ne comporte qu'une seule base et renvoie à un seul intervalle temporel (par ex. *demain* ou *le 14 novembre 2023*) tandis qu'un adverbial binaire a deux bases (par ex. *de fin mars à début avril 2022*). Un adverbial composé est quant à lui constitué de l'association de plusieurs adverbiaux temporels (par ex. *de mars à avril, dès 19h*).

Trois types d'opérations sémantiques pouvant s'appliquer sur une base sont distingués : les opérations de *régionalisation* qui, à l'aide de marqueurs tels que *vers*, *aux*

alentours de, depuis, modifient « la région pointée par l'adverbial sur l'axe du temps » (Teissedre, 2012, p. 90) ; les opérations de *focalisation* qui effectuent un « zoom » par rapport à la granularité de la base grâce à des marqueurs tels que *fin, au milieu de* ; et les opérations de déplacement qui « décrivent les décalages opérés par rapport à la zone temporelle pointée initialement » (*ibid.*) et sont visibles dans des adverbiaux tels que *depuis deux mois, il y a trois semaines*. Chacun de ces types d'opération est traduit en un trait de l'unité LocalisationTemporelle.

Trait	**Valeurs**
Type	Unaire, Binaire, Composé
Base	Calendaire, Déictique, Relative, Itérative, Anaphorique, Évènement, N/A
Régionalisation	Avant, Après, Jusque, Depuis, Vers, Id
Focalisation	Quart, Tiers, Moitié, Début, Milieu, Fin, Id
Déplacement	Avant, Après, Id
Déplacement Quantité	*Saisie libre*
Unité calendaire	Seconde, Minute, Heure, Partie de jour, Jour, Week-end, Semaine, Mois, Saison – Été, Saison – Automne, Saison – Hiver, Saison – Printemps, Semestre, Année, Décennie, Siècle, Millénaire, Unité temporelle vague

Tableau 14: Traits de l'unité LocalisationTemporelle

Relation Localisation. Pour lier une unité temporelle (Situation ou Localisation Temporelle) à une unité de type SitEmo, SitCause ou SitConséquence, le schéma prévoit la relation Localisation. Elle est caractérisée par le trait *Nature de la situation* renseignant le type est la situation localisée, « émotionnelle », « de cause » , « de conséquence » ou « autre ».

Relation Positionnement. La relation Positionnement intervient pour lier deux unités SitEmo, SitCause ou SitConséquence afin de rendre compte de la relation temporelle entre deux situations. Elle est spécifiée par un trait *Nature de la relation* prenant pour valeurs « avant », « après » ou « concomitant » pour indiquer le type de relation temporelle qu'entretiennent les deux situations reliées.

Relation Intensifie. La relation Intensifie intervient pour lier une unité LocalisationTemporelle à une unité SitEmo lorsque l'adverbial annoté influence l'intensité de l'émotion exprimée par l'unité SitEmo. Cette relation est inspirée du lien entre adverbe temporel et intensité de l'émotion mentionné dans (Romero, 2017). Elle est ca-

ractérisée par le trait *Nature de la relation* pour indiquer le type d'influence de l'adverbial sur l'émotion (« Augmente » ou « Diminue »).

Conclusion

Ce chapitre a décrit un schéma d'annotation d'une grande richesse, reposant sur une modélisation articulant quatre unités pour cerner l'expression des émotions (SitEmo, Expérienceur, SitCause et SitConséquence) et plusieurs structures permettant de lier entre elles ces unités (relation Affecte et schéma PassageEmo). Le schéma comporte également une partie dédiée à l'annotation des dimensions TAME, reposant sur deux unités (Situation et LocalisationTemporelle) et trois relations (Localisation, Positionnement et Intensifie). Cette seconde partie du schéma, à ce jour en cours d'utilisation pour annoter des textes dans le cadre de travaux menés par C. Bruneau, vise à permettre un approfondissement et une systématisation de l'analyse des interactions entre marqueurs TAME et expression des émotions. La partie émotion a été appliquée sur un grand ensemble de textes aboutissant à la constitution d'un corpus annoté (chapitre 4), à partir duquel des expériences d'apprentissage automatique (chapitre 5) et des observations linguistiques ont pu être menées (chapitre 7).

Tout au long du travail rapporté ici, il est clair que le choix de certaines annotations proposées pour l'analyse d'exemples (choix d'une catégorie émotionnelle plutôt qu'une autre, choix d'un mode d'expression plutôt qu'un autre, *etc.*) pourrait être discuté. Je souhaite cependant souligner que, sans réfuter la pertinence des corrections qui pourraient être proposées de ce point de vue, la tâche d'analyse de la composante émotionnelle d'un texte reste à ce jour encore exploratoire et fait l'objet de débats théoriques, ne serait-ce que par rapport au nombre ou aux libellés des catégories émotionnelles, comme en attestent les nombreux jeux d'étiquettes proposés dans la littérature (*cf.* chapitre 2). Ce caractère non consensuel est notamment mis en évidence au travers de calculs d'accords inter-annotateurs (AIA) généralement très bas sur la tâche d'annotation en émotions (*cf.* chapitre 2). Dans l'état actuel, l'applicabilité et la pertinence du schéma ont été illustrées au sein du projet TextToKids par sa mise en œuvre sur plus de 1 500 textes de genres différents, dont des ouvrages de fiction jeunesse entiers (de 48 à 80 pages), par des annotateurs experts formés sur cette tâche. Il reste qu'il n'y a pas à ce jour de calcul d'AIA prenant en compte l'intégralité du schéma d'annotation (délimitation des unités, des relations, *etc.*).

Chapitre 4 Constitution d'un corpus annoté manuellement en émotions

L'analyse des émotions proposée ici est fondée sur une modélisation de l'expression des émotions intégrant des notions pertinentes à la fois du point de vue linguistique, psycholinguistique et dans la perspective d'une automatisation de son analyse (*cf.* chapitres 2 & 3). Cette modélisation se concrétise sous la forme d'un schéma d'annotation dédié au repérage d'unités linguistiques participant de l'expression des émotions. Le schéma a été développé pour être utilisé *via* la plateforme d'annotation manuelle Glozz (Widlöcher & Mathet, 2012). Il articule des unités, délimitées directement dans le texte par l'annotateur pour capter des marqueurs linguistiques et des relations, permettant de lier deux unités entre elles. Chacune de ces unités et relations est caractérisée par un ensemble de traits, précisant ainsi l'annotation effectuée. Le cœur du schéma d'annotation repère de quatre types d'unités :

- les unités SitEmo, captant les unités linguistiques qui expriment une émotion relevant d'une des onze catégories suivantes : ADMIRATION, COLÈRE, CULPABILITÉ, DÉGOÛT, EMBARRAS, FIERTÉ, JALOUSIE, JOIE, PEUR, SURPRISE et TRISTESSE. Les SitEmo sont caractérisées par cinq traits principaux, les traits *Mode* (mode d'expression), *Type/Type2* (type d'émotion) et *Catégorie/Catégorie2* (catégorie émotionnelle) ;
- les unités Expérienceur, captant les unités linguistiques qui expriment l'expérienceur et qui sont caractérisées par deux traits principaux, les traits *Entité* (individu) et *Contrôle* (± animé, humain) ;
- les unités SitCause, captant les unités linguistiques qui expriment la cause d'une émotion annotée par une unité SitEmo ;
- les unités SitConséquence, captant les unités linguistiques qui expriment la conséquence d'une émotion annotée par une unité SitEmo.

Un cinquième type d'unités, les unités Autre, sans trait pour les caractériser, interviennent pour annoter les unités émotionnelles ne renvoyant pas à l'une des onze catégories émotionnelles que j'ai choisi de retenir pour les SitEmo. La relation Affecte relie enfin une unité SitEmo à son unité Expérienceur associée.

L'application du schéma a nécessité d'opérer, en amont, un ensemble de choix méthodologiques concernant la délimitation des unités annotées (section 1) et de constituer le corpus de textes à annoter (section 2). Le processus d'annotation (section 3) a conduit à la constitution d'un corpus de plus de 1 500 textes annotés (section 4).

Le formatage identifiant dans ce chapitre les unités annotées des exemples est le même que celui du chapitre 3 : les unités SitEmo sont soulignées et les unités Expérienceur sont surlignées en bleu clair. Les relations Affecte sont symbolisées par une flèche et n'apparaissent que dans les informations d'annotations situées sous les textes d'exemple.

1) Choix méthodologiques pour la délimitation d'unités annotées

Le schéma d'annotation a été conçu pour être utilisé par des annotateurs linguistes experts. Tous les choix ayant conduit à la forme actuelle du schéma ainsi que de nombreux exemples annotés et cas particuliers se présentant lors du processus d'annotation sont détaillés dans le guide associé au schéma (Étienne & Battistelli, 2021). Dans cette section, ne seront mentionnés que quelques critères et cas qui semblent particulièrement intéressants du point de vue méthodologique et ont des implications sur les analyses menées dans le reste du travail. Certaines décisions clés prises pour résoudre les difficultés liées à la délimitation des unités émotionnelles seront donc exposées. Ces difficultés émergent de la granularité de l'annotation conduite qui s'effectue à l'échelle du marqueur linguistique (token, syntagme, proposition, *etc.*).

Critères pour distinguer les modes d'expression. (Étienne & Battistelli, 2021) émet de nombreuses recommandations et explicite les critères pour identifier – délimiter – les unités SitEmo et déterminer la valeur du trait *Mode* à leur associer. Ainsi, toute unité SitEmo décrivant une manifestation physique (physiologique, comportementale, discursive) occasionnée par un ressenti émotionnel est jugée « comportementale ». Les unités SitEmo qui *désignent* l'émotion correspondent quant à elles à un terme ou à un groupe de termes dont la définition lexicographique (dans le CNRTL, le Wiktionnaire ou le Larousse en ligne) remplit au moins un des trois critères suivants : a) le terme est défini comme étant une émotion, un sentiment, un affect, une humeur, un état psychologique *etc.* ; b) le terme est défini en lien avec un terme du

lexique émotionnel classique (joie, colère, tristesse *etc.*) ; c) le terme apparaît comme un synonyme ou comme un antonyme d'un terme remplissant le critère a) et/ou le critère b). Ces critères sont repris de (Piolat & Bannour, 2009). Pour qu'elle soit considérée comme « suggérée », une SitEmo doit comporter des marqueurs décrivant une situation associable de manière conventionnelle, socio-culturelle et prototypique à un ressenti émotionnel et ainsi reconnue comme pouvant être à l'origine d'une émotion. Enfin, une SitEmo « montrée » se doit de capter une caractéristique de l'énoncé (structure syntaxique, terme du lexique, *etc.*) qui montre que l'énonciateur ressentait une émotion au moment de l'énonciation. Un arbre de décision extrait du guide (Étienne & Battistelli, 2021) conçu pour aider l'identification des modes d'expression est présenté en Erreur : source de la référence non trouvée.

Principe général de la délimitation des unités. De manière générale, les unités délimitées doivent respecter les frontières de syntagmes. Ce principe garantit une certaine compatibilité entre le découpage des unités et les analyseurs syntaxiques ou *chunkers*. L'objectif était en effet de laisser ici la possibilité de recourir à ces outils pour développer un système de détection automatique des émotions à partir du corpus annoté. Les méthodes retenues dans le présent travail n'emploient cependant pas ces techniques (*cf.* chapitre 5).

Une SitEmo, un mode d'expression. Le schéma a été conçu pour qu'un seul mode d'expression soit associé à chaque unité SitEmo. Quand des segments combinent plusieurs modes d'expression des émotions, les marqueurs linguistiques qui se rapportent à différents modes d'expression doivent donc être identifiés séparément. Cela peut conduire à des SitEmo qui se chevauchent ou qui s'emboîtent mais assure qu'aucune information linguistique n'est perdue. Dans une locution telle que *sauter de joie*, la JOIE exprimée sera par exemple annotée par deux SitEmo : l'une capte l'expression comportementale « sauter de joie » ; l'autre capte le terme « joie » qui désigne l'émotion.

Un découpage fin : la relation Discontinue. Les critères de délimitation des unités annotées œuvrent tous pour garantir le découpage le plus fin possible. Cela conduit parfois l'annotateur à délimiter des unités *discontinues*. La relation Discontinue intervient alors pour lier les différentes unités constituant l'unité discontinue, comme illustré exemple [26] où l'unité SitEmo a été délimitée de sorte à exclure l'adverbe « désormais », afin de resserrer l'annotation sur les marqueurs exprimant réelle-

ment l'émotion. La relation Discontinue intervient ensuite en post-traitement pour « recoller » les segments textuels captés par les différentes unités discontinues. Si deux unités SitEmo distinctes apparaissent dans Glozz pour l'exemple [26], une seule unité SitEmo est comptabilisée lors de l'analyse des résultats.

[26] Tous les pays du monde redoutent désormais le pire : le déclenchement d'une guerre nucléaire[35].

Expérienceur<Tous les pays du monde>{Entité : Tous les pays du monde, Contrôle : Humain, Nature : SN}

SitEmo<redoutent>+<le pire : le déclenchement d'une guerre nucléaire>{Mode : Désigné, Type : Base, Catégorie : Peur, Type2 : Aucun, Catégorie2 : Aucune, Nature : Proposition, Déclencheur : redoutent}

SitEmo<le pire : le déclenchement d'une guerre nucléaire>{Mode : Suggéré, Type : Base, Catégorie : Peur, Type2 : Aucun, Catégorie2 : Aucune, Nature : SN, Déclencheur : pire;guerre nucléaire}

Relation Discontinue :

<redoutent>----><le pire : le déclenchement d'une guerre nucléaire>

Relations Affecte :

<Tous les pays du monde> <---- <redoutent>+<le pire : le déclenchement d'une guerre nucléaire>

<Tous les pays du monde> <---- <le pire : le déclenchement d'une guerre nucléaire>

Émotions désignées et modalité. De manière générale, l'annotateur est tenu de repérer toutes les unités linguistiques de type émotion désignée. Si un marqueur désignant l'émotion est affecté par des marques de modalité (négation, marques hypothétiques, *etc.*), le label doit malgré tout être annoté par une unité SitEmo et ce même si le ressenti émotionnel est explicitement nié ou présenté comme non avéré. L'exemple [27] où les trois SitEmo désignées sont affectées par le marqueur de modalité « espèrent » illustre ce phénomène. La partie du schéma dédiée à l'analyse des caractéristiques temporelles, modales et énonciatives des textes a été pensée pour permettre, lors d'une phase d'annotation supplémentaire, d'identifier les marqueurs de modalité et de rendre compte de l'interaction entre la modalité et l'expression de l'émotion (par ex. émotion hypothétique ou avérée). Si l'exclusion des marques de modalité occasionne la séparation de l'unité SitEmo en plusieurs unités, la relation Discontinue servira à les relier.

[27] Ils espèrent tous être étonnés, éblouis, épatés[36].

35 Exemple issu de l'article « États-Unis et Corée du Nord : le bras de fer », *Albert*, n°15, 3 mai 2017.

36 Exemple issu de *La famille trop d'filles – Élisa*, Susie Morgenstern.

Expérienceur<Ils>{Entité : les spectateurs, Contrôle : Humain, Nature : ProPers}

SitEmo<étonnés>{Mode : Désigné, Type : Base, Catégorie : Surprise, Type2 : Aucun, Catégorie2 : Aucune, Nature : SAdj, Déclencheur : étonnés}

SitEmo<éblouis>{Mode : Désigné, Type : Complexe, Catégorie : Admiration, Type2 : Aucun, Catégorie2 : Aucune, Nature : SAdj, Déclencheur : étonnés}

SitEmo<épatés>{Mode : Désigné, Type : Base, Catégorie : Surprise, Type2 : Complexe, Catégorie2 : Admiration, Nature : SAdj, Déclencheur : étonnés}

Relations Affecte :

<Ils> <---- <étonnés>

<Ils> <---- <éblouis>

<Ils> <---- <épatés>

Expérienceur de type doxa ou scripteur : nécessité des unités fantômes. Quand l'expérienceur de l'émotion est la doxa ou le scripteur, il peut arriver qu'aucun segment textuel ne désigne explicitement l'entité. Afin de distinguer ce phénomène des cas où l'émotion n'a effectivement aucun expérienceur[37], le principe de l'unité « fantôme » a été introduit. Pour délimiter une unité fantôme, l'annotateur sélectionne le premier caractère qui jouxte à droite l'unité SitEmo. L'unité Expérienceur est ainsi collée à l'unité SitEmo et correspond généralement à une espace ou un signe de ponctuation. Les unités fantômes sont positionnées à la suite des unités SitEmo pour traduire le fait que l'unité n'existe pas matériellement dans le texte mais qu'elle est issue du raisonnement suivant : après avoir lu une unité émotionnelle, on se demande qui elle affecte mais la recherche infructueuse d'un segment renvoyant à l'entité qui ressent l'émotion pousse à délimiter une unité fantôme. Si le ressenti émotionnel de la doxa ou du scripteur est réalisé par plusieurs unités linguistiques, une unité fantôme sera créée pour chaque unité SitEmo.

L'absence de marqueurs linguistiques référant explicitement à l'expérienceur rend parfois difficile la distinction entre l'entité doxa et l'entité scripteur. En cas d'ambiguïté, les critères suivants s'appliquent :

– si des marqueurs montrent l'émotion en dehors de tout discours rapporté, alors l'émotion dénotée par ces marqueurs ainsi que par toutes les autres unités exprimant la même émotion – peu importe leur mode d'expression – est attribuée

37 Les cas où une émotion n'a aucun expérienceur sont très rares dans le corpus et correspondent au fait de ne pas pouvoir répondre à la question « qui ressent l'émotion ? ». Cela arrive typiquement quand un terme désignant l'émotion est employé de manière générique, par exemple dans un énoncé comme *La peur est une émotion comme une autre.*

à l'expérienceur scripteur, comme illustré exemple [28] où la JOIE du scripteur est à la fois montrée, désignée et suggérée ;

– si une structure passive marque la distance de l'énonciateur principal par rapport au contenu propositionnel et qu'aucun marqueur qui montre l'émotion n'est présent, alors l'émotion est attribuée à la doxa, comme illustré exemple [29] où la COLÈRE de la doxa est à la fois comportementale (comportement discursif) et suggérée.

[28] Ah la la, qu'est-ce qu'un bon repas peut apporter comme plaisir !![38]

SitEmo<Ah la la>{Mode : Montré, Type : Base, Catégorie : Joie, Type2 : Aucun, Catégorie2 : Aucune, Nature : Interjection, Déclencheur : Ah la la}

Expérienceur<,>{Entité : Scripteur, Contrôle : Humain, Nature : N/A}

SitEmo<qu'est-ce qu'un bon repas peut apporter comme plaisir> {Mode : Montré, Type : Base, Catégorie : Joie, Type2 : Aucune, Catégorie2 : Aucune, Nature : Énonce exclamatif, Déclencheur : qu'est-ce qu'un bon repas peut apporter comme plaisir }

Expérienceur< >{Entité : Scripteur, Contrôle : Humain, Nature : N/A}

SitEmo<bon>{Mode : Suggéré, Type : Base, Catégorie : Joie, Type2 : Aucun, Catégorie2:Aucune, Nature : SAdj, Déclencheur : bon}

Expérienceur< >{Entité : Scripteur, Contrôle : Humain, Nature : N/A}

SitEmo<plaisir>{Mode : Désigné, Type : Base, Catégorie : Joie, Type2 : Aucun, Catégorie2 : Aucune, Nature : SN, Déclencheur : plaisir}

Expérienceur< >{Entité : Scripteur, Contrôle : Humain, Nature : N/A}

SitEmo<!>{Mode : Montré, Type : Base, Catégorie : Joie, Type2 : Aucun, Catégorie2 : Aucune, Nature : Point d'exclamation, Déclencheur : !}

Expérienceur< >{Entité : Scripteur, Contrôle : Humain, Nature : N/A}

Relations Affecte :

<Ah la la> ----> <,>

<qu'est-ce qu'un bon repas peut apporter comme plaisir> ----> < >

<bon> ----> < >

<plaisir> ----> < >

<!> ----> < >

38 Exemple tiré de l'article « Qu'est-ce que la haute gastronomie ? », *P'tit Libé*, n°90, 25 au 30 janvier 2019.

[29] Les zoos et les aquariums sont souvent critiqués parce qu'ils enferment des animaux[39].

SitEmo<sont>+<critiqués>{Mode : Comportemental, Type : Base, Catégorie : Colère, Type2 : Aucun, Catégorie2 : Aucune, Nature : Proposition, Déclencheur : critiqués}

Expérienceur< >{Entité : Doxa, Contrôle : Humain, Nature : N/A}

SitEmo<ils enferment des animaux>{Mode : Suggéré, Type : Base, Catégorie : Colère, Type2 : Aucun, Catégorie2 : Aucune, Nature : Proposition, Déclencheur : enferment}

Expérienceur<.>{Entité : Doxa, Contrôle : Humain, Nature : N/A}

Relation Discontinue :

<sont>----><critiqués>

Relations Affecte :

<sont>+<critiqués> ----> < >

<ils enferment des animaux> ----> <.>

2) Constitution d'un corpus de textes pour l'analyse de la dimension émotionnelle des textes

Le corpus global sur lequel le schéma a été appliqué comporte 1 594 textes, répartis en trois genres : journalistique, encyclopédique et fictionnel. Ce corpus, qui représente lui-même une sous-partie du corpus TextToKids général (*cf.* chapitre 1), est composé de sept sous-corpus distincts :

- le corpus *P'tit Libé*, composé d'articles issus du journal *P'tit Libé*, partenaire du projet TextToKids, adressé aux 7-12 ans (un texte = un article) ;
- le corpus *Albert*, composé d'articles issus du journal *Albert*, partenaire du projet TextToKids, adressé aux 9-14 ans (un texte = un article) ;
- le corpus CovidAdultes, composé d'articles traitant de l'épidémie de Covid-19, soit de manière descriptive (qu'est-ce que le Covid-19 ? quelles sont les mesures mises en place par le gouvernement ? *etc.*), soit par le biais de témoignages (de soignants, de personnes confinées, *etc.*), issus des versions en ligne et gratuites des journaux *Le Monde* et *Libération* (un texte = un article) ;
- le corpus CovidJeunesse, composé d'articles traitant de l'épidémie de Covid-19, soit de manière descriptive, soit par le biais de témoignages, issus des versions en ligne et gratuites des journaux *Petit Quotidien* (pour les 6-10 ans), *Mon*

39 Exemple issu de l'article « Que peut-on faire pour les sauver ? », *P'tit Libé*, n°82, 30 novembre au 6 décembre 2018.

Quotidien (pour les 10-13 ans), *L'Actu* (pour les 13-18 ans) et *P'tit Libé* (un texte = un article) ;

- le corpus FictionM2, composé d'ouvrages de fiction entiers, quatre livres de la série *La famille trop d'filles*, de S. Morgenstern, adressée aux 7-11 ans ; *J'aime/je déteste le français*, de B. Friot, adressé aux 8-10 ans ; et *Le monde selon Frrrintek*, d'A. Desarthe, adressé aux 8-13 ans. Les manuscrits de ces ouvrages ont été fournis par les auteur·e·s dans le cadre du projet TextToKids (1 texte = 1 ouvrage) ;
- le corpus WP7Encyclo, composé d'extraits d'articles de l'encyclopédie en ligne Vikidia, destinée aux 8-13 ans. Ce corpus a été analysé en émotions lors du WP7 du projet TextToKids (*cf.* chapitre 1) ;
- le corpus WP7Fiction, composé d'extraits d'ouvrages de littérature jeunesse. Certains extraits proviennent du corpus du projet ANR Alector[40] : *Les boîtes de peintures*, M. Aymé (pour les 6-8 ans) ; *Émilie et le crayon* magique, H. Bichonnier (6-8 ans) ; *Le Diable et le* paysan, F. Rabelais (7-11 ans) ; *La dragonne de minuit*, A. Laroche (9-11 ans) ; Œil *du loup*, D. Pennac (9-11 ans) ; *Harry Potter à l'école des* sorciers, J. K. Rowling (9-11 ans). Deux autres extraits proviennent d'ouvrages adressés aux 6-8 ans fournis par les auteur·e·s au projet TextToKids : *Chien Pourri fait du ski,* C. Gutman et *La bande à Tristan*, M.-A. Murail. Un dernier extrait issu de *Verte*, M. Desplechin a été ajouté de manière arbitraire. L'ensemble de ces textes a été analysé en émotions lors du WP7 du projet TextToKids (*cf.* chapitre 1).

95,9 % des textes du corpus (88,6 % des tokens) correspondent à des textes issus des journaux *P'tit Libé* et *Albert*. Ces textes, disponibles rapidement et en grande quantité, tout en garantissant une diversité thématique ont été les premiers à être annotés en émotions. Pour s'assurer de la robustesse et de l'applicabilité du schéma, d'autres textes ont ensuite été ajoutés afin de diversifier les genres et les sources.

Les caractéristiques de ces sept sous-corpus et du corpus global sont résumées en terme de nombre de textes, de tokens et de phrases dans le tableau 15.

40 https://alectorsite.wordpress.com/

Corpus	Genre	Nb de textes (nb de tokens)	Nb de phrases	Nb. moy. de phrases par texte	Nb. moy. de tokens par texte
PtitLibe	Journalistique	611 (217 619)	12 130	20	356
Albert	Journalistique	917 (238 839)	11 854	13	260
CovidAdultes	Journalistique	7 (9 618)	368	53	1 374
CovidJeunesse	Journalistique	31 (15 981)	1 056	34	516
FictionM2	Fictionnel	6 (24 975)	2 075	346	4 163
WP7Encyclo	Encyclopédique	11 (4 195)	199	18	381
WP7Fiction	Fictionnel	11 (4 049)	327	30	368
Total		**1 594 (515 276)**	**28 009**	**18**	**323**

Tableau 15: Composition du corpus annoté en émotions

Il est à noter que la pandémie de Covid-19 représentait une occasion rare de constituer aisément un corpus respectant une unité thématique mais composé de sources variées. Les corpus CovidAdultes et CovidJeunesse ont été pensés pour permettre des observations linguistiques sur l'expression des émotions, opposant différents styles de texte (explicatif par opposition à témoignage), différentes sources (*Petit Quotidien, Mon Quotidien, etc.*) et différents publics cibles (adultes par opp. à enfants).

3) Processus d'annotation : équilibre entre quantité et qualité des annotations

La mise en œuvre du schéma d'annotation a dans un premier temps été consacrée à son application sur un grand nombre de textes (section 3.1). Une mesure d'accord inter-annotateurs a ensuite été effectuée sur une partie du corpus annoté afin d'évaluer le caractère consensuel de quelques notions clés du schéma (section 3.2).

3.1) Annotation en émotions d'un corpus de textes hétérogènes

Le schéma d'annotation a été appliqué par six annotatrices expertes en linguistique (A1 est l'autrice de la présente thèse, A2 à A5 des étudiantes en master TAL et A6 la directrice de thèse de A1). Les annotatrices ont disposé du guide (Étienne & Battistelli, 2021) comme point de référence lors du processus d'annotation qui a toujours débuté par une période de familiarisation avec le schéma. Pendant cette période d'environ une semaine, les annotatrices notaient les questions et doutes qui émergeaient

lors de l'analyse des premiers textes puis en discutaient avec moi. Chaque annotatrice a annoté son propre ensemble de textes comme illustré par le tableau 16. Les annotatrices A3 à A5 se sont donc réparties les textes du corpus FictionM2. Les seuls corpus ayant subi une procédure d'annotation légèrement différente sont les corpus WP7Encyclo et WP7Fiction : A1 et A6 ont discuté des annotations puis A1 a opéré l'annotation effective des textes sous Glozz.

Corpus	**Annotateur**	**Unités**
PtitLibe	A1	SitEmo & Epérienceur
Albert	A2	SitEmo & Epérienceur
CovidAdultes	A2	SitEmo & Epérienceur
CovidJeunesse	A2	SitEmo & Epérienceur
FictionM2	A3, A4 & A5	SitEmo, Epérienceur, SitCause & SitConséquence
WP7Encyclo	A1 & A6	SitEmo, Epérienceur, SitCause & SitConséquence
WP7Fiction	A1 & A6	SitEmo, Epérienceur, SitCause & SitConséquence

Tableau 16: Annotatrices et unités annotées selon le sous-corpus

Il existe un troisième corpus de textes analysés dans le cadre du WP7, selon les mêmes principes que pour les corpus WP7Encyclo et WP7Fiction : le corpus WP7Journa. Composé de neufs articles journalistiques, ce corpus a donc été annoté en SitEmo, Expérienceur, SitCause et SitConséquence après une analyse concertée entre A1 et A6. Ce corpus n'est cependant pas représenté dans les tableaux car les textes qui le composent, tous issus du journal *Albert* ou *P'tit Libé*, se recoupent soit avec le corpus Albert, soit avec le corpus PtitLibé.

La priorité du travail conduit ici était dans un premier temps de proposer un schéma complet modélisant la notion d'émotion (*cf.* chapitre 4) et de vérifier l'applicabilité et la pertinence de ce schéma[41]. La seconde priorité était d'obtenir un grand nombre de données annotées afin de pouvoir mobiliser le corpus annoté constitué pour des expériences informatiques d'analyse automatique des émotions à l'aide des techniques de l'apprentissage automatique (*cf.* chapitre 5). Une procédure de calcul d'accord inter-annotateurs (AIA) n'est donc intervenue que tardivement, une fois l'ensemble des données annotées.

41 Le schéma a également été appliqué dans le cadre du projet ANR Flyer pour annoter en émotions des textes extrémistes (Dragos et al., 2022). Il est par ailleurs mobilisé dans le cadre d'une thèse en cours conduite par C. Bruneau pour participer à l'étude de la transmission de valeurs dans des manuels scolaires d'histoire.

3.2) Mise en place d'un calcul d'accord inter-annotateurs

Un schéma difficile à évaluer. Le schéma d'annotation repose sur la délimitation manuelle d'unités dans les textes, leur typage, leur association à des traits (*Mode, Catégorie, etc.*), la création de relations pour relier certaines de ces unités entre elles (relation Affecte pour lier les unités SitEmo et Expérienceur) et le typage de ces relations. À ma connaissance, aucune méthode ne permet d'évaluer l'accord sur l'ensemble des éléments d'un schéma aussi complexe. Pour évaluer le consensus sur la délimitation des unités (parfois aussi appelée *unitizing*) l'Alpha de Krippendorf (Krippendorff, 1995) ou la mesure Gamma (Mathet et al., 2015) pourraient par exemple être utilisés, tandis que la mesure d'accord sur l'attribution d'étiquettes (par ex. pour les catégories émotionnelles associées à une SitEmo) s'effectue généralement par un calcul de Kappa de Cohen (Cohen, 1960) pour deux annotateurs ou par un Kappa de Fleiss pour plus de deux annotateurs (Fleiss et al., 1981). Quelque soit la mesure choisie, Y. Mathet et A. Widlöcher (2016) soulignent le fait que « mesure d'accord et mesure de validité[42] sont deux éléments distincts » (p. 79). En effet, « être d'accord ne signifie pas forcément avoir raison » (p. 77), même si un fort accord a tendance à garantir une bonne reproductibilité de la tâche.

Les limites des approches reposant sur l'établissement d'un corpus de référence (*gold standard*) dont la qualité est supposément assurée par un fort AIA sont également soulignées dans (Plank, 2022). La plupart des méthodes de TAL considèrent en effet qu'il est possible – voire nécessaire – d'aboutir à une étiquette de référence pour chaque item à classer, à annoter, cette étiquette reflétant la vérité (*ground truth*). B. Plank met cependant en garde en indiquant que « cette présupposition n'a souvent aucun sens » (traduit de l'anglais ; *ibid.*, p. 10 671). Cela concerne en particulier les tâches d'annotation sémantique – comme l'annotation en émotions – où les variations d'étiquetage opéré par les humains s'expliquent par « des désaccord dus à des cas difficiles, à la subjectivité ou à des cas où des réponses multiples sont plausibles » (traduit de l'anglais, *ibid.*). Selon B. Plank, ignorer ces variations pour forcer l'association d'une étiquette unique, soit-disant consensuelle, représente un problème qui gêne les progrès de la résolution automatique de ces tâches plus complexes.

42 La validité perçue comme l'« adéquation à la réalité » (Mathet & Widlöcher, 2016, p. 76) que les annotations visent à capter (en l'occurrence, l'expression des émotions dans toute leur complexité) lors de la constitution d'un corpus de référence (ou *gold standard*).

La mesure d'un accord reste cependant à ce jour relativement incontournable, ne serait-ce que pour donner une idée de l'aspect consensuel (ou non) de la tâche demandée aux annotateurs.

Mise en place d'un protocole de mesure de l'AIA. Face à la complexité de la mise en œuvre d'un calcul d'AIA complet, seule une partie du schéma d'annotation a été évaluée en s'appuyant sur une simplification de la tâche d'annotation. La visée de cette évaluation est de fournir une idée du degré de consensus sur les annotations produites pour les notions principales du schéma (*mode, type d'émotion et catégorie émotionnelle*).

Corpus. L'ensemble des textes du corpus annoté en émotions n'ont été analysés que par une seule personne (*cf.* section 3.1). Pour mener à bien l'expérience de mesure de l'AIA, il était donc nécessaire de faire annoter par une autre personne des textes déjà annotés. Cependant, au vu du temps considérable nécessaire pour obtenir de nouvelles annotations pour l'ensemble des 1 594 textes annotés par les annotatrices A1 à A6, un sous-corpus de 131 textes – ci-après nommé corpus AIA – a été constitué à partir du corpus annoté global. Le corpus AIA est constitué de cinq sous-corpus, comme illustré par le tableau 17. Il est composé des trois corpus WP7[43] dans leur intégralité et de textes provenant des deux sous-corpus majoritaires : le corpus Albert et le corpus PtitLibé. Cinquante articles issus du corpus Albert et cinquante articles issus du corpus PtitLibé ont donc été choisis aléatoirement en respectant les deux contraintes suivantes : 1) chaque texte sélectionné doit contenir au moins une unité SitEmo dans les annotations initiales ; 2) une fois les 50 textes sélectionnés, un tiers se doit d'être très dense en émotions (textes où le plus de SitEmo ont été annotées), un tiers moyennement dense et le dernier tiers très peu dense (une ou deux SitEmo maximum). Les caractéristiques des sous-corpus annotés lors de la mesure de l'AIA sont présentées tableau 17 en nombre de textes, de tokens et de phrases.

Processus d'annotation. Une septième annotatrice (A7), étudiante en Master TAL, est intervenue pour annoter le corpus AIA selon la même procédure que décrite section 3.1. A7 a donc disposé d'une semaine pour se familiariser avec Glozz et le schéma puis elle a continué l'annotation seule, toujours aidée du guide (Étienne & Battistelli, 2021). L'analyse a porté sur le repérage des unités SitEmo et Expérienceur (ainsi que les relations Affecte qui les unissent), délimités à l'échelle du marqueur lin-

43 Ces textes étant utilisés dans le cadre de protocoles expérimentaux auprès d'enfants, il paraissait intéressant d'avoir un évaluation du consensus de leur analyse.

guistique. L'AIA a donc été mesuré entre deux paires d'annotatrices : entre A1 et A7 pour les sous-corpus PtitLibé50, WP7Encyclo, WP7Fiction et WP7Journa ; entre A2 et A7 pour le sous-corpus Albert50.

Corpus	**Nb textes**	**Nb tokens**	**Nb phrases**
Albert50	50	17 684	850
PtitLibe50	50	21 610	1 148
WP7Encyclo	11	4 195	199
WP7Fiction	11	4 049	326
WP7Journa	9	3 518	177

Tableau 17: Caractéristiques des corpus annotés pour le calcul de l'AIA

Simplification de la tâche d'annotation pour le calcul de l'AIA. Afin de mesurer plus aisément et plus rapidement l'accord entre paires d'annotatrices, une simplification portant à la fois sur le schéma et sur son échelle d'application a été mise en place. Ainsi, seules les étiquettes de mode (« comportemental », « désigné », « montré », « suggéré »), de type d'émotion (« base », « complexe ») et de catégorie émotionnelle (« admiration », « colère », « culpabilité », « dégoût », « embarras », « fierté », « jalousie », « joie », « peur », « surprise », « tristesse ») ont été évaluées. De plus, les annotations ont été transposées à l'échelle de la phrase. Dans cette transposition, une phrase comportant deux unités SitEmo, l'une captant des marqueurs qui désignent la JOIE et l'autre des marqueurs qui suggèrent à la fois la JOIE et la FIERTÉ se voit donc attribuer les étiquettes « désignée », « suggérée », « joie » et « fierté ». Cette simplification s'accorde par ailleurs avec celle opérée pour les expériences de développement des classifieurs en émotions (*cf.* chapitre 5). Pour chacune des étiquettes évaluées, le nombre d'unités SitEmo ainsi que le nombre de phrases correspondant sont présentées par le tableau 18. Ces informations sont indiquées pour le corpus AIA dans son ensemble, en distinguant les annotations initiales (produites par A1 ou A2) de celles produites par A7. Le détail par sous-corpus est présenté en Annexe III.

	Nb d'unités		Nb de phrases	
Étiquette	**A1 ou A2**	**A7**	**A1 ou A2**	**A7**
SitEmo	1371	1339	782	828
comportementale	409	418	314	311
désignée	308	267	263	236
montrée	208	198	132	146
suggérée	446	433	355	340
base	1323	1310	730	787
complexe	222	104	131	84
admiration	81	32	53	29
colère	432	538	267	346
culpabilité	4	1	3	1
dégoût	12	12	9	9
embarras	85	44	51	32
fierté	49	25	39	23
jalousie	3	2	2	2
joie	136	124	96	103
peur	329	250	205	173
surprise	162	118	100	92
tristesse	208	199	135	137

Tableau 18: Modes, types et catégories émotionnelles annotés, en nombre d'unité et de phrases, selon l'annotateur

Mesure de l'AIA : Kappa de Cohen. La mesure retenue pour calculer l'AIA entre A1 ou A2 et A7 est celle du Kappa de Cohen. Cette mesure permet de comparer l'accord entre les deux annotateurs observés sur les données avec l'accord dû au hasard. Le Kappa, noté $\varkappa$, se calcule selon la formule suivante :

$$\varkappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

où *Pr(a)* est la proportion d'accord entre les annotateurs et *Pr(e)* est la probabilité d'un accord aléatoire.

		A2	
		Oui	**Non**
A1	**Oui**	a	b
	Non	c	d

Tableau 19: Matrice de comparaison d'annotations

Pour une matrice d'annotation telle que celle du tableau 19, *Pr(a)* et *Pr(e)* se calculent comme suit :

$$Pr(a) = Pr(A1 = \text{'Oui'} ; A2 = \text{'Oui'}) + Pr(A1 = \text{'Non'} ; A2 = \text{'Non'})$$
$$= \frac{a+d}{a+b+c+d}$$

$$Pr(Oui) = Pr(A1 = \text{'Oui'}) \times Pr(A2 = \text{'Oui'})^{44}$$
$$= \frac{a+b}{a+b+c+d} \times \frac{a+c}{a+b+c+d}$$

$$Pr(Non) = Pr(A1 = \text{'Non'}) \times Pr(A2 = \text{'Non'})^{45}$$
$$= \frac{c+d}{a+b+c+d} \times \frac{b+d}{a+b+c+d}$$

$$Pr(e) = Pr(Oui) + Pr(Non)$$

L'accord a été évalué séparément pour chaque étiquette, avec l'ajout de l'étiquette « émotionnelle » pour rendre compte du niveau macro de l'analyse menée par les annotatrices (ici, une phrase est considérée comme « émotionnelle » si elle comporte au moins une unité SitEmo). Les mesures de Kappa obtenues par paire d'annotatrices sont présentées tableau 20. Les cases grisées correspondent à des étiquettes qui ne sont attribuées par aucune des deux annotatrices. Les résultats de comparaison entre A1 et A7 apparaissent tous sous-corpus confondus. Les mesures de Kappa entre A1 et A7 détaillées par sous-corpus et par genre sont présentées Annexe III.

De manière générale, un accord parfait atteindra une valeur de Kappa de 1, tandis qu'un profond désaccord atteindra une valeur de 0 ou moins. Cependant, selon la tâche considérée, le seuil déterminant une valeur de « bon » accord peut varier. Dans le cas de l'annotation en émotions, l'accord ne dépasse jamais 0,70-0,80 pour deux ou trois classes et se situe toujours en-dessous avec l'ajout de classes (*cf.* chapitre 2). (Kim & Klinger, 2018) atteignent ainsi au maximum un Kappa de Cohen de 0,40 pour l'annotation de la joie.

44 Avec une indépendance supposée de A1 et A2.
45 Avec une indépendance supposée de A1 et A2.

Étiquette	**A1 *vs.* A7**	**A2 *vs.* A7**
émotionnelle	0,652	0,677
comportementale	0,652	0,749
désignée	0,785	0,67
montrée	0,611	0,743
suggérée	0,527	0,383
base	0,636	0,68
complexe	0,495	0,61
admiration	0,36	0,703
colère	0,726	0,689
culpabilité	0,5	
dégoût	0,749	1
embarras	0,525	0,498
fierté	0,493	0
jalousie	1	
joie	0,542	0,474
peur	0,652	0,626
surprise	0,599	0,853
tristesse	0,554	0,278

Tableau 20: κ de Cohen par paire d'annotateur, pour chaque étiquette – corpus entier

Au vu de la tâche réalisée, les mesures d'accord obtenues ici sont donc plutôt bonnes. Si les accords varient selon la paire d'annotatrices considérée, les mêmes tendances se dégagent. L'étiquette « émotionnelle » bénéficie ainsi toujours d'un bon accord (0,652 pour A1-A7 et 0,677 pour A2-A7). Parmi les modes d'expression, le suggéré est toujours le plus problématique (0,527 pour A1-A7 et 0,383 pour A2-A7). Ce mode étant le plus difficile à définir et à cerner, cela n'a rien de surprenant. Les autres modes bénéficient tous d'un bon accord ($>$ 0,61). L'accord est également bon pour l'étiquette « base » (0,636 A1-A7 et 0,68 pour A2-A7). L'autre étiquette de type a un atteint un certain consensus entre A2 et A7 (0,61), moindre entre A1 et A7 (0,495).

Parmi les catégories émotionnelles les plus fréquentes dans le corpus AIA, la COLÈRE est la plus consensuelle avec un accord de 0,726 entre A1 et A7 et de 0,689 entre A2 et A7. La PEUR (0,652 pour A1-A7 et 0,626 pour A2-A7) bénéficie également d'un bon accord. La SURPRISE fait plus d'accord entre A2 et A7 (0,853) qu'entre A1 et A7 (0,599). La JOIE et en particulier la TRISTESSE sont quant à elles plus problématiques (respectivement Kappa de 0,542 et 0,554 pour A1-A7 ; 0,474 et 0,278 pour A2-A7).

Il est important de préciser que, pour les étiquettes très peu représentées (« culpabilité », « dégoût » et « jalousie » principalement, « admiration », « embarras » et « fierté » dans une moindre mesure), les mesures de Kappa obtenues peuvent difficilement être interprétées.

L'évaluation partielle de l'accord montre donc que les annotations produites sont globalement cohérentes d'une annotatrice à l'autre. Certaines étiquettes émergent déjà comme plus consensuelles, plus simples à identifier (modes comportemental, désigné, montré, catégorie COLÈRE) et d'autres plus complexe (mode suggéré, type complexe). Ces observations sont approfondies chapitre 5 par la mise en œuvre d'expériences pour l'identification automatique des notions du schéma.

4) Caractérisation chiffrée du corpus annoté

Nombre d'unités annotées. Dans le corpus global, composé de 1 594 textes et annoté par six annotatrices, 6 804 unités SitEmo ont été délimitées, auxquelles sont associées 4 972 unités Expérienceur. Les émotions de type AUTRE correspondent à 1 432 unités. Dans les corpus où les causes et conséquences des émotions ont été annotées, 904 unités SitCause et 155 unités SitConséquence ont été identifiées. Les résultats détaillés par sous-corpus sont présentés par le tableau 21.

Dans les textes annotés, les causes des émotions sont ainsi bien plus fréquemment exprimées que leurs conséquences. De même, il y a toujours plus d'unités SitEmo que d'unités Expérienceur annotées : la même émotion (même catégorie émotionnelle ressentie par le même expérienceur) est en effet souvent exprimée par plusieurs unités SitEmo, toutes reliées à la même unité Expérienceur.

	Nombre d'unités				
Corpus	**SitEmo**	**Autre**	**Expérienc.**	**SitCause**	**SitConséq.**
PtitLibe	2 774	1 148	2 122	*n/a*	*n/a*
Albert	2 041	13	1 727	*n/a*	*n/a*
CovidAdultes	28	0	21	*n/a*	*n/a*
CovidJeunesse	46	1	37	*n/a*	*n/a*
FictionM2	1 587	227	858	729	103
WP7Encyclo	121	14	89	58	21
WP7Fiction	207	29	118	117	31
Total	**6 804**	**1 432**	**4 972**	**904**	**155**

Tableau 21: Nombre d'unités annotées par sous-corpus

Annotation de SitEmo par textes et phrases du corpus. Seuls 349 des textes du corpus (21,9 %) ne comportent aucune unité émotionnelle (unité SitEmo ou unité Autre). Lorsqu'un texte contient au moins une SitEmo, il en inclut au maximum 362 (textes de fiction correspondant à des ouvrages entiers) et en moyenne 5,75. Le tableau 22 expose ces informations par sous-corpus.

En revanche, la majorité des phrases du corpus (22 537 sur 28 009, soit 80,1 %) ne contiennent aucune unité émotionnelle. Les phrases comportant au moins une SitEmo en incluent en moyenne 1,54 et au maximum 8. Le tableau 23 présente les résultats détaillés par sous-corpus. Au vu de l'écrasante majorité de phrases sans unités SitEmo dans tous les sous-corpus, les nombres minimum, maximum et moyen de SitEmo par phrase ne se rapportent qu'aux phrases contenant au moins une SitEmo.

	Nombre de					
Corpus	**textes avec SitEmo**	**textes avec Autre**	**textes sans u. émo.**	**Nb. moy. SitEmo/texte**	**Nb. max. SitEmo/texte**	**Nb. min. SitEmo/texte**
PtitLibe	460	380	89	6,03	39	0
Albert	671	11	244	3,04	25	0
CovidAdultes	7	0	0	4	11	1
CovidJeunesse	15	1	16	3,07	9	0
FictionM2	6	6	0	262,67	362	193
WP7Encyclo	11	7	0	11	21	3
WP7Fiction	11	10	0	18,82	37	8
Total	**1 181**	**415**	**349**	**5,76**		

Tableau 22: Unités émotionnelles par texte

	Nombre de					
Corpus	**phrases avec SitEmo**	**phrases avec Autre**	**phrases sans u. émo.**	**Nb. moy. SitEmo/ phrase**	**Nb. max. SitEmo/ phrase**	**Nb. min. SitEmo/ phrase**
PtitLibe	1 769	1 008	9 568	1,57	6	1
Albert	1 427	13	10 318	1,43	5	1
CovidAdultes	21	0	347	1,33	3	1
CovidJeunesse	40	1	1 015	1,15	2	1
FictionM2	960	207	982	1,65	8	1
WP7Encyclo	63	14	128	1,92	8	1
WP7Fiction	127	27	179	1,63	4	1
Total	**4 407**	**1 270**	**22 537**	**1,54**		

Tableau 23: Unités émotionnelles par phrase

Modes d'expression. À l'échelle du corpus entier, les modes d'expression sont représentés de manière à peu près équilibrée, avec environ 33,1 % des unités SitEmo captant une émotion suggérée, 24,7 % une émotion désignée, 21,2 % une émotion comportementale et 21,0 % une émotion montrée. Ces informations sont détaillées pour chaque sous-corpus dans le tableau 24.

	Prop. de SitEmo			
Corpus	**comport.**	**désignées**	**montrées**	**suggérées**
PtitLibe	20,6%	34,7%	9,5%	35,2%
Albert	28,4%	19,3%	22,6%	29,8%
CovidAdultes	25,0%	57,1%	0,0%	17,9%
CovidJeunesse	4,3%	58,7%	19,6%	17,4%
FictionM2	13,7%	13,4%	37,4%	35,5%
WP7Encyclo	19,8%	18,2%	30,6%	31,4%
WP7Fiction	21,3%	23,7%	30,0%	25,1%
Total	**21,2%**	**24,7%**	**21,0%**	**33,1%**

Tableau 24: Proportions des modes d'expression parmi les SitEmo

La majorité des phrases contenant au moins une SitEmo n'expriment l'émotion que par un seul mode (3 296 phrases sur 4 407). Quand une phrase contient plusieurs modes d'expression, il arrive que les quatre modes soient représentés mais 91 % des phrases multi-modes ne contiennent que deux modes d'expression. Au sein des phrases multi-modes, l'association la plus fréquente est celle des modes désigné et suggéré (31,7 %). De manière générale, le mode suggéré est celui qui apparaît le plus souvent associé à un autre mode, comme illustré par le tableau 25.

	Parmi les phrases multi-modes	
Modes associés	**nb.**	**prop.**
désigné, suggéré	352	31,7%
comportemental, suggéré	318	28,6%
montré, suggéré	265	23,9%
comportemental, désigné	177	15,9%
désigné, montré	105	9,5%
comportemental, montré	104	9,4%

Tableau 25: Modes associés par phrase

Catégories émotionnelles. À l'échelle du corpus entier, les catégories émotionnelles les plus fréquentes sont la COLÈRE (25,2 % des unités SitEmo), la PEUR (21,8 %), la SURPRISE (18,1 %), la JOIE (17,9 %) et la TRISTESSE (14,0 %). Toutes les autres catégories représentent moins de 5 % des unités SitEmo. Ces résultats mettent sur le même plan les catégories émotionnelles renseignées aussi bien pour le trait « catégorie » que « catégorie2 ». Cependant, 89,8 % des SitEmo n'expriment qu'une seule catégorie émotionnelle (*cf.* la proportion de la catégorie « aucune » dans le tableau 26, cette valeur étant celle par défaut du trait « catégorie2 »). Les résultats détaillés par sous-corpus sont présentés tableau 26 où les catégories apparaissent mentionnées par leurs deux premières lettres, faute de place.

	Proportion de SitEmo exprimant, en trait « catégorie » ou « catégorie2 », la catégorie...											
Corpus	**AD.**	**CO.**	**CU.**	**DÉ.**	**EM.**	**FI.**	**JA.**	**JO.**	**PE.**	**SU.**	**TR.**	***auc.***
PtitLibe	4,9%	22,0%	0,4%	0,4%	4,8%	4,2%	0,0%	15,7%	30,2%	13,7%	16,7%	87,0%
Albert	3,9%	36,3%	0,2%	0,6%	0,6%	1,2%	0,0%	7,3%	18,9%	30,8%	3,7%	96,5%
CovidAdultes	3,6%	28,6%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	64,3%	3,6%	0,0%	100%
CovidJeunesse	2,2%	2,2%	2,2%	2,2%	2,2%	2,2%	2,2%	2,2%	2,2%	2,2%	2,2%	2,2%
FictionM2	2,3%	16,6%	0,6%	3,0%	5,0%	6,6%	0,4%	36,1%	9,9%	10,7%	21,7%	87,0%
WP7Encyclo	12,4%	34,7%	0,0%	0,0%	0,0%	3,3%	2,5%	5,0%	21,5%	15,7%	18,2%	86,8%
WP7Fiction	1,4%	21,3%	0,5%	4,3%	5,8%	6,8%	0,0%	21,3%	23,2%	14,5%	18,8%	82,1%
Total	**4,0%**	**25,2%**	**0,4%**	**1,2%**	**3,5%**	**3,9%**	**0,1%**	**17,9%**	**21,8%**	**18,1%**	**14,0%**	**89,8%**

Tableau 26: Proportions des catégories émotionnelles parmi les SitEmo

Sur le corpus global, 695 SitEmo articulent l'expression de deux catégories émotionnelles. Sur ce corpus, ce sont 770 phrases qui articulent l'expression de plusieurs catégories émotionnelles, avec jusqu'à quatre catégories émotionnelles dénotées par une même phrase (admiration, fierté, joie et surprise). Cependant, 90 % des phrases multi-catégories n'expriment que deux catégories à la fois. Les dix couples de catégories associées les plus fréquents sont indiqués dans le tableau 27. Le classement est présenté pour les unités SitEmo multi-catégories et les phrases multi-catégories. Bien qu'ils soient à des rangs de fréquence différents selon l'échelle considérée (unité ou phrase), les mêmes couples de catégories reviennent. Exception faite du couple formé par la TRISTESSE et la SURPRISE, 10e plus fréquent pour les SitEmo mais 15e plus fréquent pour les phrases. Le couple manquant pour les phrases – celui de la 9e position – est ADMIRATION et JOIE avec 39 occurrences, soit 5,6 % des phrases multi-catégories.

	Parmi les SitEmo multi-cat.			Parmi les phrases multi-cat.		
Catégories associées	**rang**	**nb.**	**prop.**	**rang**	**nb.**	**prop.**
FIERTÉ, JOIE	1	155	22,3%	2	122	17,6%
PEUR, TRISTESSE	2	141	20,3%	1	140	20,1%
COLÈRE, TRISTESSE	3	74	10,6%	4	66	9,5%
ADMIRATION, SURPRISE	4	70	10,1%	6	52	7,5%
EMBARRAS, TRISTESSE	5	41	5,9%	8	41	5,9%
PEUR, SURPRISE	6	40	5,8%	7	51	7,3%
COLÈRE, SURPRISE	7	27	3,9%	10	39	5,6%
JOIE, SURPRISE	8	27	3,9%	3	68	9,8%
COLÈRE, PEUR	9	20	2,9%	5	65	9,4%
SURPRISE, TRISTESSE	10	18	2,6%	15	24	3,5%

Tableau 27: Top 10 des couples des catégories émotionnelles associées

De manière générale, que l'échelle de l'unité ou de la phrase soit considérée, il est assez rare que plusieurs catégories soient exprimées en même temps. Cela s'explique sûrement d'une part par le fait que les catégories du schéma correspondent à des macro catégories (par ex. indignation, mécontentement, rage, fureur, *etc.* regroupées dans COLÈRE) et d'autre part par le fait que la grande majorité des textes du corpus sont assez courts, traitant d'une seule thématique (*cf.* tableau 15).

Lié à la catégorie émotionnelle, le type d'émotion exprimé par une unité SitEmo est majoritairement « de base » (93,1 % des unités contre 11,8 % de « complexe ») à l'échelle du corpus entier. Parmi les phrases contenant au moins une SitEmo, la majorité (93,8 %) exprime une émotion de base (seulement 12,9 % une émotion complexe). Ces chiffres prennent une fois encore en compte les valeurs des traits « type » et « type2 » indifféremment.

Marqueurs annotés par une SitEmo. Les unités linguistiques annotées par les SitEmo sont de plusieurs natures. À l'échelle du corpus entier et tous modes d'expression confondus, 50,2 % des unités correspondent à des propositions, 18,3 % à des syntagmes nominaux (SN) et 11,9 % à des syntagmes adjectivaux (SAdj). Les autres natures possibles (*cf.* tableau 28) représentent chacune moins de 10 % des segments annotés. Observée selon le mode d'expression de l'unité émotionnelle annotée, la nature permet de retrouver aisément les différents modes, comme illustré par le tableau 28.

Le mode montré apparaît ainsi clairement comme le plus hétérogène[46], le plus diversifié dans ses marqueurs, comme l'indiquait sa définition (*cf.* chapitre 3). Le comportemental qui, par essence, correspond à des descriptions d'actions ou d'états est réalisé à 83,2 % par des propositions. De même, le mode suggéré, correspondant à la description de situations, est majoritairement réalisé par des propositions (69,2 %). Le mode désigné, où propositions (26,4 %) et SN (28,5 %) sont à peu près à égalité, est celui qui est le plus exprimé par des syntagmes adjectivaux (40,0 %).

Nature	**comp.**	**désigné**	**montré**	**suggéré**
Proposition	1 202	444	210	1 558
Syntagme nominal	207	479	41	517
Syntagme prépositionnel	20	55	15	52
Syntagme adjectival	13	673	13	111
Syntagme adverbial	2	31	143	14
Interjection	0	0	70	0
Conj. de coordination	0	0	4	0
Conj. de subordination	0	0	4	0
Accumulation	0	0	11	0
Énonce averbal	0	0	14	0
Énonce clive	0	0	4	0
Énonce exclamatif	0	0	5	0
Dislocation droite	0	0	3	0
Dislocation gauche	0	0	6	0
Point d'exclamation	0	0	595	0
Points de suspension	0	0	275	0
Autre	0	0	13	0
Total	**1 444**	**1 682**	**1 426**	**2 252**

Tableau 28: Nature des unités SitEmo, selon le mode

Le tableau 29 présente, pour chaque mode d'expression, les dix déclencheurs qui lui sont le plus fréquemment associés. Les déclencheurs – ces marqueurs qui, au sein de l'unité annotée, constituent le cœur de l'expression émotionnelle – sont renseignés par les annotatrices tels qu'ils apparaissent dans le texte. Leurs formes ne sont donc pas normalisées, ce qui nécessiterait un traitement (par ex. lemmatisation) qui n'a pas été effectué ici. Les observations pouvant être faites perdent donc de leur portée, comme le souligne la présence de formes fléchies du verbe *accuser* à trois reprises dans la liste

46 Il existe de rares cas où l'annotatrice a considéré que des conjonction de coordination telles que *mais, or, mais malgré* ou des conjonctions des subordination telles que *comme si, comme* montraient en elles-mêmes l'émotion.

des déclencheurs du mode comportemental. Quelques remarques intéressantes, qui viennent confirmer celles déjà faites sur les natures des SitEmo, peuvent malgré tout être faites.

Malgré l'absence de traitements de normalisation, les déclencheurs d'émotions désignées, correspondant à des noms, des adjectifs ou des verbes, sont ainsi ceux ayant les plus fortes fréquences. Cette observation concorde une fois encore avec ce qui pouvait être attendu pour ce mode, le mieux circonscrit à l'aide de critères explicites (*cf.* section 3.1).

L'intérêt de procédures de normalisation apparaît clairement dans la liste des déclencheurs du mode montré, où les marques de ponctuation – ne disposant par de formes fléchies – atteignent les nombres d'occurrences les plus élevés, tous modes confondus. La diversité des marqueurs qui montrent l'émotion émerge également dans cette liste qui, comme pour le mode désigné, pointe vers un mode relativement bien circonscrit de part les fréquences des déclencheurs récurrents.

Les déclencheurs du mode comportemental correspondent quant à eux principalement à des verbes, ce qui est cohérent avec la prédominance de la nature « Proposition » pour les SitEmo comportementales. Le mode suggéré est celui pour lequel les nombres d'occurrences des déclencheurs sont les plus bas. Combiné à la diversité des noms et verbes déclencheurs d'une émotion suggérée, ceci souligne la diversité des situations – et par là des formes – qui réalisent ce mode difficile à appréhender.

comportemental		**désigné**		**montré**		**suggéré**	
déclencheur	**nb.**	**déclencheur**	**nb.**	**déclencheur**	**nb.**	**déclencheur**	**nb.**
manifestations	59	peur	127	!	597	guerre	22
grève	24	dangereux	55	...	251	victime	20
accusent	20	colère	52	même	40	gagné	14
fuir	19	surprise	35	...	23	découvert	14
accusé	19	danger	29	?	13	tué	14
accuse	18	heureux	24	déjà	10	victimes	13
critiqué	18	plaisir	22	juste	8	grave	13
pleurer	18	triste	21	Oh	8	exploit	12
protester	17	inquiète	20	pourtant	7	victoire	11
fête	17	craignent	18	quand même	6	remporté	11

Tableau 29: Top 10 des déclencheurs selon le mode d'expression

Expérienceurs. Que ce soit à l'échelle du corpus global ou pour chaque sous-corpus considéré séparément, la grande majorité des expérienceurs annotés sont des entités de type « Animé - Humain », comme illustré dans le tableau 30.

Corpus	**Animé-Hum.**	**Animé-Non H.**
PtitLibe	98,7%	1,3%
Albert	99,3%	0,7%
CovidAdultes	100%	0%
CovidJeunesse	100%	0%
FictionM2	99,8%	0,2%
WP7Encyclo	91,0%	9,0%
WP7Fiction	62,7%	37,3%
Total	**98,1%**	**1,9%**

Tableau 30: Contrôle de l'expérienceur selon le sous-corpus

Les unités linguistiques exprimant un expérienceur correspondent, à l'échelle du corpus global, principalement à des syntagmes nominaux (32,3 % des unités Expérienceur), des unités fantômes pointant vers la doxa ou le scripteur (25,8 %), des pronoms personnels (24,5 %) ou des noms propres (11,1 %). Les autres natures représentent moins de 3 % des unités Expérienceurs. Les résultats détaillés par sous-corpus sont présentés tableau où les corpus sont identifiés par trois lettres de leur nom. Il est tout de même intéressant de noter que la proportion d'unités fantômes chute pour les corpus FictionM2 et WP7Fiction, soient les deux corpus de textes fictionnels.

Nature	**Pti.**	**Alb.**	**CoA**	**CoJ**	**Fic.**	**WPE**	**WPF**	**Total**
N/A	27,4%	35,2%	38,1%	13,5%	4,7%	34,8%	6,8%	**25,8%**
SN	31,3%	43,3%	47,6%	24,3%	12,1%	36,0%	33,1%	**32,3%**
NP	6,5%	5,0%	0%	5,4%	34,7%	4,5%	17,8%	**11,1%**
ProPerso	28,2%	8,9%	14,3%	54,1%	44,4%	16,9%	39,8%	**24,5%**
ProRel	2,3%	3,1%	0%	0%	0,2%	3,4%	0%	**2,1%**
ProDem	1,0%	0,3%	0%	0%	0,7%	0%	0%	**0,6%**
ProIndef	1,6%	1,8%	0%	0%	0,3%	0%	0%	**1,4%**
ProPoss	0%	0,1%	0%	0%	2,7%	0%	0%	**0,5%**
Det	1,3%	1,3%	0%	2,7%	0,1%	2,2%	2,5%	**1,1%**
SAdj	0,2%	1,0%	0%	0%	0%	2,2%	0%	**0,5%**
Adverbe	0,2%	0,1%	0%	0%	0%	0%	0%	**0,1%**

Tableau 31: Nature des expérienceurs

Comme pour le déclencheur, l'entité ou les entités d'une unité Expérienceur sont renseignées en saisie libre par l'annotatrice, en mettant le texte tel qu'il apparaît dans le corpus. Pour faire une analyse plus approfondie, il faudrait une fois encore normaliser les entités pour maximiser les recoupements. Cependant, des observations intéressantes se dégagent tout de même des données non normalisées. Ainsi, dans les corpus journalistiques, la doxa et le scripteur apparaissent dans les dix entités les plus fréquentes. Les autres entités sont des syntagmes nominaux, parfois assez longs (par ex. « les syndicats Force ouvrière (FO) et la Confédération française démocratique du travail (CFDT) » est une entité du corpus CovidAdultes). Les entités sont variées et reflètent les diverses thématiques abordées au sein des corpus (par ex. « les gilets jaunes », « les animaux », « des gens », « les Français » sont des entités du corpus P'tit Libé : « la population », « les policiers », « Donald Trump » sont des entités du corpus Albert ; dans les corpus Covid, les entités « l'OMS » et « Une infirmière qui travaille en réanimation » apparaissent). Pour le corpus encyclopédique, le même constat est effectué, avec encore plus de diversité du point de vue des entités (par ex. « L'ara de Spix », « le diable », « Les comédiens »). Ce corpus étant bien plus restreint que les corpus journalistiques, chaque thématique émerge donc probablement plus facilement dans les listes d'entités – tandis que seuls les sujets les plus récurrents se donnent à voir dans les listes des entités pour les corpus journalistiques. Dans les corpus de fiction, les entités les plus fréquentes correspondent principalement à des noms propres référant aux personnages des histoires (par ex. « Flavia », « Elisa », « Anna » sont des entités du corpus FictionM2 ; « Verte », « Chien Pourri » et « Tristan » du corpus WP7Fiction). La doxa et le scripteur sont absents des tops 10 pour ces textes mais l'apparition du narrateur peut être constatée. Les résultats détaillés sont présentés en Annexe IV.

Conclusion

Grâce à un ensemble de critères regroupés dans le guide d'annotation et à l'intervention de six annotatrices, le schéma d'annotation décrit au chapitre 3 a été appliqué sur un corpus de 1 594 textes. Dans ce corpus hétérogène, constitué de textes encyclopédiques, fictionnels et journalistiques traitant tous de thématiques variées, 6 804 unités SitEmo ont été délimitées, réparties entre 4 407 phrases.

Cette systématisation de l'analyse des émotions a été évaluée grâce à la mise en place d'un calcul d'AIA partiel. Les taux d'accord obtenus pour l'annotation des catégories

émotionnelles et des modes d'expression, à l'échelle de la phrase, sont comparables voire meilleurs que ceux obtenus dans d'autres travaux de TAL (pour les catégories émotionnelles).

De premières observations portant sur l'expression des émotions peuvent être effectuées grâce à l'étude de caractéristiques chiffrées des annotations produites. Ces analyses sont approfondies chapitre 7. D'autres observations pouvant être menées à l'aide du schéma, impliquant notamment la temporalité, y sont aussi détaillées.

Le corpus annoté est également le point de départ pour la mise en œuvre d'expériences informatiques visant à évaluer la possibilité de développer un système d'analyse automatique des émotions à l'aide des notions du schéma. Ces expériences qui reposent sur l'entraînement de classifieurs sont présentées au chapitre 5.

Chapitre 5 Entraînement de modèles d'analyse automatique des émotions

Le schéma d'annotation des émotions que je propose caractérise les unités linguistiques exprimant une émotion selon trois notions principales : la catégorie émotionnelle, le type d'émotion et le mode d'expression. Le présent chapitre expose les expérimentations d'apprentissage automatique conduites à partir du corpus annoté constitué par l'application manuelle du schéma sur 1 596 textes. L'objectif de ces expériences est d'explorer les possibilités offertes par le corpus – et donc le schéma – du point de vue de l'analyse automatique de la dimension émotionnelle des textes.

L'analyse des émotions est appréhendée ici comme un problème de classification divisé en cinq tâches de classification de difficulté croissante. Une vue générale des grands principes de la mise en œuvre des expérimentations est dressée en section 1 avant de définir plus en amont, en section 2, les différentes tâches de classification traitées. Les sections 3 et 4 font respectivement état des transformations du corpus annoté en corpus d'apprentissage et des choix méthodologiques nécessaires pour mener à bien les expériences informatiques. Les résultats des expérimentations sont enfin observés par le biais de performances chiffrées à la section 5.

Il est à noter que ce chapitre a volontairement été conçu comme technique, afin de présenter avec précision les expérimentations informatiques, comme il est d'usage dans le domaine du TAL. Le chapitre 6 propose pour sa part une analyse qualitative des résultats obtenus par l'un des modèles développés lors des expérimentations.

1) Vue d'ensemble

Comme à l'accoutumée en TAL (*cf.* chapitre 2), j'aborde l'analyse automatique des émotions comme un problème de *classification* des émotions dans les textes, c'est-à-dire comme la prédiction par un modèle de catégories (appelées *classes* ou *étiquettes*) à partir d'objets donnés en entrée (en l'occurrence, des segments textuels). Il s'agit donc d'élaborer un modèle informatique – un *classifieur* – capable d'apprendre à effectuer ces prédictions sur la base d'exemples qui lui sont fournis.

L'objectif principal des expérimentations informatiques rapportées ici, directement lié à la problématique de ma thèse, est de sonder les perspectives offertes par mon schéma d'annotation du point de vue de l'analyse automatique des émotions, perçues comme un facteur de complexité des textes. Ainsi, les exemples servant à développer le modèle sont dérivés du corpus annoté en émotions construit par l'application manuelle de mon schéma. De manière plus générale, en lien avec les ambitions plus larges du projet TextToKids dans lequel s'insère ma thèse, l'idée est de proposer un système dont les analyses en émotions contribuent à l'étude de la complexité des textes pour enfants. Un classifieur issu des expérimentations relatées dans ce chapitre a donc été intégré dans la chaîne de traitement automatique des textes de TextToKids sous la forme du processeur EMOTYC, décrit chapitre 1.

1.1) Décomposition du problème

Pour explorer l'analyse automatique des émotions, j'ai étudié cinq tâches de classification de phrases. L'idée était d'expérimenter des tâches simples, puis d'aller progressivement vers des tâches plus difficiles. La liste ordonnée de ces tâches est la suivante :

A) la prédiction du caractère émotionnel (ou non) de la phrase ;

B) la prédiction des modes d'expression contenus dans la phrase ;

C) la prédiction des types d'émotions exprimés dans la phrase ;

D) la prédiction des catégories émotionnelles dénotées dans la phrase ;

E) la prédiction simultanée du caractère émotionnel, des modes d'expression, des types d'émotions et des catégories émotionnelles de la phrase.

La tâche E est celle que je vise à terme, et ce pour deux raisons. D'une part, elle devrait permettre aux différentes prédictions de s'entre-aider (selon l'hypothèse que, par exemple, le mode d'expression et la catégorie émotionnelle ne sont pas des notions indépendantes). D'autre part, il est plus pratique de disposer d'un seul modèle multi-tâches plutôt que de plusieurs modèles mono-tâches.

1.2) Développer des classifieurs

Pour développer les classifieurs en émotions, j'ai choisi d'utiliser des techniques d'*apprentissage automatique*, reconnues en TAL depuis de nombreuses années pour leur efficacité. Grâce à des principes statistiques, ces techniques produisent des modèles in-

formatiques qui, à partir d'exemples de départ, opèrent des généralisations leur permettant ensuite de classer des exemples nouveaux. Mettre en application ces techniques consiste ainsi à *apprendre* à partir de données initiales (ici, mon corpus annoté en émotions) un *modèle* capable de répondre à une tâche donnée (par exemple prédire le caractère émotionnel ou non d'une phrase). On parle aussi d'*entraîner* un modèle. Le développement de classifieurs repose sur trois grandes étapes : la constitution d'un corpus d'apprentissage, l'entraînement des modèles et enfin leur évaluation. Ces grandes étapes de mon travail sont illustrées de manière synthétique par la figure 10.

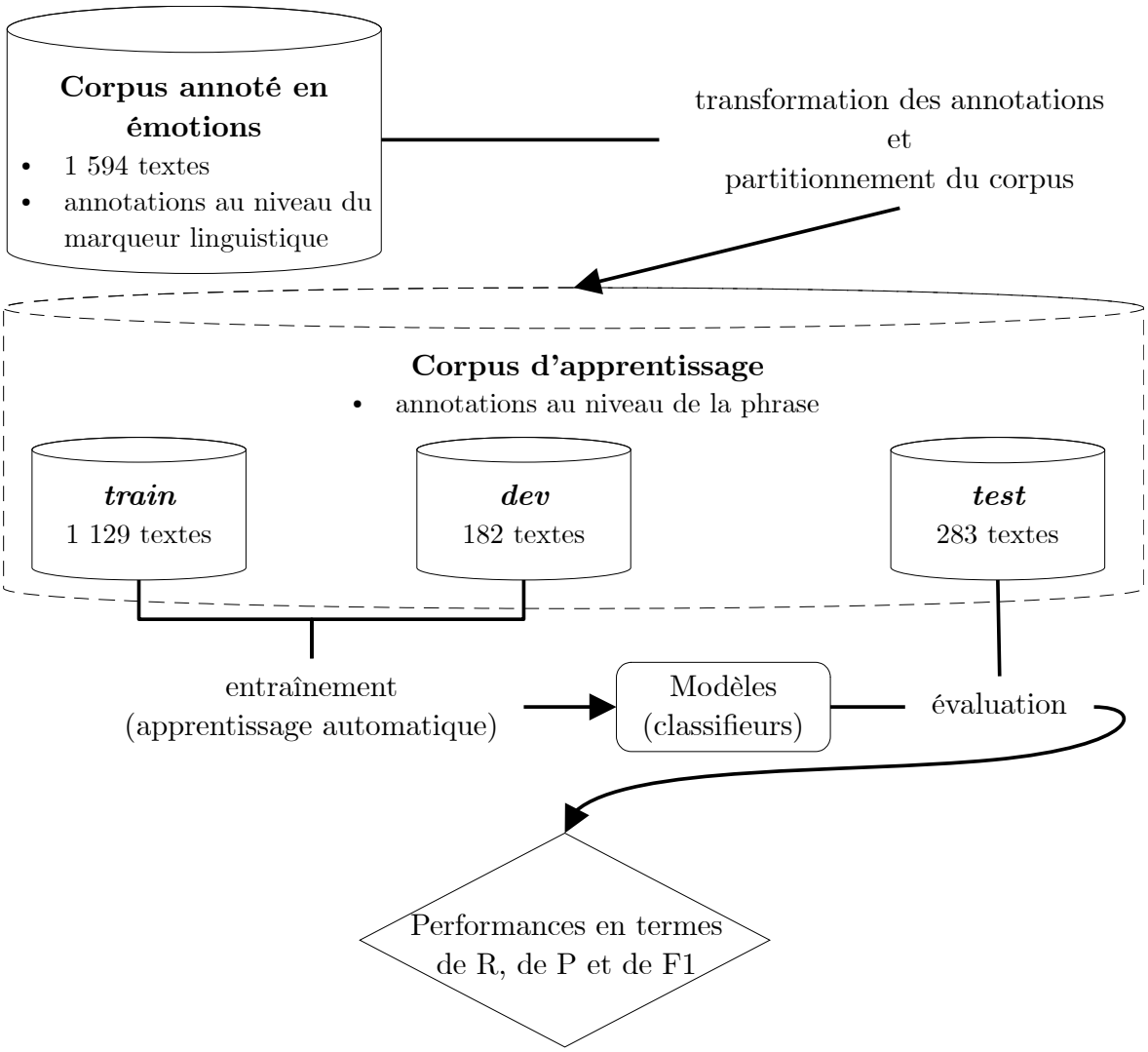


Figure 10: Grandes étapes du développement de classifieurs

Étape 1 : adaptation du corpus. Afin d'être utilisable pour l'entraînement de classifieurs, mon corpus annoté en émotions a été transformé en *corpus d'apprentissage*. Dans un premier temps, les informations d'annotation situées au niveau des marqueurs linguistiques (caractère, mot ou groupe de mots) ont été transposées au niveau de la phrase pour permettre une classification de phrases. Dans un second temps, le corpus annoté a été *partitionné* en trois ensembles, pratique nécessaire pour évaluer la capacité du modèle à généraliser à de nouveaux exemples. L'ensemble d'entraînement (noté *train*) est celui à partir duquel les modèles sont appris et concentre donc la majorité des données disponibles pour la tâche. L'ensemble de développement (noté *dev*) intervient tout au long de l'élaboration des modèles pour les évaluer et guider ainsi leur développement. Enfin, l'ensemble de test (noté *test*) n'est sollicité qu'une fois le processus de développement totalement terminé pour évaluer les modèles une dernière fois sur des données nouvelles. Les valeurs issues de cette évaluation finale sont celles retenues pour juger les performances des modèles.

Étape 2 : entraînement des classifieurs. Les classifieurs sont appris à partir des ensembles de *train* (pour alimenter l'apprentissage) et de *dev* (pour savoir quand le modèle ne progresse plus dans son apprentissage). La technique principale servant à mes expérimentations relève de l'*apprentissage profond*, c'est-à-dire de modèles utilisant des *réseaux de neurones* composés de plusieurs couches de neurones. L'outil précis que j'ai retenu est le modèle CamemBERT (Martin et al., 2020) qui appartient à la famille des *transformeurs* (architecture de type RoBERTa (Y. Liu et al., 2019)), un type de modèles d'apprentissage profond particulièrement adapté pour traiter des données dont l'ordre a une importance, comme les données langagières. CamemBERT est plus précisément un *modèle de langue contextuel du français* : entraîné sur un très grand volume de données textuelles en français (138 Go) à prédire les mots manquants dans une séquence, il a enregistré des connaissances sur la distribution des mots de la langue et leur usage en contexte. Son application la plus simple est de prédire, à partir d'une séquence de mots donnée, le mot suivant. Mais CamemBERT peut également être exploité pour faire de la classification en poursuivant son entraînement sur de nouvelles données et en ajoutant une *couche de classification* à la fin du modèle (en remplacement de la toute dernière couche qui prédisait les probabilités des mots manquants). Les couches profondes du modèle qui rendent compte de la distribution des mots de la langue restent alors intactes et servent d'appui pour la couche superficielle qui est quant à elle modifiée de sorte à répondre à la tâche de

classification voulue (par exemple prédire le caractère émotionnel d'une phrase). Ce procédé est appelé *affinage* ou *ajustement* (*fine-tuning*). Mes expérimentations reposent sur l'ajustement de CamemBERT pour chacune des cinq tâches de classification que j'ai définies.

Pourquoi CamemBERT ? Puisque CamemBERT a été pré-entraîné sur de très nombreux textes, les classifieurs résultant de son affinage sont habituellement robustes : en bénéficiant des connaissances acquises par CamemBERT lors du pré-entraînement, les classifieurs sont plus à même de généraliser correctement et donc de s'adapter à des exemples nouveaux. Ils ont par ailleurs plus de chance de capter des phénomènes linguistiques fins, tels que ceux qui m'intéressent ici (par exemple les modes d'expression des émotions). De plus, les modèles d'architecture transformeur constituent aujourd'hui l'état de l'art pour de nombreuses tâches de TAL, notamment la classification en émotions (C. Liu et al., 2019; Öhman et al., 2020).

Étape 3 : évaluation. Une étape cruciale de mes expérimentations est l'évaluation des performances des modèles, c'est-à-dire de leur capacité à répondre correctement à la tâche visée. Pour cela, j'ai choisi trois métriques fréquemment employées en TAL :

- le *rappel*, noté R. Il indique la capacité du classifieur à retrouver les éléments appartenant à une classe donnée ;
- la *précision*, notée P. Elle indique la capacité du modèle à prédire correctement l'appartenance à une classe donnée ;
- et la *F-mesure*, notée F1. Elle correspond à la moyenne harmonique du rappel et de la précision et permet ainsi d'avoir une vision d'ensemble équilibrée des performances du modèle.

Notons que pour juger de la pertinence d'utiliser des modèles transformeurs, il est nécessaire de comparer les performances de mes classifieurs avec celles de modèles plus simples, dits « *baselines* » (*cf.* section 4.5), qui indiquent le résultat atteignable en utilisant une méthode simple pour répondre à une tâche donnée.

Comment obtenir les meilleurs classifieurs ? Afin d'obtenir les classifieurs les plus performants possible pour chaque tâche, plusieurs critères paramétrables sont manipulés lors des entraînements. Certains de ces critères se rapportent aux caractéristiques des données à partir desquelles les modèles sont appris (type d'entrée fourni aux modèles, taille de l'ensemble d'entraînement) et d'autres concernent le type d'entraînement réalisé (durée de l'entraînement, pondération des classes lors de l'appren-

tissage). Je me suis appuyée sur la décomposition de la tâche d'analyse des émotions en cinq problèmes de difficulté croissante pour sélectionner de manière efficace les critères à mobiliser pour l'entraînement des classifieurs, réduisant ainsi le nombre d'expériences. Un critère n'ayant pas amélioré les performances des classifieurs sur une tâche est par exemple écarté de l'élaboration des modèles des tâches de difficulté supérieure.

La suite du chapitre détaille les choix techniques et la mise en œuvre concrète des expérimentations esquissés ici dans leurs grandes lignes.

2) Définition des tâches de classification

Le développement de classifieurs en émotions s'effectue à partir du corpus annoté qui rend compte d'une analyse manuelle des émotions, opérée grâce au schéma d'annotation décrit chapitre 3. Pour automatiser cette analyse, seules quelques notions du schéma ont été conservées. La tâche globale d'analyse automatique de la dimension émotionnelle d'un texte a quant à elle été décomposée en cinq tâches de classification. Ces adaptations sont présentées respectivement en section 2.1 et 2.2.

2.1) Sélection de notions

L'analyse en émotions que permet mon schéma d'annotation est dense et implique des concepts jusqu'ici jamais analysés automatiquement en TAL (par ex. mode d'expression, type d'émotion). L'objectif des expériences informatiques présentées dans ce chapitre est de tester quels sont les éléments du schéma qu'il est possible de retrouver automatiquement à partir de l'annotation faite au sein du corpus décrit chapitre 4. Il était donc nécessaire de la simplifier, notamment en ne conservant que quelques notions clés du schéma, telle que le mode d'expression des émotions, véritable apport de ma modélisation par rapport à d'autres approches de TAL (*cf.* chapitre 2).

Une analyse manuelle riche et complexe. Seuls les éléments relatifs au schéma et à son application jugés nécessaires pour comprendre les expériences informatiques sont rappelées ici (se reporter au chapitre 3 pour une présentation plus détaillée).

Le schéma d'annotation des émotions proposé dans cette thèse articule de nombreuses notions à travers le repérage de cinq types d'unités, spécifiées par des traits et pouvant être liées entre elles par des relations. Les informations émotionnelles à proprement parler sont captées par les unités de type SitEmo, caractérisées par cinq traits principaux (*Catégorie/Catégorie2*, *Type/Type2* et *Mode*) et par les unités Autre[47]. Lors de la mise en œuvre du schéma, l'annotateur procède à une annotation à l'échelle du marqueur linguistique. Il délimite manuellement les unités dans le texte, en respectant des consignes qui conduisent parfois à des phénomènes de chevauchement ou d'emboîtement d'unités, voire à des unités discontinues (*cf.* chapitre 4). L'analyse sémantique effectuée lors de l'application manuelle du schéma par les annotateurs humains est ainsi riche, fine et complexe.

Simplification de l'analyse pour son automatisation. Pour tester les possibilités d'automatisation de l'analyse des émotions à partir de mon schéma, la priorité a été de simplifier l'analyse en se focalisation sur l'identification de concepts clés du schéma. Seules trois notions, centrales pour la description des émotions en tant que facteur de complexité sémantique des textes (*cf.* chapitre 1), ont donc été retenues : la *catégorie émotionnelle*, le *type d'émotion* et le *mode d'expression*.

Une quatrième notion, celle de *caractère émotionnel*, a été introduite. Elle rend compte du niveau d'analyse le plus simple : l'identification de la présence d'informations émotionnelles, sans typage de ces informations.

Toujours dans une démarche de simplification, l'échelle d'analyse a été repensée : d'une tâche visant à délimiter et typer manuellement des unités dans un texte, on passe à de simples tâches de classification de phrases. Le processus de transformation des annotations de marqueurs linguistiques en informations associées à la phrase est détaillé à la section 3.1.

Les expériences informatiques rapportées ici se concentrent donc sur le repérage des informations émotionnelles contenues dans une phrase à travers quatre notions :

- le *caractère émotionnel*, qui renvoie à la présence/absence d'informations émotionnelles ;

47 Pour rappel, les unités Autre interviennent pour capter les unités émotionnelles ne relevant pas de l'une des onze catégories émotionnelles prises en charge par les SitEmo, par exemple la HAINE, le MÉPRIS ou le SOULAGEMENT. Les informations de catégorie émotionnelle, de type d'émotion et de mode d'expression ne sont pas renseignés pour les unités Autre.

- le *mode d'expression*. Cette notion renvoie à quatre façons d'exprimer les émotions : par la description d'un comportement émotionnel (mode *comportemental*), par l'emploi d'un terme du lexique émotionnel (mode *désigné*), par l'usage de caractéristiques des énoncés traduisant l'état émotionnel de l'énonciateur (mode *montré*) ou par la description de situations conventionnellement associées à un ressenti émotionnel (mode *suggéré*) ;
- la *type d'émotion*. Cette notion envoie aux concepts d'émotion *de base* et d'émotion *complexe* ;
- la *catégorie émotionnelle*. Cette notion est réalisée par les onze catégories captées par les unités SitEmo (COLÈRE, DÉGOÛT, JOIE, PEUR, SURPRISE, TRISTESSE, ADMIRATION, CULPABILITÉ, EMBARRAS, FIERTÉ, JALOUSIE) et par la catégorie AUTRE, ajoutée pour rendre comptes des annotations émotionnelles issues des unités Autre.

2.2) Cinq tâches de classification

Dans le cadre de cette thèse, l'analyse automatique des émotions est fondée sur quatre notions : le *caractère émotionnel* d'une phrase, le *mode d'expression* des émotions, le *type d'émotion* et la *catégorie émotionnelle*. Cinq tâches de classification de difficulté croissante ont ainsi été définies, de sorte à explorer progressivement la tâche d'analyse automatique de la dimension émotionnelle d'un texte.

Chaque notion correspond à un ensemble de *classes* à prédire. Pour quatre des tâches, plusieurs classes peuvent être associées à une même phrase. Il est d'usage alors de parler d'« étiquette » (et non de « classe ») et de classification « multi-étiquettes ». Les cinq tâches qui composent les expérimentations informatiques sont, de la plus simple à la plus difficile :

- tâche A – prédiction du *caractère émotionnel* de la phrase. La classe associée à cette tâche est « émotionnelle » ;
- tâche B – prédiction des *modes d'expression* mobilisés dans la phrase (tâche multi-étiquettes). Les quatre étiquettes associées à cette tâche sont « comportemental », « désigné », « montré » et « suggéré » ;
- tâche C – prédiction des *types d'émotion* exprimés dans la phrase (tâche multi-étiquettes). Les deux étiquettes associées à cette tâche sont « base » et « complexe » ;

- tâche D – prédiction des *catégories émotionnelles* exprimées dans la phrase (tâche multi-étiquettes). Les douze étiquettes associées à cette tâche sont « autre », « admiration », « colère », « culpabilité », « dégoût », « embarras », « fierté », « jalousie », « joie », « peur », « surprise » et « tristesse » ;
- tâche E – prédiction simultanée du *caractère émotionnel*, des *modes d'expression*, des *types d'émotions* et des *catégories émotionnelles* (tâche multi-étiquettes). Les étiquettes associées à cette tâche sont les dix-neufs étiquettes impliquées dans l'ensemble des tâches A à D.

Les prédictions qu'un classifieur effectue pour une classe donnée sont binaires, c'est-à-dire qu'elles prennent la forme de valeurs booléennes traduisant la présence (prédiction d'un 1) ou l'absence (prédiction d'un 0) de la classe. Une tâche de classification multi-étiquettes visant à prédire n étiquettes implique pour sa part à n prédictions binaires. Les classifieurs prédisant le mode d'expression effectuent donc quatre prédictions binaires, une pour chaque classe de mode.

La mise en œuvre des tâches de classification A à E repose sur un corpus d'apprentissage décrit section 3 et implique des choix méthodologiques exposés section 4.

3) De corpus annoté à corpus d'apprentissage

Pour permettre l'entraînement de classifieurs, le corpus annoté en émotions a subi deux modifications. Tout d'abord, les annotations manuelles ont été transformées afin d'être exploitables pour répondre aux cinq tâches de classification définies. Ensuite, le corpus a été divisé en ensembles d'entraînement (*train*), de développement (*dev*) et de test (*test*), nécessaires aux méthodes d'apprentissage automatique. Ces adaptations sont présentées respectivement aux sections 3.1 et 3.2.

3.1) De marqueurs linguistiques annotés à phrases annotées

Dans le corpus annoté, les informations d'annotation sont portées par des unités délimitées manuellement au sein des textes à l'échelle du marqueur linguistique. Afin de pouvoir être exploitées pour la mise en œuvre des tâches A à E, qui visent des prédictions à l'échelle de la phrase, ces informations doivent dans un premier temps être transposées au niveau de la phrase. Elles doivent ensuite être transformées en un format utilisable par les méthodes d'apprentissage automatique.

Transposition des informations d'annotation au niveau de la phrase. Résultant de l'application manuelle du schéma *via* la plateforme Glozz (Widlöcher & Mathet, 2012), le corpus annoté est au format Glozz. Pour chacun des textes du corpus, chaque unité annotée est donc définie par l'intervalle $[a,b]$ formé par ses indices[48] de début et de fin dans le texte. J'ai traité automatiquement[49] les données annotées de sorte que chaque phrase soit elle aussi définie par l'intervalle de ses indices de début et de fin $[i,j]$. Une unité d'indices [a,b] est alors contenue dans une phrase d'indices [i,j] si les relations suivantes sont respectées : $i \leq a \leq j$ et $i \leq b \leq j$. Chaque phrase du corpus est ainsi associée à la liste des unités SitEmo et des unités Autre qu'elle contient, autrement dit les informations d'annotation nécessaires pour la mise en œuvre des expérimentations informatiques.

Transformation des informations d'annotation en classes. Dans leur ensemble, les classifieurs décrits dans ce chapitre reposent sur dix-neufs classes (*cf.* section 2.2). Leur entraînement nécessite de savoir si une phrase donnée réalise chacune de ces classes. Pour chaque phrase du corpus, les informations émotionnelles issues de ses unités SitEmo et Autre sont transformées en une liste de dix-neuf booléens, toujours ordonnée de la même manière, où chaque booléen renvoie à une classe et indique sa présence. Par exemple, pour une phrase exprimant la catégorie JOIE selon les modes désigné et suggéré telle que « Luc est ravi de ses merveilleuses vacances. », la liste représentant les dix-neufs classes sera celle montrée figure 11. Dans leur ordre d'apparition, les chiffres 1 visibles sur cette figure traduisent la présence des classes « émotionnelle », « désigné », « suggéré », « base » et « joie ».

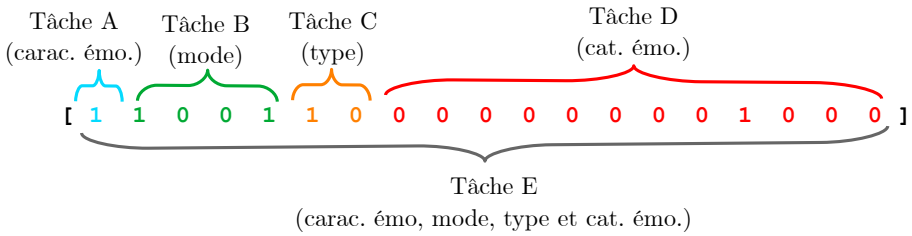


Figure 11: Représentation des informations émotionnelles d'une phrase

Les valeurs de ces listes de booléens sont ensuite filtrées afin de conserver uniquement celles utiles à la tâche de classification en cours de traitement (*cf.* figure 11).

48 Dans le cas présent, un *indice* est un numéro associé à chaque caractère du texte annoté, c'est-à-dire chaque lettre, chaque espace, *etc.* du texte. Le premier caractère a pour indice 0, le second 1, le troisième 2, *etc.*

49 Les grandes étapes de fonctionnement du script opérant ces traitements sont présentées en .

Le tableau 32 synthétise, pour chaque tâche de classification, la notion générale et les classes précises que les modèles visent à prédire ainsi que les informations d'annotation utiles pour l'entraînement de ces modèles.

Tâche	**Notion à prédire**	**Nb de classes**	**Classes à prédire**	**Informations d'annotation utiles**
A	caractère émotionnel	1	émotionnelle	Présence d'au moins une unité SitEmo ou d'au moins une unité Autre
B	mode d'expression	4	désigné, comportemental, montré, suggéré	Ensemble des modes d'expression des SitEmo de la phrase
C	type d'émotion	2	base, complexe	Ensemble des types d'émotion des SitEmo de la phrase
D	catégorie émotionnelle	12	autre, admiration, colère, culpabilité, dégoût, embarras, fierté, jalousie, joie, peur, surprise, tristesse	Ensemble des catégories émotionnelles des SitEmo de la phrase et présence d'au moins une unité Autre
E	*notions des tâches A à D*	19	*classes des tâches A à D*	*Toutes les informations utiles pour les tâches A à D*

Tableau 32: Correspondances entre tâche, notion à prédire, classes à prédire et informations d'annotation utiles

Le partitionnement du corpus a été opéré une fois les annotations adaptées.

3.2) Partitionnement du corpus

Le corpus annoté en émotions a été partitionné en trois ensembles selon les proportions suivantes : 70 % des phrases dans le *train*, 10 % dans le *dev* et 20 % dans le *test*.

Contraintes pour partitionner un corpus hétérogène. Le corpus annoté est composé de 1 594 textes, répartis en sept sous-corpus dont les caractéristiques sont détaillées au chapitre 4. Le tableau 33 présente des caractéristiques chiffrées de chaque sous-corpus, sachant qu'une phrase est dite « émotionnelle » quand elle contient au moins une unité SitEmo ou une unité Autre.

Corpus	**Genre**	**Nb textes (nb tokens)**	**Nb phrases**	**Nb phrases émotionnelles**	**Nb phrases non émotionnelles**
Albert	Journalistique	917 (238 839)	11 854	1 439	10 415
CovidAdultes	Journalistique	7 (9 618)	368	21	347
CovidJeunesse	Journalistique	31 (15 981)	1 056	41	1 015
FictionM2	Fictionnel	6 (24 975)	2 075	1 093	982
PtitLibe	Journalistique	611 (217 619)	12 130	2 562	9 568
WP7Encyclo	Encyclopédique	11 (4 195)	199	71	128
WP7Fiction	Fictionnel	11 (4 049)	327	147	180
Total		**1 594 (515 276)**	**28 009**	**5 374**	**22 635**

Tableau 33: Propriétés du corpus annoté, selon les sous-corpus

Pour répartir au mieux ces données hétérogènes, plusieurs contraintes ont été fixées :

1) toutes les phrases issues d'un même texte doivent se trouver dans le même ensemble ;
2) tous les textes des corpus Covid doivent se trouver dans le *test* ;
3) tous les textes diffusés auprès d'enfants dans le cadre des expériences du WP7 du projet TextToKids (*cf.* chapitre 1) doivent se trouver dans le *test*.

Les contraintes 1 et 2 permettent d'observer la robustesse des classifieurs. Les modèles s'appuient en effet parfois trop sur les spécificités des données sur lesquelles ils ont été appris. Leurs performances seront alors bonnes sur des textes proches du *train* et moindre sur des textes éloignés. Regrouper toutes les phrases issues d'un même texte dans le même ensemble minimise les recoupements entre *train*, *dev* et *test*. Cela assure donc que les modèles soient évalués sur des données différentes de celles d'apprentissage. Par ailleurs, les corpus Covid n'abordent qu'une seule thématique quasiment absente des autres sous-corpus, l'épidémie de Covid-19, et sont de plus majoritairement issus de sources (*Libération*, *Mon quotidien*, *etc.*) présentes nulle part ailleurs dans le corpus annoté. Avoir l'intégralité des corpus Covid dans le *test* garantit donc un fort niveau de nouveauté par rapport aux données d'entraînement.

La contrainte 3 permet quant à elle de confronter, *in fine*, les analyses des classifieurs et la réalité terrain. Si les neuf textes diffusés dans les écoles rejoignent systématiquement le *test*, les autres textes analysés dans le cadre du WP7 sont répartis entre le *train*, le *dev* et le *test* pour garantir une certaine diversité à chacun de ces ensembles.

Principe du partitionnement. Afin de faciliter le respect de la contrainte 1, la séparation des données a été effectuée sur les textes entiers. Pour les ensembles de *train*, de *dev* et de *test*, les textes ont ainsi été itérativement tirés au sort, sans remise, parmi l'ensemble des textes restant après avoir inclus dans le *test* ceux devant impérativement s'y trouver (respect des contraintes 2 et 3). La validation du partitionnement a été effectuée en deux étapes. La première, automatique, contrôlait la répartition des données selon le nombre de phrases[50]. La seconde, manuelle, vérifiait que les classes utiles aux tâches A à E étaient réparties assez équitablement entre les trois ensembles.

Caractéristiques des ensembles de *train*, de *dev* et de *test* retenus. Les caractéristiques du partitionnement retenu sont présentées dans les tableaux 34 à 36. Le tableau 34 présente le *train*, le *dev* et le *test* en termes de nombre de textes, de tokens, de phrases et de phrases émotionnelles. Il fait apparaître la part de ces ensembles représentée par chacun des sept sous-corpus annotés. La répartition des textes et phrases du corpus annoté global est présentée tableau 35 sous forme de proportions.

	Train			***Dev***			***Test***		
Corpus	**Nb textes (nb tokens)**	**Nb phrases**	**Nb phr. émo.**	**Nb textes (nb tokens)**	**Nb phrases**	**Nb phr. émo.**	**Nb textes (nb tokens)**	**Nb phrases**	**Nb phr. émo.**
Albert	655 (173 908)	8 519	1 077	120 (30 154)	1 517	160	142 (34 777)	1 721	202
CovidAdultes	0	0	0	0	0	0	7 (9 618)	368	21
CovidJeunesse	0	0	0	0	0	0	31 (15 981)	1 056	41
FictionProjetM2	5 (22 002)	1 810	899	0	0	0	1 (2 973)	265	194
PtitLibe	458 (160 393)	8 976	1 881	59 (21 502)	1 164	245	94 (35 724)	1 990	436
WP7Encyclo	5 (1 991)	90	30	1 (337)	17	6	5 (1 867)	92	35
WP7Fiction	6 (2 251)	158	65	2 (660)	72	27	3 (1 138)	96	55
Total	**1 129 (360 545)**	**19 553**	**3 952**	**182 (52 653)**	**2 770**	**438**	**283 (102 078)**	**5 588**	**984**

Tableau 34: Caractéristiques brutes du train, du dev et du test, selon les sous-corpus

	Prop. dans le *train* (%)	**Prop. dans le *dev* (%)**	**Prop. dans le *test* (%)**
Textes	70,8	11,4	17,8
Phrases	70,1	9,9	20,0

Tableau 35: Répartition des textes et phrases entre le train, le dev et le test

Enfin, le tableau 36 montre, pour chacun des ensembles, la proportion de phrases représentée par chaque classe. Le *train* comporte par exemple 20,21 % de phrases émotionnelles. Mises à part les classes « émotionnelle » et « base », toutes les autres

50 Une marge de plus ou moins 25 phrases était autorisée. Ce seuil a été choisi d'après l'écart-type de la moyenne de phrases par texte du corpus global (moyenne de 17,6 ; écart-type de 25,2).

classes sont présentes en de très faibles proportions (moins de 10 % des phrases de chaque ensemble). Six classes notamment sont représentées par moins de 1 % des phrases du *train* : « admiration », « culpabilité », « dégoût », « embarras », « fierté » et « jalousie ». Cette dernière est particulièrement rare avec seulement cinq phrases qui la réalisent dans le *train*, une dans le *dev* et une dans le *test*.

Classe	**Proportion dans le**		
	train	***dev***	***test***
émotionnelle	20,2 %	15,8 %	17,6 %
comportemental	4,6 %	3,6 %	4,3 %
désigné	5,3 %	5,2 %	5,7 %
montré	3,6 %	2,3 %	3,5 %
suggéré	7,1 %	5,8 %	6,3 %
base	15,4 %	12,6 %	13,9 %
complexe	2,0 %	2,1 %	2,3 %
admiration	0,6 %	1,1 %	1,0 %
autre	5,0 %	3,2 %	3,7 %
colère	4,6 %	3,2 %	3,4 %
culpabilité	0,1 %	0,0 %	0,1 %
dégoût	0,2 %	0,3 %	0,2 %
embarras	0,6 %	0,6 %	0,6 %
fierté	0,7 %	0,4 %	0,9 %
jalousie	0,0 %	0,0 %	0,0 %
joie	3,2 %	2,3 %	3,6 %
peur	3,8 %	3,3 %	3,8 %
surprise	3,0 %	3,1 %	2,5 %
tristesse	2,5 %	2,0 %	2,5 %

Tableau 36: Répartition des classes au sein du train, du dev et du test

4) Choix méthodologiques

La série d'expérimentations informatiques dont il est question dans ce chapitre repose sur des choix méthodologiques impliquant : les techniques pour apprendre les classifieurs ; les mesures pour évaluer leurs performances ; les facteurs à faire varier pour obtenir les classifieurs les plus efficaces possibles ; le déroulé des expérimentations ; et les modèles « simples » (*baselines*) retenus pour positionner les résultats des classifieurs. Ces choix sont présentés respectivement dans les sections 4.1 à 4.5.

4.1) Outils d'apprentissage

Choix d'un modèle d'apprentissage profond : CamemBERT. En raison de sa robustesse et de son efficacité (*cf.* section 1.2), j'ai choisi comme outil pour développer mes classifieurs en émotions un modèle d'apprentissage profond de type transformeur : CamemBERT (Martin et al., 2020), et plus précisément sa version *CamemBERT-base* diffusée par le site HuggingFace[51]. Ce modèle de langue contextuel du français comporte 1 couche de neurones qui transforme en vecteurs (plongements) les mots qui lui sont fournis en entrée et 12 couches de neurones cachées, déjà pré-entraînées et ayant ainsi enregistré des connaissances sur la distribution des mots de la langue française et leur usage en contexte. Dans le cadre de la démarche adoptée ici, l'entraînement des classifieurs consiste à *ajuster* CamemBERT (*fine-tuning*), c'est-à-dire à ajouter une couche, la *couche de classification*. Seuls les poids de cette couche sont modifiés lors de l'entraînement. Pour gérer l'apprentissage, j'ai utilisé l'API `Trainer`[52] proposée par HuggingFace avec les paramètres par défaut de l'API[53].

Les entrées des modèles : des phrases tokenisées. Les phrases fournies en entrée aux classifieurs – pour l'entraînement des modèles ou pour leur application – sont des phrases tokenisées, sans normalisation de la casse, par le tokeniseur associé à CamemBERT, disponible sur HuggingFace. La première couche de CamemBERT se charge ensuite de représenter les entrées de manière numérique (vecteurs de mots).

4.2) Mesures d'évaluation

Les mesures de rappel (notée R), de précision (notée P) et de F-mesure (notée F1) sont utilisées pour évaluer les performances des classifieurs, classe par classe. Les performances de prédiction d'une classe C donnée seront calculées comme suit :

$$\mathrm{R} = \frac{nombre\ d'items\ correctement\ prédits\ comme\ appartenant\ à\ C}{nombre\ d'items\ appartenant\ à\ C}$$

$$\mathrm{P} = \frac{nombre\ d'items\ correctement\ prédits\ comme\ appartenant\ à\ C}{nombre\ d'items\ prédits\ comme\ appartenant\ à\ C}$$

$$\mathrm{F1} = \frac{R \times P}{(R+P)}$$

51 https://huggingface.co/CamemBERT-base
52 https://huggingface.co/docs/transformers/main_classes/trainer
53 Optimiseur Adam ; taux d'apprentissage de 10-5 ; taille de batch de 8.

Le rappel mesure ainsi la capacité du modèle à retrouver correctement les éléments d'une classe donnée. Un rappel de 0,3 pour la classe « joie » indique par exemple que sur 10 phrases appartenant à la classe « joie », le modèle n'en retrouve que 3. La précision traduit quant à elle la capacité du modèle à prédire correctement une classe donnée. Une précision de 0,8 indique par exemple que sur 10 phrases prédites comme appartenant à la classe « joie », seules 8 relèvent effectivement de cette classe. La F-mesure donne une vue d'ensemble équilibrée des performances du modèle : élevée, elle indique un bon rappel et une bonne précision ; à l'inverse, si elle est basse, elle indique que le rappel, la précision ou les deux sont faibles. Par exemple, la F-mesure d'un modèle avec un R de 0,3 et une P de 0,8 n'aura une F-mesure que de 0,44.

Les performances des classifieurs multi-étiquettes (tâches B à E) sont également mesurées toutes classes confondues. Le rappel, la précision et la F-mesure sont alors calculés sous la forme de *macro-moyennes* afin d'accorder autant de poids à toutes les classes à prédire dans l'évaluation[54]. Pour un classifieur prédisant n classes, le macro-rappel, la macro-précision et la macro-F-mesure se calculent comme suit :

$$\text{macro-R} = \frac{\sum_{i=1}^{n} rappel_i}{n} \qquad \text{macro-P} = \frac{\sum_{i=1}^{n} précision_i}{n}$$

$$\text{macro-F1} = \frac{\sum_{i=1}^{n} Fmesure_i}{n}$$

4.3) Facteurs de variation des performances

Pour obtenir les classifieurs les plus performants possibles, plusieurs facteurs ont été manipulés : le nombre d'epochs, le type d'entrée fournies au modèle, la taille du corpus d'entraînement et la pondération ou non des classes lors de l'apprentissage.

Nombre d'epochs. Une *epoch* correspond à un cycle complet de traitement de l'intégralité du jeu de données lors de l'apprentissage. Plus le nombre d'epochs est élevé, plus l'apprentissage sera long. Les performances du classifieur pourront ainsi être améliorées mais le risque de *sur-apprentissage*[55] sera également plus élevé. De manière

54 Les moyennes peuvent également être calculées de manière « micro », c'est-à-dire de manière pondérée. Les classes très représentées pèsent alors plus que les classes plus rares.

55 Dans les cas de sur-apprentissage, le modèle effectue de très bonnes prédictions sur le *train* mais se trompe dès qu'il est confronté à des données nouvelles.

empirique, les classifieurs de la tâche A sont entraînés avec 3 epochs. Pour les tâches de difficulté supérieure, l'apprentissage est poussé jusqu'à 9 epochs.

Type d'entrée fourni aux modèles. Deux types d'entrée ont été testés pour l'apprentissage des classifieurs : la phrase isolée et la phrase avec son contexte. Pour la première condition, les modèles ne reçoivent en entrée qu'une seule phrase, celle à analyser. Pour la seconde, les modèles reçoivent en entrée un triplet de phrases composé de la phrase à analyser, de la phrase qui la précède et de la phrase qui la suit. Cette seconde condition a été choisie pour aider les classifieurs sur les tâches les plus difficiles et pour mieux rendre compte du processus d'annotation en contexte effectué par les humains. Fournir une phrase avec son contexte représente cependant un inconvénient majeur : le risque de ne pas savoir pour quelle phrase exactement les classes sont prédites. Pour pallier ce problème, j'utilise un formatage censé guider les modèles. Il repose sur la délimitation et la dénomination des éléments du triplet à l'aide des introducteurs « before: » (pour la phrase qui précède), « current: » (pour la phrase à analyser), « after: » (pour la phrase qui suit) et du token spécial « </s> » interprété comme une fin de phrase par le tokeniseur puis par les modèles transformeurs. Un triplet de phrases se présente donc comme suit :

```
before: phrase précédente</s>current: phrase cible</s>after: phrase suivante</s>
```

Taille du *train*. Les classes retenues pour les expériences de classification représentent des proportions faibles des phrases du *train* (*cf.* tableau 36). Cela peut gêner l'apprentissage des classifieurs, favorisant ainsi les faux-négatifs. Pour tenter de minimiser l'impact de la sous-représentation des classes, le *train* a été réduit en supprimant aléatoirement des phrases non-émotionnelles, de sorte à obtenir autant de phrases émotionnelles que non-émotionnelles : c'est le *train* tronqué (par opp. à train *entier*). Notons que le corpus *train* tronqué n'est malheureusement plus représentatif de la réalité des données textuelles.

Pondération ou non des classes lors de l'apprentissage. En plus d'être peu représentées, les classes retenues pour les tâches de classifications sont réalisées par des proportions très variables dans le *train* (*cf.* tableau 36). Ce déséquilibre entre classes impliquées dans une même tâche peut entraîner un biais d'apprentissage : les classifieurs risquent de ne prédire que les classes les plus fréquentes. Une piste pour limiter ce phénomène est de jouer sur le poids accordé à chaque classe lors de l'entraînement en augmentant celui des classes les plus rares afin d'empêcher le modèle de les ignorer

lors de l'apprentissage. Par défaut, l'API `Trainer` que j'ai utilisée pour entraîner les classifieurs gère seule les poids des classes : c'est l'apprentissage dit « non-pondéré ». Ce `Trainer` peut cependant être modifié en spécifiant le poids de chaque classe. Dans cet apprentissage dit « pondéré », j'ai affecté un nouveau poids à chaque classe, correspondant à l'inverse de la fréquence de la classe dans le *train*[56]. Ainsi, plus une classe est rare, plus elle pèsera dans l'apprentissage. Pour les tâches D et E qui reposent sur des classes très peu représentées (par ex. la classe « jalousie »), les poids corrigés sont parfois beaucoup trop grands et risquent à nouveau de gêner les modèles. Ces poids sont alors arbitrairement plafonnés à 50.

L'ensemble des facteurs manipulés lors des expérimentations et leurs valeurs associées sont résumées par le tableau 37.

Facteur	**Valeurs**
nombre d'*epochs*	3 ; 9
entrée du modèle	phrase isolée ; phrase avec contexte
taille du *train*	entier ; tronqué
type d'apprentissage	non pondéré ; pondéré ; pondéré avec un seuil fixé à 50

Tableau 37: Facteurs manipulés pour les expérimentations et leurs valeurs

4.4) Mise en œuvre des expérimentations

Principe des expérimentations : la notion de configuration. Une *configuration* correspond ici à la combinaison d'un ensemble de valeurs des facteurs de variation identifiés section 4.3 (*cf.* tableau 37). Pour la tâche A, une configuration testée est par exemple un entraînement lancé pour 3 epochs, sur le *train* entier, avec en entrée des phrases isolées et un apprentissage non pondéré. Une autre configuration est un entraînement lancé pour 3 epochs, sur le *train* entier, avec en entrée des phrases *avec leur contexte* et un apprentissage non pondéré.

Déroulé des expérimentations. L'initialisation de la couche de classification sur laquelle repose l'ajustement de CamemBERT a un fort impact sur les performances

56 Il n'y a pas de règle d'or pour calculer le poids à attribuer à une classe : l'important est de ne pas considérer que toutes les classes sont à égalité et de faire en sorte que plus une classe est rare, plus son poids sera grand.

des classifieurs et s'effectue de manière aléatoire. Afin de s'assurer que les résultats de mes expériences sont stables et reproductibles, autrement dit qu'ils reflètent bien la capacité des modèles à répondre à une tâche de classification donnée, *trois entraînements sont lancés pour chaque configuration.* À la fin de chaque entraînement, seul le modèle ayant les meilleures performances sur le *dev* est conservé. Pour une même configuration, trois modèles sont ainsi retenus. La moyenne des mesures de R, de P et de F1 de ces trois modèles est calculée pour rendre compte de la stabilité des résultats de la configuration sur le *dev*. Ces modèles sont enfin évalués sur le *test* et leurs scores moyennés pour obtenir les performances finales de la configuration sur le *test*.

La sélection du meilleur modèle généré par un entraînement s'effectue à partir de la macro-F-mesure du modèle. Si la précision était privilégiée, le « meilleur » modèle pourrait être celui qui ne prédit une classe que quand il est très sûr de lui, quitte à ne faire que peu de prédictions (rappel très bas). Au contraire, si le rappel était privilégié, le « meilleur » modèle pourrait prédire systématiquement la classe majoritaire, générant au passage beaucoup de faux positifs (précision très basse). La F-mesure constitue ainsi un bon compromis entre ces deux mesures.

Quelle que soit leur configuration, les modèles sont tous évalués sur l'intégralité du *dev* et du *test*. Le seul facteur qui influence la forme des ensembles d'évaluation est le type d'entrée fourni au modèle. Prendre en compte le contexte des phrases lors de l'apprentissage implique en effet une modification du format des données d'entrée du modèle (cf. section 4.3), format qui doit être respecté pour l'évaluation. Les modèles entraînés pour classer des phrases isolées sont donc évalués sur le *dev* et le *test* comportant exclusivement des phrases isolées et réciproquement pour ceux entraînés pour classer des phrases présentées avec leur contexte.

Hiérarchisation des tâches par ordre de difficulté : principe de sélection des entraînements. Les tâches de classification sont conçues par ordre de difficulté croissante. Ce principe permet d'alléger le coup des entraînements en sélectionnant les valeurs de facteur testées pour chaque tâche selon le besoin des prédictions visées. Ainsi, quand une valeur donne de mauvais résultats sur une tâche plus « facile », elle est écartée de la suite des expériences. À l'inverse, certaines valeurs émergent uniquement pour les tâches les plus compliquées, comme le fait d'entraîner les modèles pour 9 epochs ou de fixer un seuil à 50 pour la pondération de l'apprentissage. Le tableau 38 résume, pour chaque facteur et pour chaque tâche, les valeurs testées.

Facteur	Tâche A	Tâches B & C	Tâche D	Tâche E
nombre d'*epochs*	3	3 ; 9	3 ; 9	3 ; 9
entrée du modèle	isolée ; contexte	isolée ; contexte	isolée ; contexte	isolée ; contexte
taille du *train*	entier ; tronqué	entier	entier	entier
type d'apprentissage	non pondéré ; pondéré	non pondéré ; pondéré	non pondéré ; pondéré ; pondéré seuil 50	non pondéré ; pondéré seuil 50

Tableau 38: Valeurs de facteurs testées pour chaque tâche de classification

De la même manière, CamemBERT est employé dans son état initial uniquement pour les tâches A et B. Dès la tâche B, il peinait en effet à accrocher l'apprentissage (besoin de nombreux entraînements pour obtenir des prédiction non nulles). Puisque les tâches B à E constituent un élargissement de la tâche de prédiction initiale qu'est le repérage du caractère émotionnel d'une phrase, les classifieurs de ces tâches résultent donc tous de l'ajustement d'un modèle issu de la tâche A. Le type d'apprentissage (non pondéré ou pondéré) utilisé pour entraîner ce modèle issu de A n'a pas d'importance pour la continuation des entraînements. J'ai en revanche jugé préférable de conserver une cohérence sur le type d'entrée fourni au classifieur. Notons qu'entraîner les modèles des tâches B à E en repartant d'un modèle issu de A ne pose pas de problème pour l'évaluation : la composition du *train*, du *dev* et du *test* est identique pour toutes les tâches et toutes les configurations.

4.5) Baselines

Une question au cœur de mes expérimentations est de pouvoir juger de la pertinence d'avoir choisi de développer des classifieurs de type transformeur. Car si ces modèles sont réputés pour leur puissance, ils présentent deux inconvénients : 1) ils sont lourds (espace de stockage) et gourmands en ressources (entraînement coûteux en temps et en énergie) ; 2) ils sont opaques, c'est-à-dire qu'il est difficile de savoir comment ils « raisonnent ». Pour évaluer l'apport de l'outil transformeur sur mes tâches de classification, il était donc nécessaire de comparer les performances des classifieurs issus de l'ajustement de CamemBERT à celles de modèles dits « de *baseline* ». Ces modèles se devaient d'utiliser des techniques plus simples que les transformeurs afin de fixer le meilleur résultat atteignable « facilement » pour les tâches visées, tout en étant plus légers et moins gourmands – voire plus explicables. Trois modèles ont été retenus pour cela. Le premier est un modèle théorique et les deux autres sont des modèles

d'*apprentissage automatique de surface* : un modèle de type *Support-Vector Machine* (SVM) et un modèle reposant sur le principe des arbres de décision, XGBoost.

Première *baseline* : modèle aléatoire. Comparer les résultats de mes classifieurs à ceux d'un modèle aléatoire montre que leurs prédictions ne sont pas dues au hasard, que les tâches fixées sont réalisables. Le modèle aléatoire n'est pas un modèle informatique : ses résultats ne correspondent qu'à une simulation des productions d'un modèle qui prédirait les classes d'après leur proportion au sein du *train*. Le modèle aléatoire « prédit » donc chaque classe avec une probabilité équivalente à sa proportion dans le *train*. Pour le calcul des résultats de cette *baseline*, chaque classe à prédire est traitée séparément. Le principe de calcul est le suivant :

1. pour chaque classe c à prédire, on calcule sa proportion dans le *train* $\text{Prop}(c)_{\text{train}}$ et sa proportion dans le *dev* $\text{Prop}(c)_{\text{dev}}$;
2. à partir des ces proportions, les prédictions d'un modèle aléatoire sont mimées en créant une matrice de confusion pour chaque classe. Chaque case de la matrice est calculée en multipliant la proportion dans le *train* de la classe « prédite » par la proportion représentée par cette classe dans le *dev* :

	Classe prédite *c*	**Classe prédite non-*c***
Classe attendue *c*	$\text{Prop}(c)_{\text{train}} \times \text{Prop}(c)_{\text{dev}}$	$(1 - \text{Prop}(c)_{\text{train}}) \times \text{Prop}(c)_{\text{dev}}$
Classe attendue non-*c*	$\text{Prop}(c)_{\text{train}} \times (1 - \text{Prop}(c)_{\text{dev}})$	$(1 - \text{Prop}(c)_{\text{train}}) \times (1 - \text{Prop}(c)_{\text{dev}})$

3. à partir de la matrice de confusion, les mesures de R, de P et de F1 sont calculées pour chaque classe. Pour les tâches multi-étiquettes, les mesures globales correspondent aux macro-moyennes de ces mesures.

Deuxième *baseline* : modèle SVM. Les modèles d'apprentissage automatique de surface de type *support-vector machine* (SVM), légers et rapides à entraîner, sont employés depuis longtemps en TAL. L'objectif est ici de comparer les résultats de mes classifieurs transformeurs à ceux d'une technique historique du TAL, moins gourmande en ressources. Le modèle LinearSVC de la librairie Python sklearn[57], simple d'utilisation, a été retenu avec le paramétrage par défaut pour l'apprentissage[58]. Pour

57 Documentation disponible ici : https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html#sklearn.svm.LinearSVC.

58 L'initialisation des entraînements a été fixée à 14 pour garantir la reproductibilité des résultats.

les tâches multi-étiquettes, un modèle SVM a été entraîné pour chaque classe. Dix-neufs modèles SVM ont donc été développés au total.

Troisième *baseline* : modèle XGBoost. Les modèles de type eXtreme Gradient Boosting (XGBoost) qui reposent sur le principe des arbres de décision sont rapides à entraîner et peu gourmands en ressources, tout en atteignant les performances de l'état de l'art sur plusieurs tâches de classification (Chen & Guestrin, 2016). L'objectif est à nouveau de comparer les résultats de mes classifieurs avec ceux d'une technique peu coûteuse, cependant plus récente et généralement plus performante que les SVM. L'apprentissage est effectué en conservant tous les paramètres par défaut des modèles XGBoost[59], sauf pour celui indiquant le type de tâche à traiter. La valeur retenue pour ce paramètre est celle permettant de traiter les classifications binaires. Un modèle XGBoost est entraîné pour chaque classe, soit dix-neuf modèles en tout.

Représentation des phrases pour les modèles SVM et XGBoost. Les modèles SVM et XGBoost ne peuvent être appris qu'à partir d'une représentation numérique des phrases. J'ai choisi deux méthodes.

Méthode 1 : sac de tokens. La première consiste à représenter une phrase par un vecteur de type « sac de tokens ». Pour cela, un vocabulaire de référence de k tokens est défini. Chaque phrase du corpus est ensuite représentée par un vecteur de dimension k, où chaque position de rang i correspond au nombre de fois que le $i^{\text{ème}}$ token du vocabulaire de référence apparaît dans la phrase. Afin de rendre plus comparables les performances de mes classifieurs et celles des modèles SVM et XGBoost, le même tokeniseur, celui associé à CamemBERT, est adopté. À partir du vocabulaire de ce tokeniseur, d'une taille de 32 005 tokens (Martin et al., 2020), et de la tokenisation opérée, trois conditions sont retenues ici pour la représentation en sacs de tokens : vecteur de taille 10 000 ; vecteur de taille 15 000 ; et vecteur de taille 20 000. Pour les deux premières, seuls les N tokens les plus fréquents du *train* sont utilisés pour représenter les phrases, où N est fixé arbitrairement. Pour la troisième, ce sont cette fois tous les mots du vocabulaire – ou en l'occurrence du *train* – qui sont utilisés[60] .

***Méthode 2 : plongement de phrase avec* SentenceTransformer.** Les phrases peuvent également être représentées à l'aide d'un plongement de phrase (*embedding*).

59 https://xgboost.readthedocs.io/en/latest/parameter.html

60 Dans le *train*, le nombre de tokens différents n'est en fait que de 18 437, rendant ainsi inutile l'emploi de l'ensemble des 32 005 tokens du vocabulaire pour représenter les phrases.

J'ai pour cela eu recours à l'outil Sentence-Transformers[61], plus précisément la version *sentence-CamemBERT-base*[62], qui fournit des plongements similaires à ceux exploités lors de l'entraînement de mes classifieurs transformeurs (c'est-à-dire ceux générés par la première couche de CamemBERT). Les plongements issus de la transformation avec SentenceTransformer sont de dimension 568.

Déroulé des expérimentations pour les *baselines* SVM et XGBoost. Comme pour le développement des classifieurs transformeurs, le déroulé des expérimentations des *baselines* SVM et XGBoost est pensé pour alléger le coup des entraînements. De manière arbitraire, j'ai commencé par les modèles SVM. Une fois encore, la première tâche traitée est la plus simple : la prédiction du caractère émotionnel. Après avoir testé plusieurs représentation de phrases pour entraîner les modèles SVM sur cette tâche et évalué sur le *dev* les modèles ainsi produits (détail des résultats en Annexe VII), seules deux types de représentation de phrases ont été conservés : sac de tokens de taille 20 000 et plongements de phrases. Les modèles SVM des tâches B à E et les modèles XGBoost des tâches A à E sont donc tous entraînés seulement avec ces deux conditions pour les représentations de phrases. Tous les modèles appris sont évalués sur le *dev* et le *test* par les mesures de R, de P et de F1.

Notons que l'objectif n'est pas d'aboutir aux modèles SVM et XGBoost les plus performants mais uniquement de tester des techniques simples pour répondre aux tâches de classification, afin d'avoir un ordre de grandeur des résultats ainsi obtenus. Les protocoles expérimentaux exposés ici pourraient être affinés pour optimiser au mieux les modèles SVM et XGBoost (variation de plus de paramètres lors de l'apprentissage, usage d'autres modes de représentation des phrases). De plus, seules les entrées de type phrase isolée, plus aisément manipulables, sont exploitées pour les *baselines*.

5) Résultat des expérimentations

Les expériences dont je viens de décrire les principes méthodologiques visent à explorer les possibilités qu'offre le corpus annoté du point de vue de l'analyse automatique des émotions. Pour rappel, cet objectif général est décomposé en cinq tâches de classification de phrase de difficulté croissante. Les sections 5.1 à 5.5 exposent les performances en termes de rappel (R), de précision (P) et de F-mesure (F1) des modèles de

61 https://www.sbert.net/
62 https://huggingface.co/dangvantuan/sentence-CamemBERT-base

baseline et des classifieurs transformeurs pour chaque tâche. La section 5.6 propose une synthèse de ces résultats.

Pour les classifieurs transformeurs, les résultats sont présentés par *configuration*. Autrement dit, les performances affichées reflètent la moyenne des R, des P et des F1 obtenus sur trois entraînements (*cf.* section 4.4). Les performances des tâches de classification multi-étiquettes (tâches B à E) rapportées sont de plus des *macro*-moyennes (*cf.* section 4.1). Au vu du grand nombre de classes à prédire et de modèles développés tout au long des expérimentations, il n'était en effet pas envisageable de détailler les résultats classe par classe, classifieur par classifieur. Une analyse détaillée des résultats du meilleur classifieur de la tâche E est en revanche proposée au chapitre 6.

Tout au long de la présente section, les analyses conduites reposent principalement sur l'observation de la (macro)F-mesure. Sont alors appelés « meilleurs » modèles ou « meilleures » configuration les modèles ou configurations qui obtiennent la meilleure (macro-)F1 sur le *dev*. Les F1 de ces modèles et configurations apparaissent **en gras** dans les tableaux 39 à 48. À noter par ailleurs que seuls les résultats des meilleurs modèles de *baseline* sont *montrés* ici, les détails étant en Annexe VII.

5.1) Résultats de la tâche A

La tâche A repose sur l'identification du *caractère émotionnel* d'une phrase, soit sur la prédiction de la classe « émotionnelle ». Les performances en termes de R, de P et de F1 des meilleurs modèles de *baseline* sur cette tâche sont présentées tableau 39. Le tableau 40 résume les performances des différentes configurations de classifieurs transformeurs testées pour la tâche A. Ces performances correspondant à des moyennes, l'écart-type est indiqué à chaque fois entre parenthèses.

		Dev			***Test***		
Modèle	**Représentation**	**R**	**P**	**F1**	**R**	**P**	**F1**
aléatoire	non concerné	0,202	0,045	0,074	0,202	0,176	0,188
SVM	sac_tokens_20000	0,477	0,668	**0,557**	0,481	0,659	**0,556**
XGBoost	sac_tokens_10000	0,221	0,758	0,343	0,223	0,700	0,338

Tableau 39: Performances des meilleures modèles de baseline pour la tâche A

Sur la tâche A, tâche la plus simple, les performances des *baselines* sont basses, en particulier pour le modèle aléatoire. Le meilleur modèle de *baseline* est de type SVM mais quelle que soit la configuration observée[63], les performances des classifieurs transformeurs sont plus élevées. Cela confirme d'une part que des modèles peuvent apprendre à résoudre la tâche A grâce au corpus annoté et d'autre part que l'usage de modèles transformeur est pertinent pour cela.

train	**Entrée**	**Entraîne-ment**	**Epochs**	***Dev***			***Test***		
				R (σ)	**P (σ)**	**F1 (σ)**	**R (σ)**	**P (σ)**	**F1 (σ)**
entier	isolée	non pond.	3	0,763 (0,027)	0,708 (0,027)	0,734 (0,004)	0,737 (0,034)	0,720 (0,023)	0,728 (0,006)
		pondéré	3	0,782 (0,011)	0,702 (0,016)	0,740 (0,009)	0,756 (0,017)	0,710 (0,012)	0,732 (0,003)
	avec contexte	non pond.	3	0,784 (0,006)	0,727 (0,012)	**0,754** (0,007)	0,767 (0,001)	0,734 (0,002)	**0,750** (0,002)
		pondéré	3	0,791 (0,011)	0,721 (0,017)	**0,754** (0,006)	0,771 (0,022)	0,730 (0,012)	**0,749** (0,005)
tronqué	isolée	non pond.	3	0,896 (0,003)	0,546 (0,002)	0,679 (0,003)	0,876 (0,009)	0,540 (0,008)	0,668 (0,008)
		pondéré	3	0,896 (0,005)	0,532 (0,012)	0,668 (0,008)	0,878 (0,012)	0,530 (0,006)	0,661 (0,006)
	avec contexte	non pond.	3	0,912 (0,013)	0,561 (0,005)	0,695 (0,001)	0,887 (0,003)	0,564 (0,003)	0,689 (0,002)
		pondéré	3	0,891 (0,02)	0,573 (0,011)	0,697 (0,004)	0,876 (0,016)	0,578 (0,01)	0,696 (0,005)

Tableau 40: Performances des configurations des classifieurs transformeurs pour la tâche A

D'après le tableau 40, les configurations apprises sur le *train* entier atteignent de bien meilleures performances, du point de vue de la F1, que celles apprises sur le *train* tronqué. Seule la première condition est donc conservée pour les tâches suivantes. L'entrée phrase avec contexte offre de meilleurs résultats que l'entrée phrase isolée. Ces deux types d'entrée ont cependant chacun des avantages (meilleures performances par opp. à facilité de manipulation) et seront donc conservés pour la suite des expériences. Les résultats ne permettent pas de se prononcer sur l'impact de la (non-)pondération de l'apprentissage. Notons enfin que le R est plus élevé que la P et que les performances sont toujours (légèrement) inférieures sur le *test* par rapport au *dev*.

5.2) Résultats de la tâche B

La tâche B, tâche de classification multi-étiquettes, repose sur l'identification de la notion de *mode d'expression*, soit sur la prédiction de quatre étiquettes : « comportemental », « désigné », « montré » et « suggéré ». Les tableaux 41 et 42 montrent les performances en termes de macro-R, de macro-P et de macro-F1 des meilleurs modèles de *baseline* et des différentes configurations de classifieurs transformeurs testées.

63 Le modèle Flaubert a aussi été testé pour la tâche A mais les résultats obtenus, soit inférieurs soit similaires à ceux obtenus avec CamemBERT, ne sont pas présentés ici.

		Dev			***Test***		
Modèle	**Représentation**	**macro-R**	**macro-P**	**macro-F1**	**macro-R**	**macro-P**	**macro-F1**
aléatoire	non concerné	0,052	0,042	0,046	0,052	0,049	0,050
SVM	sac_tokens_20000	0,317	0,759	**0,408**	0,267	0,721	**0,368**
XGBoost	sac_tokens_20000	0,296	0,702	0,384	0,218	0,730	0,314

Tableau 41: Performances des meilleures modèles de baseline pour la tâche B

Sur cette tâche, située un cran de difficulté au-dessus de la tâche A, les performances des *baselines*, en baisse par rapport à la tâche précédente, sont faibles et ce en particulier pour le modèle aléatoire. Les classifieurs performent toujours mieux que les modèles de *baseline,* y compris que le modèle SVM le plus efficace. La faisabilité de la tâche et l'intérêt des modèles transformeurs sont à nouveau confirmés.

Modèle	***train***	**Entrée**	**Entraîne-ment**	**Epochs**	***Dev***			***Test***		
					macro-R (σ)	**macro-P (σ)**	**macro-F1 (σ)**	**macro-R (σ)**	**macro-P (σ)**	**macro-F1 (σ)**
camembert	entier	isolée	non pond.	3	0,563 (0,052)	0,738 (0,065)	0,596 (0,057)	0,477 (0,066)	0,717 (0,058)	0,535 (0,074)
				9	0,677 (0,04)	0,638 (0,015)	0,654 (0,014)	0,590 (0,044)	0,636 (0,013)	0,611 (0,031)
			pondéré	3	0,413 (0,148)	0,474 (0,172)	0,408 (0,142)	0,568 (0,025)	0,649 (0,014)	0,605 (0,014)
				9	0,678 (0,006)	0,659 (0,014)	0,666 (0,006)	0,584 (0,026)	0,644 (0,006)	0,612 (0,014)
		avec contexte	non pond.	3	0,468 (0,244)	0,562 (0,19)	0,484 (0,243)	0,594 (0,02)	0,656 (0,016)	0,622 (0,013)
				9	0,684 (0,016)	0,670 (0,009)	0,672 (0,01)	0,628 (0,007)	0,667 (0,013)	0,646 (0,007)
			pondéré	3	0,433 (0,139)	0,519 (0,165)	0,448 (0,142)	0,374 (0,135)	0,514 (0,147)	0,416 (0,137)
				9	0,702 (0,016)	0,661 (0,008)	0,678 (0,005)	0,633 (0,003)	0,655 (0,021)	0,642 (0,012)
modèle A	entier	isolée	non pond.	3	0,665 (0,008)	0,677 (0,009)	0,667 (0,007)	0,614 (0,004)	0,656 (0,011)	0,633 (0,007)
				9	0,688 (0,027)	0,636 (0,01)	0,658 (0,016)	0,619 (0,021)	0,650 (0,004)	0,633 (0,01)
			pondéré	3	0,675 (0,003)	0,656 (0,022)	0,662 (0,013)	0,629 (0,011)	0,675 (0,01)	0,650 (0,007)
				9	0,683 (0,019)	0,659 (0,022)	0,668 (0,017)	0,613 (0,022)	0,644 (0)	0,627 (0,012)
		avec contexte	non pond.	3	0,696 (0,006)	0,694 (0,013)	**0,692 (0,009)**	0,623 (0,022)	0,669 (0,009)	**0,644 (0,014)**
				9	0,704 (0,008)	0,670 (0,003)	0,686 (0,005)	0,645 (0,008)	0,659 (0,011)	0,651 (0,01)
			pondéré	3	0,708 (0,009)	0,681 (0,001)	0,691 (0,004)	0,621 (0,014)	0,686 (0,004)	0,651 (0,006)
				9	0,692 (0,02)	0,671 (0,001)	0,68 (0,011)	0,629 (0,009)	0,667 (0,01)	0,647 (0,006)

Tableau 42: Performances des configurations des classifieurs transformeurs pour la tâche B

Sur le tableau 42, les observations effectuées pour la tâche A se maintiennent : la condition phrase avec contexte génère de meilleures performances que la condition phrase isolée ; aucun effet majeur n'émerge de la (non-)pondération de l'apprentissage ; le macro-R est souvent supérieur à la macro-P (pour les classifieurs correspondant à l'affinage d'un modèle issu de la tâche A) ; et les performances baissent sur le *test*. Au vu de la stabilité des résultats permise par l'ajustement d'un modèle issu de A, CamemBERT ne sera plus mobilisé par la suite. Enfin, l'augmentation du nombre d'epochs améliore les performances.

5.3) Résultats de la tâche C

La tâche C (classification multi-étiquettes) repose sur l'identification de la notion de *type d'émotion*, soit sur la prédiction de deux étiquettes: « base » et « complexe ». Les tableaux 43 et 44 montrent les performances respectivement des meilleurs modèles de *baseline* et des différentes configurations de classifieurs transformeurs[64].

		Dev			***Test***		
Modèle	**Représentation**	**macro-R**	**macro-P**	**macro-F1**	**macro-R**	**macro-P**	**macro-F1**
aléatoire	non concerné	0,087	0,073	0,079	0,087	0,081	0,084
SVM	sac_tokens_20000	0,216	0,342	0,265	0,211	0,343	0,261
XGBoost	sac_tokens_20000	0,165	0,856	**0,274**	0,120	0,659	**0,200**

Tableau 43: Performances des meilleures modèles de baseline pour la tâche C

Modèle	***train***	**Entrée**	**Entraîne-ment**	**Epochs**	***Dev***			***Test***		
					macro-R (σ)	**macro-P (σ)**	**macro-F1 (σ)**	**macro-R (σ)**	**macro-P (σ)**	**macro-F1 (σ)**
modèle A	entier	isolée	non pond.	3	0,529 (0)	0,620 (0)	0,558 (0)	0,507 (0)	0,620 (0)	0,548 (0)
				9	0,559 (0)	0,677 (0)	0,601 (0)	0,528 (0)	0,604 (0)	0,562 (0)
			pondéré	3	0,343 (0)	0,358 (0)	0,350 (0)	0,328 (0)	0,352 (0)	0,340 (0)
				9	0,341 (0)	0,345 (0)	0,343 (0)	0,334 (0)	0,341 (0)	0,337 (0)
		avec contexte	non pond.	3	0,587 (0)	0,648 (0)	0,610 (0)	0,553 (0)	0,623 (0)	0,583 (0)
				9	0,565 (0)	0,744 (0)	**0,624** (0)	0,547 (0)	0,644 (0)	**0,588** (0)
			pondéré	3	0,274 (0)	0,213 (0)	0,240 (0)	0,322 (0)	0,218 (0)	0,260 (0)
				9	0,221 (0)	0,268 (0)	0,243 (0)	0,275 (0)	0,267 (0)	0,271 (0)

Tableau 44: Performances des configurations des classifieurs transformeurs pour la tâche C

Sur cette tâche, les performances des *baselines*, toujours faibles, sont encore en baisse par rapport aux tâches précédentes (sauf pour le modèle aléatoire). Le meilleur modèle de *baseline* sur le *dev* est cette fois de type XGBoost mais, à nouveau, les classifieurs transformeurs performent systématiquement mieux, indiquant la faisabilité de la tâche et l'intérêt des transformeurs.

Le tableau 44 montre que seules certaines observations effectuées sur les tâches les plus simples se confirment ici, à savoir de meilleures performances pour la condition phrase avec contexte (sauf pour l'apprentissage pondéré) et une baisse des performances sur le *test*. Pour la première fois, la pondération de l'apprentissage fait chuter les performances et les macro-P sont supérieures aux macro-R. L'augmentation du nombre d'epochs améliore les résultats lorsque l'apprentissage n'est pas pondéré.

64 Pour une raison inconnue, tous les entraînements lancés pour chaque configuration donnaient exactement les mêmes résultats. Les résultats présentés ne correspondent donc pas à des moyennes.

5.4) Résultats tâche D

La tâche D (classification multi-étiquettes) repose sur l'identification de la notion de *catégorie émotionnelle*, soit sur la prédiction de douze étiquettes : « autre », « admiration », « colère », « culpabilité », « dégoût », « embarras », « fierté », « jalousie », « joie », « peur », « surprise » et « tristesse ». Les performances des meilleurs modèles de *baseline* et celles des différentes configurations de classifieurs transformeurs sont présentées respectivement tableaux 45 et 46.

		Dev			***Test***		
Modèle	**Représentation**	**macro-R**	**macro-P**	**macro-F1**	**macro-R**	**macro-P**	**macro-F1**
aléatoire	non concerné	0,020	0,016	0,018	0,020	0,019	0,019
SVM	sac_tokens_20000	0,144	0,494	0,209	0,125	0,487	0,186
XGBoost	sac_tokens_20000	0,179	0,557	**0,263**	0,192	0,565	**0,272**

Tableau 45: Performances des meilleures modèles de baseline pour la tâche D

Modèle	***train***	**Entrée**	**Entraîne-ment**	**Epochs**	***Dev***			***Test***		
					macro-R (σ)	**macro-P (σ)**	**macro-F1 (σ)**	**macro-R (σ)**	**macro-P (σ)**	**macro-F1 (σ)**
modèle A	entier	isolée	non pond.	3	0,291 (0,019)	0,334 (0,007)	0,308 (0,017)	0,275 (0,013)	0,420 (0,072)	0,299 (0,014)
				9	0,388 (0,024)	0,495 (0,006)	0,419 (0,026)	0,379 (0,01)	0,482 (0,014)	0,411 (0,006)
			pondéré	3	0,302 (0,049)	0,396 (0,077)	0,319 (0,041)	0,291 (0,046)	0,389 (0,07)	0,316 (0,042)
				9	0,403 (0,029)	0,461 (0,015)	0,415 (0,023)	0,376 (0,011)	0,440 (0,012)	0,397 (0,013)
			pond. 50	3	0,290 (0,024)	0,351 (0,026)	0,315 (0,008)	0,259 (0,012)	0,333 (0,025)	0,288 (0,006)
				9	0,403 (0,017)	0,460 (0,033)	0,419 (0,022)	0,371 (0,004)	0,432 (0,011)	0,392 (0,006)
		avec contexte	non pond.	3	0,296 (0,023)	0,339 (0,015)	0,310 (0,023)	0,278 (0,019)	0,343 (0,027)	0,296 (0,02)
				9	0,417 (0,009)	0,500 (0,026)	**0,440 (0,01)**	0,393 (0,027)	0,494 (0,021)	**0,422 (0,025)**
			pondéré	3	0,267 (0,022)	0,299 (0,011)	0,278 (0,019)	0,252 (0,019)	0,294 (0,006)	0,266 (0,014)
				9	0,386 (0,007)	0,479 (0,013)	0,405 (0,004)	0,366 (0,004)	0,437 (0,012)	0,390 (0,005)
			pond. 50	3	0,286 (0,005)	0,366 (0,041)	0,315 (0,012)	0,264 (0,012)	0,345 (0,046)	0,295 (0,013)
				9	0,409 (0,014)	0,472 (0,027)	0,424 (0,022)	0,391 (0,004)	0,457 (0,012)	0,413 (0,002)

Tableau 46: Performances des configurations des classifieurs transformeurs pour la tâche D

Sur cette tâche, les performances des *baselines*, toujours faibles, sont encore en baisse par rapport aux tâches précédentes et le meilleur modèle de *baseline* de type XGBoost. Les performances supérieures des classifieurs transformeurs indiquent à nouveau la faisabilité de la tâche et l'intérêt de ces modèles.

Le tableau 46 montre que les observations suivantes, effectuées sur les tâches précédentes se confirment : meilleures performances pour la condition phrase avec contexte (sauf pour l'apprentissage pondéré) ; baisse des performances sur le *test*. Comme pour la tâche C, les macro-P sont supérieures aux macro-R. En revanche, l'effet de la pondération de l'apprentissage est ici mitigé : sans seuil de poids, les performances sont globalement diminuées par la pondération ; avec un seuil à 50, elles restent stables ou

augmentent. Seule cette dernière condition est maintenue pour la tâche suivante. Enfin, l'augmentation du nombre d'epochs améliore ici systématiquement les résultats.

5.5) Résultats tâche E

La tâche E (classification multi-étiquettes), repose sur la prédiction de dix-neufs étiquettes, celles des tâches A à D. Les performances des meilleurs modèles de *baseline* et celles des différentes configurations de classifieurs transformeurs sont présentées respectivement tableaux 47 et 48.

		Dev			***Test***		
Modèle	**Représentation**	**macro-R**	**macro-P**	**macro-F1**	**macro-R**	**macro-P**	**macro-F1**
aléatoire	non concerné	0,043	0,029	0,033	0,043	0,040	0,041
SVM	sac_tokens_20000	0,205	0,543	0,275	0,183	0,530	0,252
XGBoost	sac_tokens_20000	0,204	0,629	**0,294**	0,191	0,617	**0,277**

Tableau 47: Performances des meilleures modèles de baseline pour la tâche E

Modèle	***train***	**Entrée**	**Entraîne-ment**	**Epochs**	***Dev***			***Test***		
					macro-R (σ)	**macro-P (σ)**	**macro-F1 (σ)**	**macro-R (σ)**	**macro-P (σ)**	**macro-F1 (σ)**
modèle A	entier	isolée	non pond.	3	0,339 (0,03)	0,428 (0,037)	0,360 (0,031)	0,312 (0,032)	0,426 (0,027)	0,345 (0,027)
				9	0,486 (0,014)	0,552 (0,013)	0,501 (0,01)	0,443 (0,014)	0,524 (0,013)	0,468 (0,013)
			pond. 50	3	0,328 (0,047)	0,408 (0,075)	0,343 (0,046)	0,298 (0,048)	0,407 (0,084)	0,326 (0,053)
				9	0,475 (0,027)	0,562 (0,013)	0,497 (0,018)	0,433 (0,025)	0,537 (0,01)	0,467 (0,016)
		avec contexte	non pond.	3	0,323 (0,027)	0,400 (0,026)	0,342 (0,026)	0,291 (0,027)	0,411 (0,03)	0,323 (0,025)
				9	0,493 (0,003)	0,566 (0,004)	0,512 (0,002)	0,448 (0,014)	0,530 (0,003)	0,471 (0,015)
			pond. 50	3	0,372 (0,029)	0,443 (0,028)	0,391 (0,035)	0,340 (0,035)	0,470 (0,03)	0,374 (0,036)
				9	0,504 (0,009)	0,576 (0,006)	**0,524 (0,005)**	0,466 (0,018)	0,547 (0,006)	**0,496 (0,013)**

Tableau 48: Performances des configurations des classifieurs transformeurs pour la tâche E

Sur cette tâche la plus difficile, les performances des *baselines*, toujours faibles, remontent légèrement par rapport aux tâches C et D, sans atteindre le niveau des tâches B et A. Si le meilleur modèle de *baseline* est à nouveau de type XGBoost, les performances plus élevées des classifieurs transformeurs indiquent une fois encore la faisabilité de la tâche et l'intérêt de l'outil choisi pour développer mes classifieurs.

Le tableau 48 montre que les observations suivantes, effectuées sur les tâches antérieures se confirment : meilleures performances globales pour la condition phrase avec contexte ; baisse des performances sur le *test*. Comme pour les tâches C et D, les macro-P sont supérieures aux macro-R. Comme pour la tâche D, l'augmentation du nombre d'epochs améliore systématiquement les résultats. L'effet de la pondération de l'apprentissage reste ambivalent : diminution des performances pour les phrases isolées mais amélioration pour les phrases avec contexte.

L'évolution des performances sur les tâches A à D présentait une diminution cohérente avec le niveau de difficulté croissant des tâches. La meilleure configuration de E a en revanche une macro-F1 supérieure à celle constatée pour la tâche D. Cette amélioration des performances de E par rapport à D s'explique certainement par le fait que la tâche E mobilise les classes plus faciles à reconnaître des tâches A, B et C.

5.6) Synthèse

Baselines. Sur le *dev* ou sur le *test* et quelle que soit la tâche considérée, les classifieurs transformeurs sont toujours bien plus performants que le modèle aléatoire. Cela confirme la faisabilité des cinq tâches de classification à partir de mes données annotées. Les écarts de performance avec les modèles SVM et XGBoost soulignent quant à eux la pertinence des transformeurs pour répondre efficacement aux tâches visées.

Configurations des modèles transformeurs. La sélection des meilleurs modèles et configurations s'effectue toujours sur les valeurs de (macro-)F1. Il est à noter cependant que si les valeurs de R sont toujours supérieures à celle de P pour les tâches les plus simples, c'est le contraire pour les tâches C à E.

***Taille du* train.** Les classifieurs de la tâche A appris sur le *train* entier ont une meilleure F1 que ceux appris sur le *train* tronqué, ainsi écarté pour les autres tâches.

Type d'entrée. Les classifieurs appris à partir de phrases avec contexte performent mieux que ceux appris sur des phrases isolées. Les deux conditions sont cependant conservées, au vu de leurs avantages respectifs (performances par opp. à maniabilité).

Modèle initial. Les classifieurs de la tâche A résultent de l'ajustement de CamemBERT. Pour la tâche B cependant, CamemBERT peine à accrocher l'apprentissage et donne des résultats instables. Les classifieurs des tâches C à D proviennent donc uniquement de l'ajustement d'un modèle issu de la tâche A, cette technique ayant prouvé son efficacité sur la tâche B.

Pondération ou non lors de l'apprentissage. Aucun impact clair ne semble se dégager de la (non-)pondération de l'apprentissage sur les performances des classifieurs, exception faite de la tâche C où la pondération fait chuter les résultats.

Nombre d'epochs. L'augmentation du nombre d'epochs n'occasionne une amélioration prononcée des résultats que pour les tâches D et E. Il semblerait que plus la tâche est complexe, plus il est intéressant d'entraîner longtemps les classifieurs.

Performances sur le *test*. Les tendances observées sur le *dev* (comparaison aux *baselines*, impact du type d'entrée, *etc.*) se confirment sur le *test*. De plus, la relative stabilité des performances sur le *test* démontre la robustesse des classifieurs transformeurs. Si elles restent du même ordre de grandeur, les performances sont cependant globalement moindre sur le *test*. Toutes tâches confondues, j'ai plus précisément constaté une baisse moyenne de F1 de 0,013 fondée principalement sur une diminution du R (baisse moyenne de 0,0027 pour le R contre 0,0003 pour la P). Ainsi, si les classifieurs sont toujours aussi pertinents dans leurs analyses, ils ne sont pas en mesure de retrouver certains cas du *test*, *a priori* nouveaux par rapport au *train* et au *dev*. Les raisons de ces différences de performances sont examinées chapitre 6.

Conclusion

Dans le présent chapitre, cinq tâches de classification de phrase de difficulté croissante ont été imaginées pour explorer progressivement les possibilités d'utilisation de mon schéma d'annotation du point de vue de l'analyse automatique des émotions. Dans leur ensemble, ces tâches visaient la prédiction de dix-neufs classes rendant compte de quatre notions clé de la description des émotions : le caractère émotionnel ou non d'une phrase, le mode d'expression sollicité, le type d'émotion et la catégorie émotionnelle dénotés.

Pour répondre à ces tâches, des classifieurs transformeurs ont été entraînés sur mon corpus annoté. L'observation tâche par tâche des performances de ces classifieurs a confirmé l'ordre de difficulté des tâches (identification du caractère émotionnel, puis du mode d'expression, puis du type d'émotion et enfin de la catégorie émotionnelle). De plus, leurs performances sont systématiquement supérieures à celles des *baselines* des expérimentations (modèle aléatoire, modèles d'apprentissage de surface de type SVM ou XGBoost), démontrant ainsi la faisabilité des tâches et la pertinence de l'usage des modèles transformeurs.

Une analyse détaillée du meilleur classifieur de la tâche E, tâche la plus difficile définie ici (prédiction de l'ensemble des dix-neufs classes) est opérée au chapitre 6.

Chapitre 6 Analyse détaillée des résultats du modèle E

Pour sonder les perspectives d'analyse automatique offertes par mon schéma d'annotation des émotions, cinq tâches de classification de difficulté croissante ont été définies chapitre 5. Elles ont été mises en œuvre grâce à des techniques d'apprentissage profond (modèle de type transformeur) à partir du corpus annoté présenté chapitre 4, préalablement transformé en corpus d'apprentissage (transposition des informations d'annotation à l'échelle de la phrase, partitionnement en *train*, en *dev* et en *test*). La plus difficile d'entre elles, la tâche E, vise à prédire, pour une phrase donnée, un ensemble de dix-neuf classes correspondant aux notions de *caractère émotionnel*, de *mode d'expression*, de *type d'émotion* et de *catégorie émotionnelle*.

Le présent chapitre est consacré à l'exploration qualitative des résultats du classifieur le plus performant de la tâche E, ci-après nommé « modèle E », afin d'avoir une vue d'ensemble des prédictions effectuées pour les dix-neuf classes. Ce modèle, sélectionné d'après sa macro-F-mesure sur le *dev* (moyenne des F-mesures obtenues pour les dix-neufs classes), a été entraîné à partir de phrases avec leur contexte[65].

Les performances chiffrées du modèle E par classe sont brièvement exposées, avant de sonder les raisons de la baisse d'efficacité des classifieurs sur le *test* par rapport au *dev* (*cf.* chapitre 5). Pour mieux cerner les résultats du modèle E, des matrices de confusion sont examinées puis l'impact des interactions entre deux notions, le *mode d'expression* et la *catégorie émotionnelle*, est observé. Les prédictions du modèle E sont ensuite mises en regard des annotations humaines de référence sur quelques phrases. Un retour chiffré sur les prédictions du modèle et les annotations humaine de référence est également proposé *via* une expérience de validation perceptuelle menée auprès d'experts. Enfin, les performances du classifieur sont comparées à celles du processeur ÉMOTIONS de la chaîne TextToKids, introduit chapitre 1. Les sections 2 à 5 présentent l'ensemble de ces analyses, effectuées sur le *test*.

À titre indicatif, le tableau 49, déjà présent au chapitre 5, montre la répartition des dix-neufs classes prédites par le modèle E, en termes de proportion de phrases, pour chacun des ensembles du corpus d'apprentissage.

65 La *configuration* exacte du modèle E est : ajustement du modèle issu de A, *train* entier, phrases avec contexte, apprentissage pondéré à seuil 50, 9 epochs.

Classe	**Proportion dans le**		
	train	***dev***	***test***
émotionnelle	20,2 %	15,8 %	17,6 %
comportemental	4,6 %	3,6 %	4,3 %
désigné	5,3 %	5,2 %	5,7 %
montré	3,6 %	2,3 %	3,5 %
suggéré	7,1 %	5,8 %	6,3 %
base	15,4 %	12,6 %	13,9 %
complexe	2,0 %	2,1 %	2,3 %
admiration	0,6 %	1,1 %	1,0 %
autre	5,0 %	3,2 %	3,7 %
colère	4,6 %	3,2 %	3,4 %
culpabilité	0,1 %	0,0 %	0,1 %
dégoût	0,2 %	0,3 %	0,2 %
embarras	0,6 %	0,6 %	0,6 %
fierté	0,7 %	0,4 %	0,9 %
jalousie	0,0 %	0,0 %	0,0 %
joie	3,2 %	2,3 %	3,6 %
peur	3,8 %	3,3 %	3,8 %
surprise	3,0 %	3,1 %	2,5 %
tristesse	2,5 %	2,0 %	2,5 %

Tableau 49: Répartition des classes au sein du train, du dev et du test

1) Performances chiffrées par classes pour le modèle E

Pour rappel, les performances des classifieurs sont évaluées à l'aide des mesures de rappel – noté R –, de précision – notée P – et de F-mesure – notée F1. Le rappel rend compte de la capacité d'un modèle à retrouver correctement les éléments d'une classe donnée tandis que la précision traduit sa capacité à prédire correctement une classe donnée. La F-mesure, moyenne harmonique du rappel et de la précision, donne une vue d'ensemble équilibrée des performances du modèle.

Sur le tableau 50 qui présente les performances (R, P, F1) du modèle E pour chacune des dix-neuf classes, il apparaît que trois d'entre elles ne sont jamais prédites : les classes « dégoût », « culpabilité » et « jalousie », les plus rares du *train* (*cf.* tableau 53). Trois modèles dits « de *baseline* », plus simples que les modèles transformeurs, ont été introduits chapitre 5 pour positionner les résultats des classifieurs

transformeurs. À titre de comparaison, ces trois modèles – dont le détail des performances est disponible en Annexe VII – affichent tous une F1 égale à zéro pour la classe « jalousie ». En revanche, certains trouvent parfois les classes « dégoût » (F1 de 0,002 sur le *dev* et le *test* avec le modèle aléatoire ; F1 de 0,588 sur le *test* avec un modèle XGBoost) et « culpabilité » (F1 de 0,001 sur le *test* avec le modèle aléatoire). Si certains modèles de *baseline* prédisent mieux ou aussi bien que le modèle E les classes les plus rares du *train*, leurs performances pour toutes les autres classes sont en revanche toujours inférieures.

Les performances du modèle E peuvent être groupées pour rendre compte de la capacité du modèle à répondre aux tâches de classification A à D définies chapitre 5. Dans la suite de la section, seule les F1 sur le *test*, **en gras** tableau 50, sont mentionnées.

	Dev			***Test***		
Classe	**R**	**P**	**F1**	**R**	**P**	**F1**
émotionnelle	0,785	0,748	0,766	0,765	0,745	**0,755**
comportemental	0,778	0,706	0,740	0,597	0,664	**0,628**
désigné	0,793	0,833	0,813	0,820	0,810	**0,815**
montré	0,794	0,694	0,741	0,672	0,732	**0,701**
suggéré	0,406	0,455	0,429	0,412	0,468	**0,438**
base	0,761	0,751	0,756	0,702	0,730	**0,716**
complexe	0,386	0,815	0,524	0,425	0,621	**0,505**
admiration	0,194	0,750	0,308	0,281	0,500	**0,360**
autre	0,764	0,660	0,708	0,745	0,594	**0,661**
colère	0,844	0,738	0,788	0,665	0,679	**0,672**
culpabilité	0	0	0	0	0	**0**
dégoût	0	0	0	0	0	**0**
embarras	0,529	0,818	0,643	0,364	0,632	**0,462**
fierté	0,250	0,333	0,286	0,333	0,571	**0,421**
jalousie	0	0	0	0	0	**0**
joie	0,540	0,654	0,591	0,540	0,722	**0,618**
peur	0,670	0,670	0,670	0,712	0,668	**0,689**
surprise	0,721	0,775	0,747	0,704	0,735	**0,719**
tristesse	0,509	0,609	0,554	0,420	0,504	**0,458**

Tableau 50: Performances par classe du modèle E

Tâche A : prédiction du *caractère émotionnel*. Avec une F1 de 0,755 pour la classe « émotionnelle », le modèle E répond de manière très satisfaisante à la tâche A.

Tâche B : prédiction du *mode d'expression*. La classe « désigné » est la mieux reconnue, toutes classes confondues, par le modèle E. Les classes « montré » et « comportemental » sont globalement bien prédites. En revanche, la classe « suggé-

ré » est correctement analysée par le modèle dans moins de la moitié des cas. Ces résultats concordent avec les spécificités des *modes d'expression*. Des quatre modes, le désigné est en effet le mieux circonscrit car son expression repose uniquement sur le lexique émotionnel. Le montré, bien que recouvrant des phénomènes très hétérogènes, s'actualise concrètement dans le corpus avec un petit nombre de marqueurs linguistiques très typiques (par ex. points de suspension ; *cf.* chapitre 4) *a priori* très facilement identifiables par le modèle E. Le mode comportemental, malgré des frontières plus souples que le désigné, repose sur des marqueurs lexicaux relativement récurrents dans le corpus annoté (par ex. manifester, crier, *etc.*). Le mode suggéré, le plus difficile à cerner, est le mode le plus dur à annoter pour les humains. Il n'est donc pas étonnant qu'il soit également le moins bien prédit.

Tâche C : prédiction du *type d'émotion*. Si la classe de *type d'émotion* « base » est le plus souvent bien analysée par le modèle E, ce n'est pas le cas de la classe « complexe » dont il retrouve à peine plus de la moitié des occurrences. Cet écart s'explique sûrement par la fréquence plus élevée des émotions de base dans le corpus annoté (*cf.* tableau 49).

Tâche D : prédiction de la *catégorie émotionnelle*. Les classes de catégories émotionnelles les mieux reconnues sont « surprise », « peur » et « colère ». Les autres catégories émotionnelles, peu fréquentes dans le corpus, sont bien analysées dans moins de la moitié des cas (F1 $<$ 0,5). Il n'est pas étonnant que « peur » et « colère », les deux catégories émotionnelles les plus fréquentes du *train* (*cf.* tableau 49), soient bien prédites. Le critère de la fréquence ne suffit cependant pas à expliquer pourquoi la « surprise » est mieux reconnue que la « joie », pourtant plus représentée. Le facteur qui semble pouvoir justifier cette différence est le *mode* employé pour exprimer les émotions. Les interactions entre mode d'expression et catégorie émotionnelles du point de vue des prédictions sont explorées section 4. Notons par ailleurs que l'écart de performance entre le *dev* et le *test* à nouveau constaté, notamment pour la classe « colère ».

Positionnement par rapport à d'autres travaux de TAL sur l'analyse des émotions. La diversité des démarches (schémas d'annotation, corpus, échelles de prédiction des émotions, systèmes d'analyse automatique, *etc.*) employées en TAL pour la détection des émotions a été présentée en détails en section 3 du chapitre 2. Cette diversité et la spécificité de mon approche – en particulier la prise en compte explicite

de quatre modes d'expression des émotions là où la plupart des travaux n'observent que les informations émotionnelles véhiculées par le lexique émotionnel – rendent difficile la comparaison des performances du modèle E avec celles d'autres classifieurs en émotions. À titre d'exemple cependant, quelques résultats de modèles mentionnés en section 3.3. du chapitre 2 sont rappelés ci-dessous :

- (Öhman et al., 2020) propose des classifieurs transformeurs pour sept langues différentes (mais pas le français), pour prédire à l'échelle de la phrase huit classes de catégorie émotionnelle, dont seulement six sont comparables à celles du modèle E (« colère », « dégoût », « joie », « peur », « surprise » et « tristesse »). Le détail des performances par classe n'est malheureusement pas fourni. Ces classifieurs ont une macro-F1 comprise entre 0,46 et 0,52. Le modèle E a quant a lui une macro-F1 générale de 0,51 ; de 0,42 pour les douze classes de catégorie émotionnelle ; et de 0,52 pour les cinq classes comparables au classifieur de (Öhman et al., 2020) groupées ensemble ;
- pour le français, (Fraisse & Paroubek, 2015) propose un classifieur de type SVM pour prédire des catégories émotionnelles à l'échelle du paragraphe. Les F1 obtenues par ce modèle pour les classes similaires à celles que j'étudie sont : « colère » 0,50 ; « peur » 0,16 ; « tristesse » 0,37. Les performances du modèle E sont bien meilleures pour ces trois classes (*cf.* tableau 50) ;
- pour l'anglais, (Kim & Klinger, 2018) proposent deux classifieurs (un modèle symbolique et un modèle de type réseau de neurones plus simple que les transformeurs) pour prédire huit classes de catégories émotionnelles à l'échelle d'un triplet de phrases, dont seulement six (les mêmes que listées pour (Öhman et al., 2020)) sont comparables aux miennes. Le détail des résultats n'est à nouveau pas communiqué. Les deux classifieurs atteignent une macro-F1 de 0,31. Notons que l'annotation servant de fondement à l'apprentissage de ces classifieurs est l'une des plus proches de celle que j'ai mise en œuvre dans ma thèse (annotation manuelle multi-étiquettes à l'échelle du marqueur linguistique).

Si la comparaison du modèle E à d'autres classifieurs en émotions ne peut être que partielle, il semble donc atteindre des performances similaires voire largement supérieures à celles de ces autres classifieurs.

2) Différences de performances entre le *dev* et le *test*

Une observation récurrente tout au long de la présentation des performances chiffrées (par tâche ou par classe pour le modèle E, *cf.* chapitre 5) est la baisse globale des résultats sur le *test*. Si elle s'explique *a priori* par des spécificités du *test* par rapport au *dev* (et au *train*), il paraissait nécessaire de mieux cerner la nature de ces spécificités. Deux pistes ont été explorées : la composition du *test* en termes de sources de textes, puis la distribution des différentes classes au sein du *test*. Les analyses menées ne sont présentées ici que dans leurs grandes lignes. Le détail est disponible Annexe VIII.

2.1) Piste 1 : composition du *test*.

La piste la plus évidente expliquant la différence de performances entre le *dev* et le *test* est le contenu du *test*. Au sein de cet ensemble, deux sous-corpus peuvent de fait être distingués :

- le *test vus*, composé de textes issus de sources déjà présentes dans le *train* et dans le *dev* ;
- le *test non vus*, composé des textes des corpus Covid, dont les sources et la thématique sont majoritairement absentes du *train* et du *dev* (*cf.* chapitre 5). Ce sous-corpus présente donc un niveau de nouveauté beaucoup plus élevé que le *test vus*. À noter que trois classes sont totalement absentes du *test non vus* : « culpabilité », « dégoût » et « jalousie ».

Les performances du modèle E ont été calculées par classe mais également par tâche (voir le détail des performances en Annexe VIII) sur chacun de ces sous-corpus et sur le *dev* afin de pouvoir les comparer entre elles. Il apparaît ainsi qu'en moyenne, les F1 sur le *test vus* sont supérieures à celles sur le *test non vus*, pour les tâches (+0,198) et les classes (+0,222). L'observation des F1 montre également, toutes tâches confondues, une plus grande proximité des performances entre le *dev* et le *test vus* (+0,006 en moyenne pour le *dev*) qu'entre le *dev* et le *test non vus* (+0,204 en moyenne pour le *dev*). Le constat est le même pour les classes.

L'observation de la variation des performances du modèle E sur le *dev*, le *test vus* et le *test non vus* met en lumière l'impact des données d'évaluation sur la qualité des résultats, au niveau de la tâche comme de la classe. Le modèle E apparaît plus performant sur celles proches (sources, thématiques) du *train*. Le fort niveau de nouveauté

du *test non vus* semble ainsi expliquer une partie de l'écart de performance entre le *dev* et le *test*. Cela suggère un léger phénomène de sur-apprentissage.

2.2) Piste 2 : distance entre les distributions de classes selon les couples (*c,m*).

Une seconde piste pour expliquer la spécificité du *test* repose sur l'observation des distributions des classes sur le *train,* le *dev,* le *test vus* et le *test non vus.* L'hypothèse posée ici est que les prédictions des classifieurs appris sur le *train* sont influencées par la distribution des classes au sein du *train*. Autrement dit, plus un corpus aura une distribution proche de celle du *train*, plus il occasionnera des performances élevées.

Les distributions comparées ici sont celles obtenues à partir de la répartition des phrases selon les couples (*c,m*) qu'elles contiennent, où *c* est une catégorie émotionnelle dénotée par la phrase et *m* le mode qui l'exprime. Une phrase comportant une unité qui désigne la JOIE contiendra donc le couple (joie, désigné). Pour pouvoir être comparées, les distributions sont normalisées en rapportant les nombres d'occurrences au nombre de phrases de l'ensemble. Deux types de distributions sont observés : les distributions dites « attendues », c'est-à-dire celles des corpus *train, dev, test vus* et *test non vus* ; et les distributions dites « prédites », obtenues à partir des prédictions du modèle E. Une distribution fictive est aussi calculée : une distribution uniforme où chaque couple (*c,m*) aurait exactement une occurrence. Les comparaisons s'appuient sur le calcul de la distance entre deux distributions d'après la distance de Bhattacharyya[66] (notée B ci-après). La distance entre la distribution du *train* et la distribution uniforme, d'une valeur de 1,82, sert de point de référence pour évaluer l'éloignement de deux distributions. Le détail des distances calculées est en Annexe VIII.

Des conclusions mitigées se dégagent de l'observation des distances entre les distributions des phrases selon les couples (*c,m*) des différents ensembles. Le *test non vus* se distingue nettement des autres corpus : de loin le plus distant par rapport à la distribution du *train* (B 2,41), c'est également celui pour lequel l'écart entre la distribution attendue et la distribution prédite est le plus grand (B 2,66). Cette spécificité du corpus *test non vus* qui apparaît ici est certainement la cause de la chute de performance qu'il subit. L'analyse des distributions n'apporte cependant pas de réponse quant à la légère baisse de performance sur le *test vus* par rapport au *dev*. Ce dernier est en effet plus distant de la distribution du *train* (B 1,59 entre le *dev* et le *train* ; B 1,36 entre

66 https://fr.wikipedia.org/wiki/Distance_de_Bhattacharyya

le *test vus* et le *train*) et il y a un plus grand écart entre sa distribution attendue et sa distribution prédite (B 1,73) que pour le *test vus* (B 1,38).

Si les analyses menées confirment la spécificité du corpus *test non vus*, certainement en grande partie responsable de la baisse de performance des modèles sur le *test*, les écarts de F1 entre *dev* et *test vus* tendent à souligner un léger phénomène de sur-apprentissage.

3) Matrices de confusion

En proposant une représentation visuelle des croisements entre les classes *de référence* et les classes *prédites*, les matrices de confusion sont un outil précieux pour comprendre des performances des classifieurs. Pour l'analyse conduite dans cette section, les prédictions du modèle E ont été groupées de sorte à obtenir une matrice de confusion par tâche. Les tableaux 52 à 54 présentent les matrices ainsi construites sur le *test*[67], où les chiffres correspondent à des nombres de phrases. Sur ces matrices, les noms de classe sont abrégés. La classe factice « aucune » (abrégé « auc. ») a par ailleurs été ajoutée dans les matrices des tâches B à E pour rendre compte des cas où les classes apparaissent seules dans une phrase.

Croisements « naturels ». Le modèle E répond à une tâche de classification multi-étiquettes, où plusieurs classes peuvent donc être associées à une même phrase. Afin de mieux interpréter les matrices de confusion de cette section, il est nécessaire d'avoir connaissance des classes qui se croisent « naturellement » dans le *test*. Dans ce corpus, les phrases comportent généralement un seul mode d'expression. Les classes de mode les plus souvent associées dans une phrase sont « désignée » et « suggérée », « montrée » et « suggérée », puis « comportementale » et « suggérée ». Pour les types d'émotion, la classe « base » apparaît principalement seule, tandis que la classe « complexe » est majoritairement associée à la classe « base ». Enfin, mises à part «culpabilité », « fierté » et « jalousie », les classes de catégorie émotionnelle sont le plus souvent seules dans une phrase. Les classes les plus associées sont : « admiration » et « joie », « autre » et « joie », « fierté » et « joie », « joie » et « surprise », « peur » et « tristesse ». Ces croisements sont illustrés en Annexe IX.

67 Des matrices de confusion par tâche pour le *test vus* et le *test non vus* sont disponibles en Annexe IX.

Références \ *Prédictions*	émo.	non-émo.
émo.	753	231
non-émo.	258	4338

Tableau 51: Matrice de confusion - tâche A

Références \ *Prédictions*	comp.	dés.	mont.	sugg.	auc.
comp.	142	45	19	63	48
dés.	30	260	33	78	30
mont.	15	25	131	36	53
sugg.	42	79	44	145	132
auc.	60	39	38	88	4536

Tableau 52: Matrice de confusion - tâche B

Tâche A. Le modèle E prédit majoritairement la classe « émotionnelle » quand elle est attendue (*cf.* tableau 51). Les nombres de faux-positifs et de faux-négatifs étant équivalents, le modèle ne prédit pas de manière privilégiée l'absence ou la présence de cette classe.

Tâche B. Les classes de mode sont principalement prédites là où elles sont attendues (*cf.* tableau 52). En observant la ligne de référence « aucune » (phrases sans mode d'expression), il apparaît que les classes « comportemental » et « suggéré » sont les plus prédites à tort. Ce constat est cohérent avec les observations de la section 1, où les performances chiffrées étaient mises en regard des caractéristiques linguistiques des modes. Par ailleurs, la classe « suggéré » est à la fois la plus prédite pour des phrases où d'autres modes sont attendus et la référence pour laquelle d'autres modes sont souvent prédits. Sur la matrice de confusion 52, elle est associée aux classes « désigné » et « comportemental » en particulier. Ces remarques concordent avec le fait que, dans le *test*, le mode suggéré est celui le plus souvent accompagné d'un autre mode. Cependant, la faible performance sur la classe « suggéré » (*cf.* tableau 50) ainsi que l'entremêlement des classes « comportemental », « désigné » et « suggéré » illustré tableau 52 soulignent la nécessité de préciser les limites de ces trois classes les unes par rapport aux autres. Une analyse de phrases extraites du corpus et l'évaluation perceptuelle menée auprès d'experts sont nécessaires pour identifier les erreurs relevant de la référence et celles relevant des prédictions (voir sections 5 et 6).

Références \ *Prédictions*	base	comp.	auc.
base	547	48	205
comp.	35	54	50
auc.	195	17	4536

Tableau 53: Matrice de confusion - tâche C

Tâche C. La matrice de confusion de l'analyse du type d'émotion (tableau 53) confirme les performances chiffrées exposées section 1 : le modèle E peine à trouver la classe « complexe ». Pour cette classe, de fait peu prédite par le modèle, il y a presque autant de vrais-positifs que de faux-négatifs. Si la classe « base » est majoritairement prédite là où elle était attendue, le nombre de faux-positifs reste élevé. La matrice 53 illustre donc la difficulté de la tâche C dans son ensemble. Plusieurs facteurs semblent en être à la source. En premier lieu, les catégories d'émotion complexes, déjà moins nombreuses que celles de base, sont peu représentées (*cf.* tableau 49). Elles sont ensuite le plus souvent associées à des catégories de base (*cf. supra*). Les classes de type sont enfin sémantiquement très hétérogènes puisqu'elles regroupent des catégories émotionnelles parfois très différentes (par ex. joie et dégoût).

Références \ ***Prédictions***	**adm.**	**aut.**	**col.**	**culp.**	**deg.**	**emb.**	**fier.**	**jal.**	**joie**	**peur**	**surp.**	**tris.**	**auc.**
adm.	16	5	0	0	0	0	7	0	7	0	10	1	23
aut.	3	155	9	0	0	2	3	0	18	12	3	11	35
col.	1	12	125	0	0	5	0	0	3	10	6	6	44
culp.	0	0	0	0	0	1	0	0	0	0	0	1	2
deg.	0	4	0	0	0	2	0	0	1	3	0	6	1
emb.	0	6	2	0	0	12	0	0	0	2	4	6	9
fier.	4	4	0	0	0	1	16	0	8	0	5	0	20
jal.	0	0	1	0	0	0	0	0	0	0	0	0	0
joie	9	26	2	0	0	1	15	0	109	4	20	6	52
peur	1	17	11	0	0	3	0	0	3	151	8	20	41
surp.	6	6	7	0	0	1	5	0	9	8	100	2	29
tris.	4	15	9	0	0	5	0	0	5	26	4	58	41
auc.	9	74	45	0	0	1	5	0	29	60	21	35	4340

Tableau 54: Matrice de confusion - tâche D

Tâche D. Le modèle E prédit surtout les classes de la tâche D là où elles sont attendues. Si les associations entre catégories émotionnelles au sein du *test* (*cf. supra*) apparaissent dans les prédictions illustrées par la matrice 54, aucune confusion entre classes n'émerge. Quand aucune catégorie émotionnelle n'est attendue, les classes « autre », « colère », « peur » et « tristesse » sont les plus prédites. Ces faux-positifs ne détonent pas avec la classe « autre », très hétérogène, mais surprennent pour les autres classes, normalement mieux circonscrites. Pour trancher entre erreur du modèle et annotations humaines perfectibles, des analyses complémentaires sont nécessaires.

Les matrices de confusion analysées ici précisent les profils des prédictions du modèle E et mettent en lumière des difficultés d'analyse (tâche C, mode suggéré, catégo-

rie AUTRE). Ces difficultés peuvent traduire des limites du modèle E dues aux données annotées ou à une nécessité de perfectionner l'apprentissage. Elles peuvent également révéler des imperfections dans les annotations humaines de référence (cohérence des annotations, définition de certaines classes, *etc.*). Les sections 5 et 6 visent à apporter des éléments de réponses pour tenter de trancher entre ces deux explications.

4) Interactions entre catégorie émotionnelle et mode

Les évaluations chiffrées des performances du modèle E présentées section 1, suggéraient des interactions entre la catégorie émotionnelle et le mode d'expression dans les prédictions. J'explore ici l'impact du mode d'expression sur la prédiction de la catégorie émotionnelle et celui de la catégorie émotionnelle sur la prédiction du mode.

4.1) Observation 1 : Impact du mode sur la prédiction de la catégorie émotionnelle.

L'objectif est ici de savoir si les prédictions de catégories émotionnelles sont influencées par le mode exprimant l'émotion. Par exemple, le modèle E retrouve-t-il mieux la JOIE quand elle est désignée que quand elle est suggérée ?

Pour répondre à cette question, les prédictions du modèle E sont évaluées mode d'expression par mode d'expression. La démarche ayant permis de filtrer les informations nécessaires à ces observations est détaillée en Annexe X.

	comp.	**dés.**	**mont.**	**sugg.**
admiration	0,108	0,449	0,042	0,164
colère	0,597	0,292	0,078	0,348
culpabilité	0	0		0
dégoût		0	0	0
embarras	0,160	0,571	0	0,313
fierté		0,450	0,061	0,246
jalousie	0	0		
joie	0,262	0,531	0,256	0,275
peur	0,295	0,626	0,026	0,341
surprise	0,014	0,409	0,578	0,216
tristesse	0,187	0,438	0,042	0,281

Tableau 55: F-mesure des classes de catégories émotionnelles selon le mode d'expression, modèle E sur le test

	Prop. de phrases où la cat. émo est			
	comp.	**désignée**	**montrée**	**suggérée**
admiration	14,3 %	37,9 %	15,0 %	32,9 %
colère	**47,6 %**	12,0 %	10,9 %	29,4 %
culpabilité	6,7 %	46,7 %	20,0 %	26,7 %
dégoût	11,9 %	16,7 %	16,7 %	54,8 %
embarras	15,5 %	36,4 %	16,3 %	31,8 %
fierté	8,4 %	12,0 %	17,4 %	62,3 %
jalousie	0,0 %	0,0 %	33,3 %	66,7 %
joie	18,5 %	24,8 %	15,7 %	41,0 %
peur	14,5 %	40,6 %	4,3 %	40,5 %
surprise	2,7 %	23,1 %	**52,3 %**	21,9 %
tristesse	14,0 %	28,6 %	10,9 %	46,5 %

Tableau 56: Pour chaque catégorie émotionnelle, proportion représentée par les modes d'expression dans le train

Le tableau 55 présente les F1 du modèle E sur le *test* pour chaque classe de catégorie émotionnelle, en fonction du mode d'expression des informations émotionnelles à retrouver. Les cases barrées correspondent à des combinaisons absentes du *test*. Le tableau Erreur : source de la référence non trouvée expose quant à lui, pour chaque catégorie émotionnelle, la proportion des différents modes d'expression parmi les phrases du *train* exprimant cette émotion.

Le tableau 55 montre que la F1 est globalement plus élevée quand la classe émotionnelle est exprimée par le mode désigné. Cette observation, cohérente avec la meilleure reconnaissance de ce mode par le modèle E (*cf.* section 1), n'est cependant pas avérée pour les classes « colère » et « surprise ». La première est mieux retrouvée quand elle est exprimée par le mode comportemental et la seconde quand elle est montrée. Ces différences sont expliquées par le tableau Erreur : source de la référence non trouvée où la COLÈRE apparaît fortement associée au mode comportemental (48 % des phrases exprimant de la COLÈRE) et la SURPRISE au mode montré (52 %).

Ce facteur seul n'est cependant pas suffisant pour expliquer la meilleure capacité du modèle E à retrouver la COLÈRE comportementale et la SURPRISE montrée. Le DÉGOÛT est par exemple fortement associé au mode suggéré (55 %). Cette classe n'est pourtant jamais prédite à cause de sa sous-représentation dans le *train* (0,18 % des phrases du *train*). L'exemple de la JOIE indique qu'un troisième facteur explique la variation des performances : les caractéristiques linguistiques du mode. Si cette émotion est plus souvent suggérée (41 %) que désignée (24,8 %) dans le *train*, le modèle E la retrouve mieux quand elle est désignée (F1 de 0,531 contre 0,275). De manière similaire, la PEUR, presque autant associée au désigné (40,6 %) qu'au suggéré (40,5 %) est mieux prédite quand elle est désignée (F1 de 0,626 contre 0,341). Le mode suggéré est en effet le plus difficile à reconnaître, tant pour la machine que pour les humains.

Le mode d'expression influence donc la capacité du modèle à reconnaître les catégories émotionnelles. Si une catégorie et un mode sont fortement associés dans le *train*, le modèle retrouve mieux l'émotion quand elle est exprimée par ce mode (par ex. COLÈRE comportementale, SURPRISE montrée). Cependant, ce phénomène est tempéré par le nombre d'occurrences disponibles dans le *train* (par ex. DÉGOÛT) et par la nature du mode d'expression impliqué dans l'association (par ex. désigné *vs.* suggéré).

4.2) Observation 2 : Impact de la catégorie émotionnelle sur la prédiction du mode.

L'objectif est de déterminer si les prédictions de modes d'expression sont influencées par la catégorie émotionnelle exprimée. Par exemple, le modèle E retrouve-t-il mieux le mode comportemental quand il exprime de la COLÈRE ? Pour explorer cette question, les informations émotionnelles du *test* sont évaluées catégorie émotionnelle par catégorie émotionnelle. La démarche ayant permis de filtrer les informations nécessaires à ces observations est détaillée en Annexe X.

	adm.	**col.**	**culp.**	**deg.**	**emb.**	**fier.**	**jal.**	**joie**	**peur**	**surp.**	**tris.**
comp.	0,037	0,472	0		0		0,009	0,173	0,226	0,028	0,109
désigné	0,059	0,180	0,012	0,018	0,071	0,054	0	0,268	0,501	0,208	0,175
montré	0,072	0,129		0,011	0,043	0,033		0,329	0,021	0,530	0,155
suggéré	0,053	0,179	0	0,025	0,025	0,058		0,177	0,241	0,082	0,140

Tableau 57: F-mesure des classes de mode selon la catégorie émotionnelle, modèle E sur le test

Le tableau 57 présente les variations de F-mesure obtenues par le modèle E sur le *test* pour la prédiction du mode, selon la catégorie émotionnelle exprimée par l'information émotionnelle à retrouver. Le mode comportemental y apparaît mieux retrouvé quand de la COLÈRE est exprimée. De même, pour le mode montré et la SURPRISE. Les associations fortes entre catégorie émotionnelle et mode d'expression, déjà constatées pour l'impact du mode sur la prédiction de la catégorie émotionnelle, apparaissent à nouveau ici. Pour le désigné, c'est l'expression de la PEUR qui semble faciliter la reconnaissance du mode. Parmi les émotions les plus fréquentes du corpus, c'est en effet celle qui lui est le plus souvent associée dans le *train* (*cf.* tableau Erreur : source de la référence non trouvée). Pour le suggéré, aucune catégorie émotionnelle ne se démarque : ce mode est toujours mal prédit. Il semble tout de même mieux reconnu pour les catégories les plus fréquentes du *train*, soit la COLÈRE, la PEUR, la JOIE et la TRISTESSE (*cf.* tableau 55), à l'exception de la SURPRISE, sans doute trop fortement associée au mode montré.

Ainsi, les prédictions du mode et de la catégorie émotionnelle apparaissent, bien entendu, influencées par les caractéristiques du *train*. Le modèle retrouve mieux la catégorie émotionnelle quand elle est exprimée par un mode auquel elle est fortement associée dans le *train* et inversement pour la prédiction du mode. Cette effet disparaît

en revanche pour le mode suggéré, toujours mal reconnu. Cela souligne la difficulté du modèle E à capter ce mode, de fait le plus difficile à annoter pour les humains.

5) Confrontation des annotations humaines et des prédictions machine

Les analyses des résultats du modèle E des sections 1 et 3 en particulier font émerger le besoin d'étudier de plus près des phrases extraites du corpus de *test* pour mieux comprendre les erreurs produites par le modèle. De l'observation des prédictions du modèle E à l'échelle de la phrase se dégagent plusieurs cas de figure de (dés)accord entre les annotations humaines et les prédictions du modèle.

Accord total sur l'analyse émotionnelle. Dans le meilleur des cas, l'analyse humaine et l'analyse machine coïncident totalement. Certaines combinaisons semblent particulièrement consensuelles. Les énoncés [30] à [32] sont par exemple tous annotés comme comportant de la SURPRISE montrée et prédits comme réalisant les classes « surprise » et « montrée ». La présence des points d'exclamation et/ou de points de suspension semble ici être le critère déterminant pour l'analyse émotionnelle.

[30] Après avoir donné des (faux) coups à son adversaire, il lui a... rasé la tête !
[31] Mais au lieu d'ouvrir la porte des toilettes, l'étourdie a ouvert... l'issue de secours !
[32] Il avait été transmis par... une civette.

La COLÈRE comportementale fait également l'unanimité entre annotations humaines et prédictions machine, comme pour les énoncés [33] et [34] (marqueurs « ont accusé », « manifestent », « bloquent des routes » en [33] et [34] « est critiqué »). Cette émotion est aussi suggérée en [33] (« avoir menti », « le Président prend des décisions injustes ») et analysée comme telle par le modèle E.

[33] Dans le monde entier, des gens ont accusé l'armée américaine d'avoir menti.
[34] Une volte-face qui interroge sur la fiabilité des chiffres officiels, alors que le manque de transparence dans la communication du Parti communiste chinois est critiqué par la population.

L'accord concerne parfois des cas très complets, comme l'énoncé [35] où la PEUR est à la fois désignée (« dangereux »), comportementale «(« fuir ») et suggérée (« tuer ceux qui partent », « punir leurs familles ») mais s'applique également au fait de considérer comme non-émotionnelle la phrase, comme pour l'énoncé [36].

[35] Plus de 360 000 Erythréens ont réussi à fuir leur pays mais c'est très dangereux : la police a pour mission de tuer ceux qui partent et de punir leurs familles.

[36] Généralement, les éditions ont lieu tantôt sur un parcours américain, tantôt sur un parcours européen.

Accord partiel sur l'analyse émotionnelle. Quand humain et machine considèrent tous deux la phrase comme émotionnelle, ils peuvent tout de même effectuer une analyse divergente. L'énoncé [37] est par exemple analysé avec de la TRISTESSE comportementale par l'annotateur humain, tandis que la machine prédit les classes « tristesse » et « désignée ». Le désaccord porte donc ici sur le mode d'expression de l'émotion. L'analyse humaine semble s'être focalisée ici sur l'action de « rendre hommage » comme acte de tristesse, tandis que la machine a probablement repéré les termes « blessés » et « traumatisées », proche du lexique émotionnel de la tristesse. L'analyse proposée par le modèle E invite ici à repenser l'annotation de la phrase[68].

[37] Les familles des victimes décédées, les blessés, ainsi que de nombreuses personnes traumatisées par cet événement se sont donc réunies dimanche, exactement un an après les faits, pour rendre hommage aux personnes tuées.

Le désaccord est parfois sur la catégorie émotionnelle. L'énoncé [38] est par exemple analysé, par l'humain comme par la machine, avec les modes montré (notons au passage la présence de points d'exclamation) et désigné. L'annotateur voit cependant en [38] de l'EMBARRAS, là où le modèle E prédit la classe « tristesse ». L'analyse humaine paraît ici plus pertinente que l'analyse machine.

[38] C'était très gênant parce que toute l'équipe du film était déjà montée sur scène !

L'accord entre humain et machine n'est parfois que sur le caractère émotionnel de la phrase. Le modèle E prédit par exemple la classe « autre » (sans classe de mode d'expression) pour [39], alors qu'il comporte du DÉGOÛT désigné. La rareté de cette catégorie dans le *train* explique la mauvaise analyse effectuée par le classifieur.

[39] Au menu : 80 spécialités culinaires du monde entier réputées pour leur capacité à susciter le dégoût.

Désaccord sur l'analyse émotionnelle. Humain et machine sont parfois en total désaccord sur l'analyse émotionnelle d'une phrase. L'annotateur identifie par exemple

68 S'il n'est présent ni dans les annotations humaines, ni dans les prédictions machine, le mode suggéré pourrait aussi être employé pour décrire cette phrase (« victimes », « tuées »).

dans l'énoncé [40] de la PEUR comportementale et désignée et de l'EMBARRAS exprimé par ces deux modes. Le modèle considère la phrase comme non-émotionnelle.

[40] Elle avance timidement et s'installe avec la troupe.

Si l'analyse produite par le classifieur pour l'énoncé [40] est clairement erronée, il est plus difficile de trancher pour l'énoncé [41] où l'annotateur a vu de la PEUR comportementale (« fait plus attention ») et suggérée (« au cas où il y aurait une bombe dedans »). La phrase, bien que comportant des termes négatifs, reste assez descriptive. Pour cette raison peut-être, le modèle a considéré qu'elle n'était pas émotionnelle.

[41] Par exemple, il y a plus de policiers et de militaires dans les grandes gares, et on fait plus attention s'il y a un bagage abandonné : il sera plus vite détruit, au cas où il y aurait une bombe dedans.

Mais en cas de désaccord total sur l'analyse émotionnelle, les propositions du modèle E sont parfois très pertinentes. Tous les énoncés [42] à [44] sont considérés comme non émotionnels par l'annotateur humain. En [42], le modèle prédit pour sa part les classes « tristesse » et « désignée », certainement d'après le terme « douleur ». Les classes « joie » et « suggérée » associées à [43] sont probablement liées à la présence du syntagme « as vraiment de la chance ». L'analyse de [44], où les classes « tristesse » et « suggérée » sont prédites, est particulièrement pertinente : rentrer « bredouille » implique la notion d'échec, tout à fait apte à suggérer des ressentis émotionnels négatifs comme la tristesse. Notons que les analyses proposées pour ces énoncés, sont considérées comme erronées dans l'analyse chiffrée des résultats.

[42] Il ne faut pas qu'ils sentent la douleur quand on les tue, donc les personnes qui travaillent dans l'abattoir doivent les rendre inconscients.

[43] Tu as vraiment de la chance, lecteur de ce P'tit Libé, parce que normalement on ne fait de la philosophie que la dernière année du lycée, quand on a 17 ans.

[44] Ceux-ci sont finalement repartis bredouilles.

Les cas de (dés-)accord entre les analyses humaines et les analyses du modèle E mettent en avant la difficulté de la tâche d'analyse des émotions. Ils poussent parfois à se demander qui de l'humain ou de la machine a « raison ». Pour cela, une expérience de validation perceptuelle auprès d'experts a été menée section 6.

6) Évaluation perceptuelle par des experts

Les prédictions du modèle E et les annotations humaines de référence ne concordent pas toujours. Si ces prédictions sont quelquefois visiblement erronées, elles invitent parfois à la réflexion et poussent à questionner les annotations de référence (*cf.* section 5). Il est alors compliqué de déterminer qui, de l'humain ou de la machine, effectue l'analyse la plus pertinente, en raison de la difficulté de la tâche d'annotation des émotions.

Pour apporter des éléments de réponse à cette question, les annotations humaines et les prédictions du modèle E ont été confrontées dans le cadre d'une expérience de validation perceptuelle menée auprès de trois experts. Le protocole mis en place est détaillé en section 6.1 et des premiers résultats sont présentés en section 6.2.

6.1) Protocole expérimental

L'expérience mise en œuvre consiste à présenter à l'évaluateur humain une phrase associée à un ensemble d'étiquettes d'annotation émotionnelle et de lui demander s'il est d'accord ou non avec elles. Ces étiquettes correspondant soit à des annotations humaines, soit à des prédictions, l'objectif est ainsi de voir si l'évaluateur est plus souvent d'accord avec l'humain ou avec la machine. La source des étiquettes proposées n'est bien entendue par communiquée à l'évaluateur.

Informations évaluées lors de l'expérience. Seules les informations relatives aux classes de catégories émotionnelles et de modes d'expression sont évaluées lors de l'expérience. Voici les jeux d'étiquettes soumis aux évaluateurs, où « aucune » et « aucun » permettent de rendre compte du caractère non émotionnel des phrases :

- « aucune », « admiration », « autre », « colère », « culpabilité », « dégoût », « embarras », « fierté », « jalousie », « joie », « peur », « surprise », « tristesse » ;
- « aucun », « comportemental », « désigné », « montré », « suggéré ».

Pour chaque phrase du corpus expérimental, deux ensembles d'étiquettes sont soumis aux évaluateurs : l'un pour l'analyse de la catégorie émotionnelle ; l'autre pour celle du mode. Pour chaque étiquette soumise, l'expert indique s'il est « d'accord », « plutôt d'accord », « plutôt pas d'accord » ou « pas d'accord » avec la valeur proposée.

Matériel expérimental. Le matériel expérimental a été constitué à partir du *test*. Les phrases sont présentées avec leur contexte (une phrase avant, une phrase après) aux évaluateurs pour garantir des conditions proches de celles du processus d'annotation des données de référence (annotation en contexte). Ce type d'entrée génère de plus les meilleures performances pour les classifieurs. Le modèle dont les prédictions sont mobilisées dans l'expérience est le modèle E, soit le meilleur classifieur de la tâche E, entraîné à partir de phrases avec contexte. En choisissant ce modèle, les prédictions faites pour les classes de catégorie émotionnelle et pour celles de mode peuvent être observées simultanément.

Constitution du corpus expérimental. Le corpus expérimental contient 150 phrases, 100 émotionnelles et 50 non-émotionnelles, pour une durée d'expérience d'une demie-journée maximum. Les 50 phrases non-émotionnelles ont été sélectionnées sur un seul critère : l'accord entre l'humain et la machine sur le caractère non-émotionnel. La sélection des 100 phrases émotionnelles, plus complexe, s'est faite en tirant aléatoirement des phrases du *test* pour respecter les contraintes suivantes :

- toutes les classes évaluées doivent être représentées si elles sont présentes dans le *test*, même les plus rares ;
- tous les couples (c,m) formé par une catégorie émotionnelle c exprimée selon un mode m présents dans le *test* doivent être représentés ;
- pour chaque classe, lorsque cela est possible, tous les types d'accord entre humain et machine doivent être représentés, à savoir : *accord sur la présence* de la classe ; *désaccord sur la présence* de la classe : l'humain considère qu'elle est présente, la machine non ; *désaccord sur la présence* de la classe : l'humain considère que la classe n'est pas présente, la machine oui ;
- avoir les proportions les plus équilibrées possibles entre les différentes classes de catégories émotionnelles, les différentes classes de modes et les différents types d'accord.

Les caractéristiques du corpus expérimental ainsi constitué sont présentées dans les tableaux 58 et 59 en proportions de phrases.

	Annotation humaine	**Prédiction machine**
émo.	55,3 %	58,0 %
non-émo.	44,7 %	42,0 %
comp.	20,7 %	19,3 %
dési.	22,0 %	20,0 %
montré	20,0 %	18,7 %
sugg.	22,0 %	24,0 %
adm.	4,7 %	2,0 %
autre	4,7 %	9,3 %
colère	9,3 %	10,7 %
culp.	2,7 %	0,0 %
dég.	5,3 %	0,0 %
emb.	7,3 %	4,0 %
fierté	2,7 %	2,0 %
jalou.	0,7 %	0,0 %
joie	11,3 %	10,7 %
peur	12,7 %	14,7 %
surpr.	11,3 %	11,3 %
tris.	9,3 %	9,3 %

Tableau 58: Proportion de phrases selon chaque classe – corpus expérimental

Accord entre humain et machine			**Prop.**
accord		absence d'émotion	33,33 %
		présence d'une cat. émo.	35,33 %
		présence d'un mode	40,67 %
désaccord	présence d'une cat. émo.	oui pour l'humain	33,33 %
		oui pour la machine	30,67 %
	présence d'un mode	oui pour l'humain	26,00 %
		oui pour la machine	24,67 %

Tableau 59: Accords entre analyses humaine et machine, en proportions de phrases du corpus expérimental

Choix des évaluateurs. Trois évaluateurs ont été sollicités, sélectionnés sur leur maîtrise du schéma d'annotation employé pour constituer les données de référence.

Mise en œuvre de l'expérience. Le corpus expérimental a été mis en forme dans un tableur, dont une feuille est consacrée à l'évaluation de l'analyse de la catégorie émotionnelle et une autre feuille à l'évaluation de l'analyse du mode. Ce tableur a été diffusé par mail aux évaluateurs, accompagné d'un document de consignes de sorte que l'expérience puisse être effectuée en totale autonomie. Ce document est disponible en Annexe XII. Lorsqu'une même phrase dénote plusieurs catégories émotionnelles, autant d'étiquettes de catégorie émotionnelle sont proposées. Dans le tableur, cela se traduit par une duplication de la phrase, où chaque ligne correspond à une valeur proposée. Le même raisonnement est appliqué pour le mode d'expression. Au total, 224 étiquettes de catégorie émotionnelle et 218 de mode ont été soumises aux experts.

6.2) Résultats

L'ensemble des 150 phrases, soit des 224 étiquettes de catégorie émotionnelle et des 218 étiquettes de mode, a été évalué par chacun des trois experts. Face à une proposition d'annotation, les évaluateurs devaient indiquer s'ils étaient d'accord ou non avec la valeur proposée. Quatre avis différents pouvaient être renseignés : « d'accord », « plutôt d'accord », « pas d'accord » ou « plutôt pas d'accord ». Pour une très grande majorité des étiquettes évaluées, les experts sont d'accord avec la valeur proposée. Cet accord est cependant plus prononcé pour l'analyse du mode d'expression (81 % des étiquettes de catégorie émotionnelle et 94 % des étiquettes de mode). Le détail, en nombre d'étiquettes, est présenté en Annexe XI.

Les valeurs « d'accord » et « plutôt d'accord » d'une part, les valeurs « pas d'accord » et « plutôt pas d'accord » d'autre part peuvent être regroupées. La proportion d'étiquettes pour lesquelles au moins deux évaluateurs sont d'accord avec la valeur proposée augmente alors (90% pour l'analyse de la catégorie émotionnelle, 95% pour celle du mode) mais reste supérieure pour le mode. Les désaccords avec la valeur proposée portent alors sur 22 étiquettes de catégorie émotionnelle et 11 étiquettes de mode. Les résultats analysés par la suite s'appuient sur ces regroupement d'opinions pour répondre à deux questions principales.

Quelles sont les valeurs les plus consensuelles ? Les tableaux 60 et 61 présentent, pour chaque tâche d'évaluation, les proportions d'étiquettes soumises pour lesquelles au moins deux évaluateurs s'accordent avec la valeur proposée. Ces résultats ne différencient pas les étiquettes selon leur source (humain ou machine).

Toutes valeurs confondues, les experts sont d'accord avec la valeur proposée au moins trois fois sur quatre pour les catégories émotionnelles et huit fois sur dix pour les modes. Les tableaux 60 et 61 montrent que le *désaccord* le plus élevé est observé sur les classes « tristesse » et « embarras » pour la catégorie émotionnelle, et sur la classe « montré » pour le mode. Les proportions *d'accord* les plus hautes apparaissent sur les classes « admiration », « colère », « culpabilité », « fierté » et « jalousie », et sur le mode « suggéré » (98,1 %). L'observation faite sur les modes est étonnante puisque le mode suggéré est le plus difficile à annoter et le moins bien prédit par la machine. De manière générale, aucune valeur d'étiquette ne semble occasionner un consensus particulièrement fort en contraste avec l'ensemble des valeurs évaluées à partir du corpus expérimental, que ce soit pour la catégorie émotionnelle ou pour le mode.

Valeur proposée	Opinion des évaluateurs	Prop. parmi valeur (nb. étiquettes)
aucune	D'accord	94,0 % (47 étiq.)
	Pas d'accord	6,0 % (3 étiq.)
adm.	D'accord	100,0 % (8 étiq.)
autre	D'accord	82,4 % (14 étiq.)
	Pas d'accord	17,6 % (3 étiq.)
colère	D'accord	100,0 % (20 étiq.)
culpabilité	D'accord	100,0 % (4 étiq.)
dégoût	D'accord	87,5 % (7 étiq.)
	Pas d'accord	12,5 % (1 étiq.)
embarras	D'accord	78,6 % (11 étiq.)
	Pas d'accord	21,4 % (3 étiq.)
fierté	D'accord	100,0 % (5 étiq.)
jalousie	D'accord	100,0 % (1 étiq.)
joie	D'accord	91,3 % (21 étiq.)
	Pas d'accord	8,7 % (2 étiq.)
peur	D'accord	90,0 % (27 étiq.)
	Pas d'accord	10,0 % (3 étiq.)
surprise	D'accord	91,7 % (22 étiq.)
	Pas d'accord	8,3 % (2 étiq.)
tristesse	D'accord	75,0 % (15 étiq.)
	Pas d'accord	25,0 % (5 étiq.)

Tableau 60: Accord et désaccord selon la classe de catégorie émotionnelle

Valeur proposée	Opinion des évaluateurs	Prop. parmi valeur (nb. étiquettes)
aucun	D'accord	94,0 % (47 étiq.)
	Pas d'accord	6,0 % (3 étiq.)
comp.	D'accord	95,2 % (40 étiq.)
	Pas d'accord	4,8 % (2 étiq.)
désigné	D'accord	97,5 % (39 étiq.)
	Pas d'accord	2,5 % (1 étiq.)
montré	D'accord	88,2 % (30 étiq.)
	Pas d'accord	11,8 % (4 étiq.)
suggéré	D'accord	98,1 % (51 étiq.)
	Pas d'accord	1,9 % (1 étiq.)

Tableau 61: Accord et désaccord selon la classe de mode

Quand il y a une divergence d'analyse entre l'humain et la machine, qui a raison ? Autrement dit, avec quelle analyse les évaluateurs sont-ils principalement d'accord ? Le tableau 62 indique pour chaque tâche d'évaluation les proportions d'accord selon la source ayant proposé la valeur : humain, machine ou les deux. Comme mentionné plus haut, les experts sont de manière générale d'accord avec les valeurs proposées, quelles que soient la source et la tâche considérée (au moins 76,5 % d'accord). Pour les deux tâches d'évaluation, les proportions d'accord les plus élevées sont observées pour les étiquettes proposées à la fois par l'humain et par la machine. Pour le mode, les proportions d'accord sur les étiquettes humaines et les étiquettes machine sont très proches : cela illustre la capacité du modèle à bien reconnaître les modes d'expression, soit en retrouvant l'annotation humaine de référence, soit en proposant une analyse différente mais malgré tout pertinente (selon les évaluateurs). Pour la catégorie émotionnelle cependant, la proportion d'accord est plus faible quand la valeur

est proposée par la machine : cela souligne une fois encore que le modèle E a plus de mal à identifier les catégories émotionnelles.

Source de la valeur proposée	**Opinion des évaluateurs**	**Prop. parmi les propositions de la source (nb d'étiquettes)**	
		cat. émo.	***mode***
humain & machine	D'accord	95,5 % (105 étiq.)	97,7 % (129 étiq.)
	Pas d'accord	4,5 % (5 étiq.)	2,3 % (3 étiq.)
humain	D'accord	92,1 % (58 étiq.)	91,1 % (41 étiq.)
	Pas d'accord	7,9 % (5 étiq.)	8,9 % (4 étiq.)
machine	D'accord	76,5 % (39 étiq.)	90,2 % (37 étiq.)
	Pas d'accord	23,5 % (12 étiq.)	9,8 % (4 étiq.)

Tableau 62: Proportion d'accord selon la source des valeurs proposées, pour chaque tâche d'évaluation

Les brèves analyses présentées dans cette section ne font qu'effleurer toutes les possibilités offertes par les résultats de l'expérience de validation perceptuelle. Leur exploration mériterait d'être approfondie, par exemple en étudiant les valeurs de catégorie émotionnelle et de mode proposées par les experts en cas de désaccord avec l'étiquette ; en mesurant l'influence de l'individu (expert qui évalue mais aussi humain source des annotations) ; ou en effectuant un retour sur les données textuelles.

Les premières analyses rapportées mettent malgré tout au jour un phénomène important : la grande capacité de généralisation du modèle E. Elles soulignent ainsi que parmi les prédictions effectuées uniquement par la machine – donc considérées comme erronées dans l'évaluation chiffrée des performances – une grande majorité est en fait validée par les experts humains. Ce constat invite à faire un retour sur le corpus annoté, voire sur le processus d'annotation et le guide eux-mêmes.

7) Comparaison avec le processeur ÉMOTIONS de la chaîne TextToKids

La chaîne d'extraction de descripteurs de complexité des textes développée dans le cadre du projet TextToKids comporte un processeur d'analyse des émotions nommé ÉMOTIONS (*cf.* chapitre 1). Il s'appuie sur deux modules : l'un fait appel à l'outil TextBlob et l'autre repose sur le lexique EMOTAIX (Piolat & Bannour, 2009), remanié dans le cadre du projet TextToKids. L'objectif est de comparer les performances de ces deux modules avec celles du meilleur classifieur de la tâche E entraîné à partir

de phrases isolées, le modèle E'. Il s'agit en effet du classifieur sous-jacent au processeur EMOTYC que j'ai développé pour l'analyse des marqueurs émotionnels de complexité des textes, d'ores et déjà intégré à la chaîne TextToKids (*cf.* chapitre 1).

7.1) Principes des comparaisons

Module basé sur TextBlob. Le premier module – ci-après nommé module TextBlob – s'appuie sur la bibliothèque TextBlob[69] pour effectuer une analyse en sentiments. TextBlob assigne à chaque phrase un score de polarité compris dans [-1,1][70]. Ce score peut être mis en regard avec l'analyse opérée par le modèle E' sur deux aspects : le *caractère émotionnel* (avec assimilation à une phrase « émotionnelle » de toute phrase considérée comme non-neutre par le module TextBlob) ; et la *polarité* (avec extrapolation des classes « positive » et « négative » depuis les classes de catégories émotionnelles du modèle E'). Le processus rendant comparable les prédictions du modèle E' (valeurs booléennes) et les scores du module TextBlob (valeurs continues) est détaillé en Annexe XIII.

Module basé sur le lexique EMOTAIX remanié. Le second module – ci-après nommé module EmotaixRemanié – analyse la présence de termes du lexique émotionnel d'après un remaniement du lexique EMOTAIX (Piolat & Bannour, 2009) que j'ai effectué, détaillé en Annexe XIV. Ce lexique remanié comporte 24 catégories pouvant être réparties entre les douze classes de catégories émotionnelles du modèle E' comme illustré tableau 63. Le lexique contient également une catégorie 'comportement' à mettre en relation avec la classe « comportementale » du modèle E'.

Les critères ayant servi la constitution du lexique EMOTAIX sont repris dans le guide d'annotation (Étienne & Battistelli, 2021) où ils interviennent pour circonscrire et identifier le mode désigné. Mise à part ceux de la catégorie 'comportement', l'ensemble des termes inclus du lexique EMOTAIX *désignent* donc une émotion.

Plusieurs points de comparaison peuvent ainsi être établis entre les analyses du module EmotaixRemanié et les prédictions du modèle E' : le mode comportemental ; le mode désigné ; la catégorie émotionnelle AUTRE ; onze catégories émotionnelles dési-

69 https://textblob.readthedocs.io/en/dev/

70 Pour effectuer ses analyses, TextBlob s'appuie en fait sur une autre librairie (https://github.com/clips/pattern/wiki/pattern-fr). Cette seconde librairie mobilise un lexique dont les termes sont associés à des scores de subjectivité et de polarité. Ce lexique est disponible ici : https://github.com/clips/pattern/blob/master/pattern/text/fr/fr-sentiment.xml.

gnées ; le caractère émotionnel. Les regroupements d'analyses du module EmotaixRemanié permettant ces comparaisons sont exposés en Annexe XIII. Pour être comparées aux prédictions du modèle E', les analyses du module EmotaixRemanié, exprimées sous la forme d'un vecteur de proportions, doivent être traduites en valeurs booléennes. La démarche de conversion est détaillée en Annexe XIII.

Classe prédite par le modèle E'	**Catégorie du lexique EMOTAIX révisé**
admiration	admiration
autre	amour, apaisement, audace, désir, empathie, impassibilité, inhumanité, mépris, non spécifiée, ressentiment
colère	colère, déplaisir
culpabilité	culpabilité
dégoût	dégoût
embarras	embarras
fierté	fierté, orgueil
jalousie	jalousie
joie	joie
peur	peur
surprise	surprise
tristesse	tristesse
comportementale	comportement

Tableau 63: Correspondances entre les catégories du lexique EMOTAIX révisé et les classes du modèle E'

Sept tâches de comparaison. Des points de comparaison identifiés entre le module TextBlob et le modèle E' d'une part, puis entre le module EmotaixRemanié et le modèle E' d'autre part, sept tâches de comparaison sont distinguées : (1) l'identification de phrases émotionnelles ; (2) l'identification de phrases positives ; (3) l'identification de phrases négatives ; (4) l'identification du mode comportemental ; (5) l'identification du mode désigné ; (6) l'identification de la classe « autre » ; (7) l'identification de onze catégories émotionnelles désignées. La tâche 1 implique les deux modules de ÉMOTIONS, les tâches 2 uniquement le module TextBlob et les autres uniquement le module EmotaixRemanié.

7.2) Résultats des comparaisons

Les performances obtenues pour les sept tâches de comparaison sur le *test vus* et le *test non vus* sont synthétisées par le tableau 64. L'évaluation de la tâche d'identification de catégories émotionnelles désignées correspond à des macro-mesures. Le détail des performances est présenté en Annexe XIII.

Quelle que soit la tâche considérée, les F1 sont systématiquement inférieures sur le *test non vus* par rapport au *test vus*. Cette observation confirme encore la particularité du *test non vus* (*cf.* section 2). La suite des analyses s'appuie uniquement sur les performances obtenues sur le *test vus*. Sur ce corpus, les F1 du modèle E' sont systématiquement les meilleures.

		test vus			***test non vus***		
Tâche	**Outil**	**(macro-)R**	**(macro-)P**	**(macro-)F1**	**(macro-)R**	**(macro-)P**	**(macro-)F1**
phrases émo.	Modèle E'	0,735	0,798	0,765	0,935	0,372	0,532
	Module TextBlob	0,987	0,221	0,361	1,000	0,044	0,085
	Module EmotaixRem.	0,374	0,604	0,462	0,532	0,232	0,324
phrases pos.	Modèle E'	0,474	0,730	0,575	0,813	0,448	0,578
	Module TextBlob	0,145	0,206	0,170	0,250	0,080	0,121
phrases nég.	Modèle E'	0,649	0,749	0,695	0,864	0,422	0,567
	Module TextBlob	0,959	0,113	0,202	0,977	0,032	0,061
mode comp.	Modèle E'	0,607	0,768	0,678	0,667	0,261	0,375
	Module EmotaixRem.	0,022	0,714	0,042	0,000	0,000	0,000
mode dés.	Modèle E'	0,780	0,854	0,815	0,975	0,565	0,716
	Module EmotaixRem.	0,635	0,545	0,587	0,675	0,314	0,429
classe « autre »	Modèle E'	0,652	0,649	0,651	1,000	0,026	0,051
	Module EmotaixRem.	0,386	0,296	0,335	1,000	0,016	0,031
cat. désignées	Modèle E'	0,428	0,526	0,463	0,525	0,348	0,400
	Module EmotaixRem.	0,483	0,527	0,440	0,302	0,230	0,249

Tableau 64: Comparaison des performances du modèle E' et des modules TextToKids sur le test vus et le test non vus

Module TextBlob. Pour les trois tâches où il est impliqué, le module TextBlob obtient de bien médiocres performances (F1 toujours inférieure à 0,5). Sur l'identification de phrases émotionnelles et de phrases négatives, son rappel est cependant ex-

cellent (plus de 0,95). Ce phénomène s'explique peut-être par la différence des champs d'analyse des notions de polarité et d'émotion. La question qui émerge ici – et qui reste sans réponse – est de savoir si toutes les phrases positives/négatives contiennent une émotion et sont donc à considérer comme émotionnelles. Les performances du module TextBlob restent malgré tout décevantes et l'intégration du modèle E' *via* EMOTYC dans la chaîne TextToKids offre des analyses plus pertinentes.

Module EmotaixRemanié. Ce module intervient dans six tâches de comparaison.

Identification de phrases émotionnelles. Comme pour le module TextBlob, l'identification de phrases émotionnelles par le module EmotaixRemanié est insatisfaisante, en particulier du point de vue du rappel. Le modèle E' reconnaît plus d'informations émotionnelles (suggérées et montrées notamment). L'avantage de prendre en compte des informations ne reposant pas uniquement sur le lexique émotionnel (mode désigné) pour analyser la dimension émotionnelle des textes est ainsi illustré.

Identification du mode comportemental. Les performances du module EmotaixRemanié chutent à nouveau à cause du rappel pour l'identification du mode comportemental. De toute évidence, la collection de termes étiquetés 'comportement' dans le lexique EMOTAIX remanié est insuffisante pour capter toute la diversité des comportements émotionnels annotés dans le corpus *test vus*, que le modèle E' est en revanche capable d'identifier. Pour cette tâche, c'est cette fois la précision du module qui chute, peut-être à cause de la polysémie de termes inclus dans le lexique EMOTAIX remanié. L'annotation manuelle des textes ayant été faite en contexte, un terme du lexique employé dans le corpus pour son sens non émotionnel ne sera pas annoté.

Identification de la classe « autre ». Le module obtient un mauvais rappel et une mauvaise précision pour l'identification de la classe « autre ». Le rappel bas est dû au fait que le lexique EMOTAIX remanié ne saurait être exhaustif. La précision basse s'explique quant à elle sûrement encore par un problème de polysémie des termes. Remarquons tout de même que pour cette tâche, les performances du modèle E' et du module EmotaixRemanié sont très comparables sur le *test non vus*.

Identification de catégories émotionnelles désignées. Pour la tâche d'identification de catégories émotionnelles désignées, le modèle E' et le module EmotaixRemanié ont des performances comparables, avec un léger avantage du module sur le macro-R et la macro-P[71]. Ce sont en fait les classes rares du corpus – « culpabilité »,

71 Le modèle E' a malgré tout une meilleure macro-F1 car les F-mesures pour chaque classe sont en moyenne plus élevées pour le modèle E', voir détail des résultats en Annexe XIII.

« dégoût », « embarras » et « jalousie » que le module retrouve mieux (voir détail des résultats en Annexe XIII). Là où le modèle E' n'avait pas suffisamment d'occurrences pour apprendre à reconnaître ces classes, le module parvient toujours à repérer les termes du lexique émotionnel. L'autre classe mieux retrouvée par le module est « surprise », probablement à cause de la très forte association entre cette catégorie et le mode montré dans le *train*, qui gêne la reconnaissance de l'émotion par le modèle E' quand elle est exprimée par d'autres modes (*cf.* section 4). Les six autres classes sont mieux reconnues par le modèle E'.

Ainsi, la comparaison du modèle E' avec le module TextBlob souligne l'importance d'utiliser un modèle optimisé pour la reconnaissance d'informations émotionnelles pour analyser la dimension émotionnelle des textes, plutôt que de simplement employer un outil d'analyse de sentiments. Celle avec le module EmotaixRemanié pointe quelques faiblesses du classifieur (classes rares, catégories émotionnelles fortement associées à un mode) mais révèle surtout la robustesse du modèle E'. La modélisation de la notion d'émotion sur laquelle repose le classifieur apparaît ici comme un atout majeur : en ne se limitant pas au lexique émotionnel, le modèle E' est capable de retrouver un plus grand nombre d'informations émotionnelles et donc de fournir une analyse plus approfondie, plus fine de la dimension émotionnelle des textes. Les écarts de performances entre le modèle E' et le module EmotaixRemanié illustrent également l'avantage d'avoir recours à un modèle issu de techniques d'apprentissage automatique. La non-exhaustivité inhérente aux lexiques émotionnels, de même que le phénomène de polysémie apparaissent ici comme des entraves puissantes aux performances du système fondé sur l'usage d'un lexique émotionnel. Le modèle E', de par sa capacité à généraliser, n'est quant à lui pas affecté par ces contraintes.

Conclusion

L'observation classe par classe des performances d'un classifieur transformeur de la tâche E, le modèle E, a tout d'abord permis de réaffirmer la faisabilité des tâches de classification définies chapitre 5, tout en soulignant des disparités d'efficacité de repérage entre classes au sein d'une même tâche. Les modes désigné et montré sont par exemple bien mieux reconnus que le mode suggéré. Les classes « surprise » et « colère » sont également bien identifiées, contrairement aux classes « jalousie », « dégoût » et « culpabilité », jamais prédites.

Une analyse qualitative des prédictions du modèle E a permis d'identifier des facteurs expliquant ces différences de performance : la fréquence, dans le *train*, de la classe à prédire ; les caractéristiques linguistiques des modes d'expression ; et l'utilisation privilégiée de certains modes pour exprimer certaines catégories (par ex. COLÈRE comportementale, SURPRISE montrée).

Par ailleurs, une confrontation des prédictions du modèle E avec des analyses humaines, d'une part avec l'observation de quelques phrases d'exemple tirées du corpus annoté et d'autre part avec la mise en place d'une expérience de validation perceptuelle auprès d'experts, met en lumière le fait que les prédictions du modèle sont parfois plus justes que les annotations humaines de référence. Des pistes d'amélioration des critères guidant l'annotation pourraient ainsi être dégagées.

Enfin, la comparaison des résultats du modèle E' avec le processeur ÉMOTIONS de la chaîne TextToKids met en avant les forces de E' pour l'analyse de la dimension émotionnelle des textes. Ce modèle, déjà intégré dans la chaîne au sein de EMOTYC, retrouve en effet beaucoup plus d'informations émotionnelles que les modules de ÉMOTIONS et en propose une analyse plus fine, notamment grâce à la prise en compte de quatre modes d'expression des émotions. Autre avantage majeur : l'analyse opérée par le modèle E' rend compte de marqueurs émotionnels de complexité identifiés dans la littérature psycholinguistique concernant les modes d'expression, les types d'émotions et les catégories émotionnelles[72].

72 EMOTYC et EMOTIONS sont actuellement tous deux utilisés dans le cadre d'un autre projet, le projet ANR FLYER, sur des textes dits « extrémistes ». L'objectif est de produire une évaluation précise de ces deux processeurs en termes de repérage d'informations émotionnelles mais aussi une évaluation en termes d'apport dans la tâche de classification en textes extrémistes *vs.* non-extrémistes.

Chapitre 7 Quelques observations menées à partir du corpus annoté en émotions

Parmi les notions impliquées pour décrire les émotions, le schéma présenté chapitre 3 distingue quatre modes d'expression (le comportemental, le désigné, le montré et le suggéré) et onze catégories émotionnelles, perçues comme des macro-catégories regroupant plusieurs émotions plus fines. Ces macro-catégories sont ADMIRATION COLÈRE, CULPABILITÉ, DÉGOÛT, EMBARRAS, FIERTÉ, JALOUSIE, JOIE, PEUR, SURPRISE et TRISTESSE). Le schéma a été appliqué manuellement *via* la plateforme Glozz par six annotatrices sur un corpus de 1 594 textes abordant des thématiques variées et de trois genres différents (encyclopédique, fictionnel, journalistique ; *cf.* chapitre 4). Dans ce chapitre, je rapporte un ensemble d'observations linguistiques menées à partir de ce corpus annoté.

Les annotations émotionnelles permettent d'abord une analyse approfondie de l'usage des modes d'expression (section 1). Le rôle que jouent les émotions dans la structuration d'un texte apparaît également grâce à l'annotation des émotions, de leurs causes et de leurs conséquences (section 2). Des observations préliminaires peuvent enfin être menées pour explorer les liens unissant expression des émotions et dimensions TAME (section 3).

1) Analyse des usages des modes d'expression des émotions

L'application du schéma d'annotation sur un corpus de textes écrits relevant de différents genres a permis de dégager des tendances sur les façons d'exprimer les émotions, avec la mise en évidence d'interactions entre d'une part le mode d'expression et le genre de texte et, d'autre part, deux notions du schéma : le mode d'expression et la catégorie émotionnelle. Ces interactions permettent de faire un retour sur quelques tendances déjà dégagées dans d'autres travaux de TAL. À titre indicatif, les analyses présentées dans cette section s'appuient sur un regroupement en genre de texte des sous-corpus composant le corpus annoté global (*cf.* chapitre 4). Le genre journalistique correspond aux sous-corpus PtitLibé, Albert, CovidAdultes et CovidJeunesse

(4 889 unités SitEmo au total) ; le genre fictionnel aux corpus FictionM2 et WP7Fiction (1 794 SitEmo) ; et le genre encyclopédique repose uniquement sur le corpus WP7Encyclo (121 SitEmo). Si les corpus fictionnel et encyclopédique présentent une taille et une diversité trop restreintes pour prétendre à une réelle généralisation des observations effectuées, les analyses pointent malgré tout certaines tendances intéressantes à souligner.

Les observations présentées ici impliquant les catégories émotionnelles se concentrent uniquement sur les cinq plus représentées dans le corpus annoté, soit la COLÈRE (25,2 % des unités SitEmo sur le corpus global), la PEUR (21,8 %), la SURPRISE (18,1 %), la JOIE (17,9 %) et la TRISTESSE (14,0 %)[73]. Il est à noter que si ces cinq catégories restent toujours les plus fréquentes, quelque soit le genre de texte considéré, leurs proportions varient. La COLÈRE est donc par exemple la catégorie la plus représentée dans les corpus journalistique et encyclopédique (respectivement 27,9 % et 34,7%) mais n'est que la troisième plus fréquente dans le corpus fictionnel (17,2 %). Dans ce corpus, la catégorie la plus représentée est la JOIE (34,4 %). Cette variation s'explique très certainement par les types de thématiques traitées dans ces différents corpus.

La significativité de certaines interactions présentées ici est mise en évidence par l'application d'un test exact de Fisher, avec un seuil à 0,99.

Interaction entre mode d'expression et genre de texte. Si à l'échelle du corpus entier les modes d'expression sont répartis à peu près équitablement (*cf.* chapitre 4), les proportions des modes varient selon le genre de texte, comme illustré figure 12. Les émotions sont ainsi plus montrées (36,5 %) dans les textes fictionnels que dans les autres genres (30,6 % pour l'encyclopédique et 15,0 % pour le journalistique). Les émotions désignées et comportementales sont quant à elles plus fréquentes dans les textes journalistiques (respectivement 28,6 % et 23,7 %, contre 14,6 % et 14,5 % dans les textes fictionnels, et 18,2 % et 19,8 % dans les textes encyclopédiques). Les émotions suggérées ne paraissent en revanche pas être spécifiques à un genre, puisqu'elles représentent toujours environ un tiers des SitEmo. L'application d'un test exact de Fisher souligne la significativité du lien entre genre fictionnel et émotions montrées[74], genre journalistique et émotions désignées et comportementales[75]. Le mode suggéré n'a en effet quant à lui aucune association significative.

73 Toutes les autres catégories représentent moins de 5 % des unités SitEmo.

74 p-valeur = 2,79e-73

75 Respectivement p-valeur = 5,46e-34 et p-valeur = 4,95e-16.

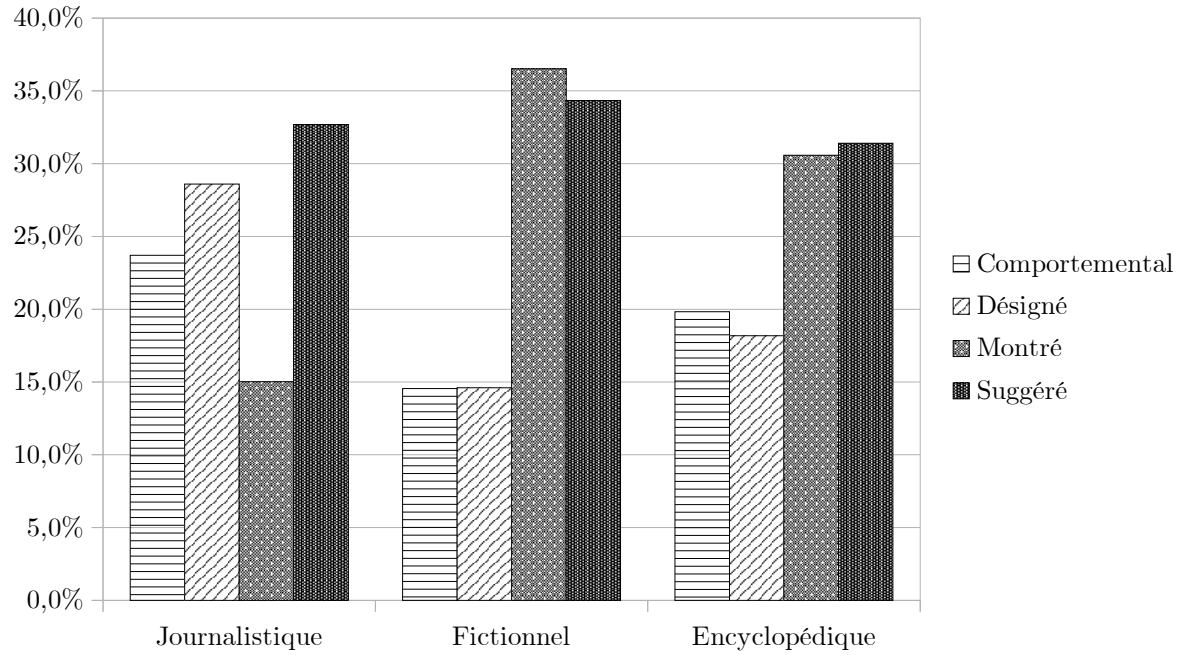


Figure 12: Proportion des modes d'expression selon le genre

Dans le corpus encyclopédique annoté en émotions, 30,6 % des unités SitEmo captent une émotion montrée, majoritairement ressentie par le scripteur (70,3 % des émotions montrées)[76]. Ce constat est surprenant car les émotions montrées se trouvent habituellement dans les discours rapportés directs ou interviennent pour montrer les émotions du scripteur, ce qui est principalement le cas dans le corpus WP7Encyclo. Pourtant, les textes encyclopédiques sont supposés décrire des faits avec un ton neutre, sans tenir compte des émotions du scripteur. Je suppose que le phénomène rencontré dans ce corpus est dû à la nature de la source des textes annotés. En effet, tous ces textes sont issus de l'encyclopédie en ligne collaborative Vikidia (*cf.* chapitre 4). Les personnes écrivant les articles ne sont ainsi pas nécessairement des experts du style encyclopédique. Ils essayent au contraire souvent de rendre les textes plus vivants, plus attrayants pour les enfants à l'aide de points d'exclamation et en indiquant clairement leur opinion sur le sujet traité par l'article. Tout cela prend la forme de marqueurs linguistiques qui montrent leurs émotions. Il est tout de même intéressant de noter que les émotions montrées restent plus fréquentes dans les textes de fiction (36,5 %).

76 Pour rappel, les textes encyclopédiques annotés en émotions dans le cadre de cette thèse comportent en moyenne 11 SitEmo. Les textes journalistiques n'en contiennent en moyenne qu'entre 3,04 et 6,03 ; les textes fictionnels entre 18,82 et 262,67 (*cf.* chapitre 4, section 4).

Ce résultat était attendu dans la mesure où ces textes contiennent de nombreux dialogues, propices à la monstration des émotions.

Le lien qui transparaît entre modes d'expression et genres de textes est l'une des interactions clés soulignées par l'application du schéma d'annotation. Ces interactions étaient déjà perçues par L.-A.-M. Bostan et R. Klinger (2018) lorsqu'ils déclarent que « les journalistes tendent idéalement à être objectifs quand ils écrivent des articles, [...] et l'on peut supposer que les expressions émotionnelles dans les contes sont plus subtiles et implicites » (traduit de l'anglais ; p. 2 105). De la même manière, E. Kim et R. Klinger (2018) attribuaient partiellement la difficulté d'annoter les émotions dans des triplets de phrases à la nature des textes auxquels ils étaient confrontés. Selon eux, « les textes fictionnels sont hautement métaphoriques et plein d'allusions et de métonymies, ce qui nécessite une lecture attentive (souvent même de lire entre les lignes) et un contexte plus large » (traduit de l'anglais ; p. 1 354).

Contrairement à ce qui est proposé dans (Bostan & Klinger, 2018; Kim & Klinger, 2018), les émotions suggérées ne sont pas spécifiques des textes fictionnels dans le corpus annoté en émotions présenté ici. Les interactions mises en évidence à partir du corpus concernent, comme mentionné précédemment, les émotions montrées et les textes fictionnels, les émotions désignées et comportementales et les textes journalistiques.

Interaction entre modes d'expression et catégorie émotionnelle. La proportion que représente chaque mode d'expression varie selon l'émotion exprimée, comme l'illustre la figure 13. La COLÈRE est ainsi majoritairement exprimée à travers des expressions comportementales de l'émotion (48,4 % des unités exprimant la COLÈRE), la JOIE et la TRISTESSE à travers des unités suggérées (respectivement 39,5 % et 47,6 %) et la SURPRISE à travers des unités montrées (60,9 %). La PEUR est presque autant suggérée (39,6%) qu'elle est désignée (40,3 %). L'application d'un test exact de Fisher confirme la significativité de ces liens[77].

L'étude des caractéristiques de phrases émotionnelles conduite par C. Alm (Alm, 2010) (*cf.* chapitre 2) suggérait déjà que « certaines caractéristiques peuvent montrer une affinité particulière avec certains affects » (traduit de l'anglais ; p. 121). Plus précisément, elle souligne que la catégorie « peur » semble être principalement véhiculée par des mots d'affect ou des descriptions de comportements et la catégorie « sur-

77 COLÈRE et mode comportemental : p-valeur = 4,09e-161 ; JOIE et suggéré : p-valeur = 1,19e-06 ; TRISTESSE et suggéré : p-valeur = 1,28e-21 ; SURPRISE et montré : p-valeur = 1,72e-228 ; PEUR et suggéré : p-valeur = 2,19e-08 ; PEUR et désigné : p-valeur = 7,60e-85.

prise » par des marqueurs trouvés dans des discours directs ou des descriptions d'observations inattendues. Pour leur part, D. Demszky et ses collègues (Demszky et al., 2020) ont trouvé que « certaines émotions sont plus verbalement implicites et peuvent nécessiter plus de contexte pour être interprétées » (traduit de l'anglais ; p. 6). Par exemple, dans leur corpus de commentaires Reddit, la catégorie « gratitude » était hautement corrélée au token *thanks* et la catégorie « amusement » au token *lol*, tandis que les catégories « deuil » et « nervosité » n'étaient associées de manière significative à aucun token.

Ces observations corroborent les résultats de l'annotation présentée ici puisque, sur l'ensemble du corpus, la SURPRISE est principalement exprimée par des unités montrées et la PEUR par des unités désignées et suggérées. Les résultats montrent également une association entre la COLÈRE et les expressions comportementales ainsi qu'une association entre le mode suggéré et la JOIE et la TRISTESSE.

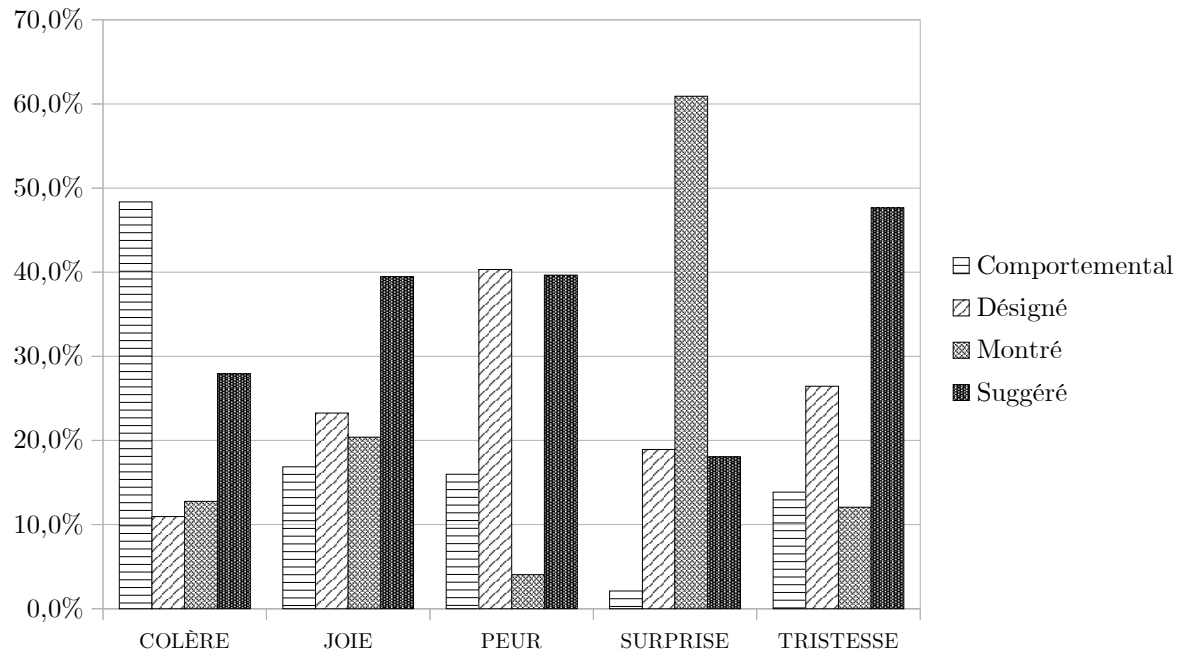


Figure 13: Proportion des modes d'expression selon la catégorie émotionnelle – Corpus global

Interaction entre mode d'expression et catégorie émotionnelle, selon le genre. Les associations entre mode d'expression et catégorie émotionnelle varient parfois selon le genre de texte considéré, comme illustré Figure 14. La COLÈRE est par exemple significativement associée au mode montré dans les textes de fiction (53,9 %

des unités SitEmo), alors qu'elle est en interaction avec le mode comportementale dans les textes journalistique (55 %) et encyclopédiques[78] (45,2 %).

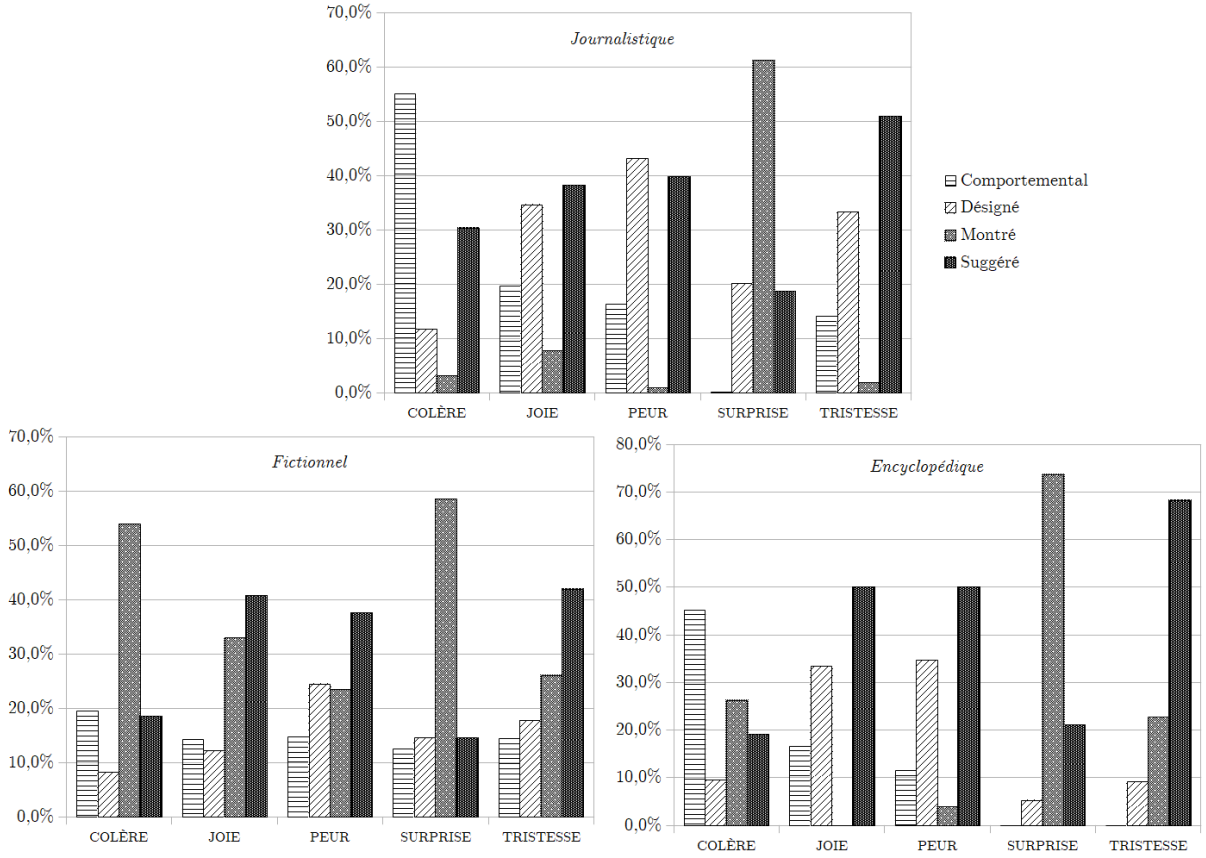


Figure 14: Proportion des modes d'expression selon la catégorie émotionnelle pour trois genres de textes

Dans les exemples [45] à [47] extraits de textes de fiction du corpus, les marqueurs soulignés montrent la COLÈRE. Ce sont cette fois des expressions comportementales de la COLÈRE qui sont soulignées dans les exemples [48] et [49], extraits respectivement d'un article journalistique et d'un article encyclopédique.

[45] Vous faites beaucoup trop de fautes ![79]

[46] « M'agace, celui-là... »[80]

[47] - C'est très joli, mais enfin, j'ai quatre pattes moi.[81]

78 Respectivement p-valeur = 8,27e-11, p-valeur = 2,96e-163 et p-valeur = 2,75e-05.

79 Exemple extrait de *La bande à Tristan*, M.-A. Murail.

80 Exemple extrait de *L'œil du loup*, D. Pennac.

81 Exemple extrait des *Boîtes de peintures*, M. Aymé.

[48] Certains films récompensés ont été très critiqués et leurs réalisateurs parfois sifflés à l'annonce des résultats.[82]

[49] Il est surtout connu pour la variété de ses nombreuses insultes, qu'il utilise lorsqu'il est en colère (autrement dit très souvent).[83]

Mode montré, mode comportemental et auteur·e de texte fiction. Au sein du corpus fictionnel, des variations de recours aux modes montré et comportemental sont observées selon l'auteur·e du texte. Par exemple, dans l'extrait de l'ouvrage de M. Aymé, l'emploi du mode comportemental (38,5 % des unités SitEmo) est significativement plus élevé que dans les textes des autres auteurs, tandis que dans les œuvres de S. Morgenstern, ce mode est significativement moins représenté[84] (11,3 %). L'extrait de M. Desplechin comporte moins d'émotions montrées que le reste du corpus fiction[85] (8,7 %). En revanche, les associations entre le mode montré et les autres auteurs ne sont pas significatives. Dans les exemples [50] à [52] extraits du texte *Les boîtes de peintures*, M. Aymé, les marqueurs montrant une émotion sont soulignés.

[50] Marinette rougit et l'âne poursuivit [...]

[51] Tout à coup, il eut comme un sursaut et s'écria [...]

[52] Delphine se mit à rire et Marinette rit aussi, quoiqu'elle eût un peu de remords.

Les résultats de l'annotation conduite sur le corpus de 1 594 textes variés en genres mettent ainsi en évidence le fait que certains modes d'expression ont tendance à être employés dans des genres de textes spécifiques. Des associations entre modes d'expression et catégories émotionnelles émergent aussi (par ex. entre la SURPRISE et le mode montré), pour certaines variant selon le genre de texte considéré (par ex. la COLÈRE est principalement montré dans les textes de fiction alors qu'elle est surtout comportementale dans les textes journalistiques et encyclopédiques). Des associations se dégagent également entre modes d'expression et auteur·e·s des textes fictionnels (par ex. M. Aymé emploie particulièrement le mode comportemental et A. Desplechin très peu de montré). Pour être pleinement confirmées, ces tendances nécessiteraient l'annotation d'un corpus mieux équilibré en genres, plus diversifié pour les sources encyclopédiques et comprenant des textes plus longs pour chaque auteur·e.

82 Exemple extrait de l'article « Le Festival de Cannes », *P'tit Libé*, n°21, mai 2017.

83 Exemple extrait de l'article « Tintin » de l'encyclopédique *Vikidia*.

84 Respectivement p-valeur = 0,0023 et p-valeur = 3,59e-07.

85 p-valeur = 0,0039

2) La fonction structurante des émotions observée à partir d'annotations

Par le repérage des émotions mais également de leurs expérienceurs, de leurs causes et de leurs conséquences, le schéma d'annotation permet d'analyser le fonctionnement textuel de l'expression des émotions. Cette analyse repose sur l'identification de blocs de phrases contiguës liées entre elles selon un critère sémantique, celui des émotions – et de leurs éventuelles causes et conséquences si elles sont exprimées – ressenties par des entités apparaissant au fur et à mesure d'un texte. L'étude de la fonction structurante des émotions s'appuie sur un raffinement du schéma présenté dans le guide (Étienne & Battistelli, 2021) (*cf.* chapitre 3). Les unités SitCause et SitConseq ne sont plus délimitées à l'échelle de la phrase graphique mais à l'échelle du marqueur linguistique, en suivant les mêmes principes que pour la délimitation des unités SitEmo ou Expérienceur. Deux nouvelles relations sont également introduites, CauseDe et ConseqDe, afin de lier ces unités à la ou les SitEmo auxquelles elles se rapportent. L'analyse complète permettant l'étude de la fonction structurante des émotions est détaillée dans (Battistelli, Étienne, et Lecorvé 2022) : seules les grandes lignes de cette étude seront présentées ici.

Si la mise au jour de la fonction structurante des émotions dans un texte est fondée sur les annotations réalisées sous Glozz, il est nécessaire de passer par une étape intermédiaire de réalisation de certaines inférences à partir de ces annotations pour dégager des blocs textuels. Ces inférences s'appuient à la fois sur le rang d'apparition d'un expérienceur dans un texte, sur la catégorie émotionnelle en jeu et sur le rang d'apparition de la phrase concernée dans le texte. Pour mieux montrer au lecteur le processus d'identification des blocs de phrases, une représentation permettant de saisir d'un seul regard à la fois les informations relatives aux unités (frontières, types et traits) et relatives aux relations entre ces unités, ainsi que l'ordre linéaire d'apparition de ces informations a été proposée dans (Battistelli, Étienne, et al., 2022). Cette représentation prend la forme d'une procédure de transformation des annotations répondant à deux objectifs : 1) donner à voir l'ensemble des informations produites par le schéma d'annotation ; 2) rendre lisible le processus inférentiel sous-jacent à la mise en évidence de blocs textuels qui s'appuie sur trois caractéristiques formellement représentées par des indices : le rang d'apparition de la phrase (noté i), le rang d'apparition de l'expérienceur dans le texte (noté j), la catégorie émotionnelle (notée k). Ce sont les liens établis entre ces indices qui permettent la délimitation de blocs de

phrases reliées entre elles par l'expression d'une même émotion (soit une catégorie émotionnelle ressentie par une même entité), de ses causes ou de ses conséquences.

Les conventions d'écriture principales sont les suivantes (se référer à (Battistelli, Étienne, et al., 2022) pour une présentation plus détaillée) :

- chaque unité annotée présente dans une phrase sera notée SitEmo, SitCause ou SitConseq ;
- chacun de ces trois types d'unités est caractérisé par un Experienceur et une catégorie émotionnelle prise dans l'ensemble au moins des onze catégories prévues par le schéma (Étienne & Battistelli, 2021) à savoir l'ensemble : {Co,De,Jo,Pe,Su,Tr,Ad,Cu,Em,Fi,Ja} où COLÈRE correspond à Co, DÉGOÛT à De, JOIE à Jo, PEUR à Pe, SURPRISE à Su, TRISTESSE à Tr, ADMIRATION à Ad, CULPABILITÉ à Cu, EMBARRAS à Em, FIERTÉ à Fi, JALOUSIE à Ja. Les catégories AMOUR notée Am et INDÉTERMINÉE notée In ont été ajoutées[86] ;
- trois indices seront mobilisés, avec $i \in \{1,...,m\}$, $j \in \{1,...,n\}$ et $k \in$ {Co,De,Jo,Pe,Su,Tr,Ad,Cu,Em,Fi,Ja,Am,In} :
  - l'indice i désignera le rang d'apparition d'une phrase dans le texte ;
  - l'indice j désignera le rang d'apparition d'un expérienceur dans un texte ;
  - l'indice k désignera une catégorie émotionnelle ;
  - une situation émotionnelle sera désignée par SitEmo_j_k ;
  - une situation de cause sera désignée par SitCause_j_k ;
  - une situation de conséquence sera désignée par SitConseq_j_k ;
  - un bloc de phrases sera noté Bloc_j_k.

Les exemples [54] et [56] donnent respectivement à voir la transformation des annotations des exemples [53] et [55], où les unités Expérienceur apparaissent surlignées en bleu, les unités SitEmo soulignées et les unités SitCause en gras. Les informations d'annotation sont par ailleurs allégées pour ne faire figurer que les éléments pertinents pour la transformation. Il est à noter que la typologie des modes d'expression retenue (comportemental, désigné, montré, suggéré) n'intervient pas comme critère de décision dans la délimitation de blocs textuels (mais est nécessaire lors des annotations initiales en assurant la prise en compte du plus grand nombre possible d'unités émo-

86 Ces catégories sont ajoutées afin de proposer une analyse plus complète de la dimension émotionnelle des textes Une annotation suivant strictement le schéma les auraient captées avec des unités Autre, excluant ainsi le repérage de leurs expérienceur, cause et conséquence.

tionnelles). L'information du mode d'expression figure donc dans les annotations des exemples mais n'est plus mentionnée pour l'analyse des blocs textuels fondés sur l'expression d'une émotion.

[53] Paul est heureux et Zoé est en colère, et même folle furieuse.

Expérienceur<Paul>{Entité : Paul}

SitEmo<heureux>{Mode : Désignée, Catégorie : Joie, Catégorie2 : Aucune}

Expérienceur<Zoé>{Entité : Zoé}

SitEmo<en colère>{Mode : Désignée, Catégorie : Colère, Catégorie2 : Aucune}

SitEmo<folle furieuse>{Mode : Désignée, Catégorie : Colère, Catégorie2 : Aucune}

Relations Affecte :

<Paul> <--Affecte-- <heureux>

<Zoé> <--Affecte-- <en colère>

<Zoé> <--Affecte-- <folle furieuse>

[54] représentation transformée de [53]

Phrase 1 : SitEmo_1_Jo SitEmo_2_Co SitEmo_2_Co

[55] **Cette histoire** a fait peur à Léo.

SitCause<Cette histoire>

SitEmo<peur>{Mode : Désignée, Catégorie : Peur, Catégorie2 : Aucune}

Experienceur<Léo>{Entité : Léo}

Relation Affecte :

<peur> --Affecte--> <Léo>

Relation CauseDe :

<Cette histoire> --CauseDe--> <peur>

[56] représentation transformée de [55]

Phrase 1 : SitCause_1_Pe SitEmo_1_Pe

La procédure de transformation des annotations constitue le fondement de l'identification de blocs textuels qui peuvent se succéder, être emboîtés ou se chevaucher. Cette démarche est appliquée dans (Battistelli, Étienne, et al., 2022) à deux exemples d'extraits de textes. Chaque extrait est présenté en deux étapes. La première montre le résultat de l'analyse du texte en terme de découpage en phrases et de repérage des quatre types d'unités linguistiques impliquées dans l'expression d'une émotion (SitEmo, SitCause, SitConseq et Experienceur). La seconde étape donne le résultat de la transformation formelle des annotations à partir de laquelle émergent les blocs tex-

tuels calculés en fonction du rang d'apparition d'un expérienceur dans un texte (indice i), de la catégorie émotionnelle en jeu (indice j) et du rang d'apparition de la phrase concernée dans le texte (indice k).

Le premier texte analysé est un texte journalistique issu de l'article « Quelles sont les conséquences du mal-logement ? », *P'tit Libé*, n°91, 1 au 7 février 2019. Ce texte est tout d'abord découpé en phrases et les unités SitEmo, Expérienceur, SitCause et SitConseq y sont repérées, comme illustré par la figure 15. Les mêmes conventions d'identifications sont appliquées ici : les unités SitEmo sont soulignés, les SitCause en gras, les SitConseq en italique et les Expérienceur en surlignage bleu clair.

Quelles sont les conséquences du mal-logement ? /Phrase 1/
«**Certains logements rendent les gens malades au lieu de les protéger**», regrette Manuel Domergue, le directeur des études de la Fondation Abbé-Pierre./Phrase 2/ Les personnes qui vivent dans des habitations insalubres ont plus de risque d'être en mauvaise santé./Phrase 3/ Le plus souvent, elles ont des maladies respiratoires, qui peuvent parfois se transformer en maladies très graves comme des pneumonies./Phrase 4/ «La trop grande humidité abîme les poumons des gens», assure Jean-Baptiste Eyraud, le porte-parole de l'association Droit au logement./Phrase 5/
Certains vivent dans des habitations sans fenêtre, où il y a très peu de lumière et d'aération./Phrase 6/ Leur corps se fragilise et ils tombent plus souvent malades./Phrase 7/ Les victimes du mal-logement souffrent aussi souvent **de graves problèmes psychologiques**. /Phrase 8/«**Quand il y a des grandes fissures dans les murs d'une maison**, ses habitants **ont peur en permanence qu'elle s'écroule**./Phrase 9/ **Etre tout le temps angoissé, ce n'est pas une vie !**» *explique Jean-Baptiste Eyraud, la voix pleine de colère. /Phrase 10/ Dans les familles, cette situation crée des tensions. /Phrase 11/* «**A force d'être les uns sur les autres dans un logement trop petit ou d'être confrontés au froid, on s'énerve et on se dispute sans cesse**», *insiste-t-il*. /Phrase 12/ Quand **des gens habitent dans un logement en très mauvais état,** *c'est aussi plus compliqué d'avoir une vie sociale.*/Phrase 13/ «Ils n'osent plus inviter des amis chez eux car **ils ont honte**./Phrase 14/ Du coup *ils n'osent pas aller chez les autres non plus et se recroquevillent sur eux-mêmes*», affirme Manuel Domergue./Phrase 15/ **C'est aussi très difficile** pour les enfants./Phrase 16/ **Comment grandir dans un logement trop petit pour y faire ses devoirs ou pour jouer et pas assez chauffé pour bien y dormir ?**/Phrase 17/Manuel Domergue explique que «les plus jeunes vont à l'école fatigués, ils ont plus de mal à se concentrer et à apprendre leur leçon»./Phrase 18/

Figure 15: Découpage en phrases et annotations du texte 1

Ce premier texte est donc composé de dix-huit phrases au sein desquelles quatre expérienceurs ont été identifiés. Au total, sept émotions sont associées à ces expérienceurs :

- la TRISTESSE de l'expérienceur 1, notée à l'aide des indices 1 et Tr, soient les étiquettes SitEmo_1_Tr, SitCause_1_Tr et SitConseq_1_Tr ;

- la TRISTESSE de l'expérienceur 2, notée à l'aide des indices 2 et Tr, soient les étiquettes SitEmo_2_Tr, SitCause_2_Tr et SitConseq_2_Tr ;
- la PEUR de l'expérienceur 2, notée à l'aide des indices 2 et Pe, soient les étiquettes SitEmo_2_Pe, SitCause_2_Pe et SitConseq_2_Pe ;
- la COLÈRE de l'expérienceur 2, notée à l'aide des indices 2 et Co, soient les étiquettes SitEmo_2_Co, SitCause_2_Co et SitConseq_2_Co ;
- l'EMBARRAS de l'expérienceur 2, notée à l'aide des indices 2 et Em, soient les étiquettes SitEmo_2_Em, SitCause_2_Em et SitConseq_2_Em ;
- la COLÈRE de l'expérienceur 3, notée à l'aide des indices 3 et Co, soient les étiquettes SitEmo_3_Co, SitCause_3_Co et SitConseq_3_Co ;
- et la TRISTESSE de l'expérienceur 4, notée à l'aide des indices 4 et Tr, soient les étiquettes SitEmo_4_Tr, SitCause_4_Tr et SitConseq_4_Tr.

À l'aide des indices j et k des étiquettes de chaque phrase, des phrases contiguës sont regroupées entre elles. Sont alors déduits à cette étape huit blocs textuels, représentés par des boîtes colorées sur la figure 16. Un code couleur permet à la fois de différencier les différents blocs repérés et de mettre en évidence les étiquettes ayant servi de support à leur délimitation. Chaque couleur symbolise une des sept émotions exprimées dans le texte. Les blocs de la figure 16 correspondent aux regroupements suivants :

- le premier bloc Bloc_1_Tr (boîte verte sur la figure 16) est constitué de la phrase 2, seule phrase à comporter au moins une étiquette d'indices 1 et Tr ;
- le deuxième bloc Bloc_2_Tr (boîte bleu foncé) est constitué de la phrase 8, seule phrase à comporter au moins une étiquette d'indices 2 et Tr ;
- le troisième bloc Bloc_2_Pe (boîte violette) regroupe les phrases 9 à 11 qui comportent chacune au moins une étiquette d'indices 2 et Pe ;
- les quatrième et sixième blocs Bloc_3_Co (boîtes rouges) regroupent respectivement les phrases 9 à 10 et la phrase 12 qui comportent chacune au moins une étiquette d'indices 3 et Co ;
- le cinquième bloc Bloc_2_Co (boîte bleu clair) regroupe les phrases 11 à 12 qui comportent chacune au moins une étiquette d'indices 2 et Co ;

- le sixième bloc Bloc_2_Em (boîte turquoise) regroupe les phrases 13 à 15 qui comportent chacune au moins une étiquette d'indices 2 et Em ;
- le septième bloc Bloc_4_Tr (boîte orange) regroupe les phrases 16 à 17 qui comportent chacune au moins une étiquette d'indices 4 et Tr.

Plusieurs phénomènes sont mis en évidence sur le texte 1. Tout d'abord, des blocs textuels y sont délimités à partir des annotations, mettant ainsi au jour le rôle des émotions dans la structuration d'un texte. Des zones sans émotions formées par les phrases n'appartenant à aucun bloc émergent ainsi. Il apparaît également des chevauchements entre plusieurs émotions, s'expliquant notamment par les relations de causalité entre différentes émotions. La PEUR de l'expérienceur 2 cause par exemple la COLÈRE de l'expérienceur 3, comme l'illustre l'enchâssement de plusieurs étiquettes d'indices 2 et Pe dans l'unité SitCause_3_Co à la phrase 9. Cela illustre bien le lien étroit qu'entretiennent entre elles la dimension émotionnelle et la dimension causale d'un texte (*cf.* chapitre 2) et, de ce fait, l'importance que peuvent revêtir les émotions dans la compréhension d'un texte, puisqu'elles interviennent pour relier sémantiquement plusieurs situations décrites dans un texte.



Figure 16: Résultat de la transformation formelle des annotations du texte 1 et segmentation en boîtes

Le second texte analysé est un extrait du roman pour enfants *Dragonne de minuit*, d'A. Laroche, éd. Rageot. Pour mettre au jour les blocs textuels reposant sur l'expression des émotions qui le composent, la même démarche que celle mise en œuvre sur le texte 1 est appliquée. Dans un premier temps, le texte est donc découpé en phrases et annoté en émotions. Ces annotations sont ensuite transformées afin d'expliciter les liens d'inférences qui les unissent et de faciliter la délimitation de blocs textuels. Seul le découpage en bloc est représenté ici, se reporter à (Battistelli, Étienne, et al., 2022) pour une présentation plus détaillée de l'analyse.

Dans ce texte, composé de trente phrases, ce sont cette fois sept expérienceurs qui sont identifiés. Treize émotions leur sont associées :

- la TRISTESSE de l'expérienceur 1, notée à l'aide des indices 1 et Tr ;
- la PEUR de l'expérienceur 1, notée à l'aide des indices 1 et Pe ;
- la TRISTESSE de l'expérienceur 2, notée à l'aide des indices 2 et Tr ;
- l'ÉMOTION INDÉTERMINÉE de l'expérienceur 2, notée à l'aide des indices 2 et In ;
- l'AMOUR de l'expérienceur 3, notée à l'aide des indices 3 et Am ;
- la JOIE de l'expérienceur 3,notée à l'aide des indices 3 et Jo ;
- la SURPRISE de l'expérienceur 3, notée à l'aide des indices 3 et Su ;
- la PEUR de l'expérienceur 4, notée à l'aide des indices 2 et Pe ;
- l'AMOUR de l'expérienceur 4, notée à l'aide des indices 4 et Am ;
- la COLÈRE de l'expérienceur 4, notée à l'aide des indices 4 et Co ;
- le mélange de PEUR et de TRISTESSE de l'expérienceur 5, noté à l'aide des indices 5 et Pe_Tr ;
- le mélange de TRISTESSE et de PEUR de l'expérienceur 6, noté à l'aide des indices 6 et Tr_Pe ;
- et l'EMBARRAS de l'expérienceur 7, noté à l'aide des indices 7 et Em.

De ces annotations sont déduits quinze blocs textuels, représentés par des boîtes colorées sur la figure 17. Comme pour l'extrait de texte journalistique, un code couleur différencie les blocs repérés et identifie les étiquettes servant de support à leur délimitation. Les blocs identifiés sont les suivants :

- le premier, le quatrième et le onzième blocs Bloc_1_Tr (boîtes noires) regroupent des phrases qui comportent chacune au moins une étiquette d'indices 1 et Tr ;
- le deuxième bloc Bloc_2_Tr (boîte marron foncé) est constitué de la phrase 4 qui comporte au moins une étiquette d'indices 2 et Tr ;
- le troisième bloc Bloc_2_In (boîte marron clair) est constitué de la phrase 5 qui comporte au moins une étiquette d'indices 2 et In ;
- le cinquième bloc Bloc_1_Pe (boîte gris clair) est constitué de la phrase 6 qui comporte au moins une étiquette d'indices 1 et Pe ;
- le sixième bloc Bloc_3_Am (boîte mauve) est constitué de la phrase 7 qui comporte au moins une étiquette d'indices 3 et Am ;
- le septième bloc Bloc_4_Pe (boîte bleu foncé) regroupe des phrases qui comportent chacune au moins une étiquette d'indices 4 et Pe ;
- le huitième bloc Bloc_4_Am (boîte bleu clair) est constitué de la phrase 10 qui comporte au moins une étiquette d'indices 4 et Am ;
- le neuvième bloc Bloc_5_Pe_Tr (boîte verte) est constitué de la phrase 10 qui comporte au moins une étiquette d'indices 5 et Pe_Tr ;
- le dixième bloc Bloc_3_Jo (boîte rose) regroupe des phrases qui comportent chacune au moins une étiquette d'indices 3 et Jo ;
- le douzième bloc Bloc_6_Tr_Pe (boîte orange) est constitué de la phrase 16 qui comporte au moins une étiquette d'indices 6 et Tr_Pe ;
- le treizième bloc Bloc_7_Em (boîte rouge) regroupe des phrases qui comportent chacune au moins une étiquette d'indices 7 et Em ;
- le quatorzième bloc Bloc_4_Co (boîte turquoise) regroupe les phrases 20 à 30 qui comportent chacune au moins une étiquette d'indices 4 et Co ;
- le quinzième bloc Bloc_3_Su (boîte violette) regroupe les phrases 21 à 29 qui comportent chacune au moins une étiquette d'indices 3 et Su.



Figure 17: résultat de la transformation formelle des annotations du texte 2 et segmentation en boîtes

Ce texte de fiction contient une plus grande diversité d'expérienceurs et d'émotions que le texte 1, de nature journalistique. Le même fonctionnement textuel de la dimension émotionnelle que celui observé pour le texte 1 est malgré tout mis en lumière. Des zones sans expression d'émotion, par exemple les phrases 14 à 15 apparaissent ainsi figure 17. Ces zones sont moins étendues que celles qui présentes pour l'analyse du texte 1 (*cf.* figure 16). Les phénomènes de structuration qui se dégagent de l'analyse du texte 2 sont cependant de plus grande ampleur que pour le texte 1. Les blocs textuels mis au jour sont notamment plus étendus : un bloc du texte 2 regroupe par exemple 11 phrases tandis que les blocs du texte 1 ne regroupent au maximum que 3 phrases. Plus de zones de forte densité émotionnelle sont également constatées dans le texte 2 (chevauchements de blocs textuels). Cette densification plus forte de la dimension émotionnelle du texte 2 par rapport au texte 1 est sans doute à mettre en rapport avec le genre de texte auquel il appartient, à savoir le genre fictionnel.

Le travail présenté ici relève de l'observation d'une dimension d'analyse inter-phrastique des émotions qui met au jour des phénomènes de structuration textuelle où des phrases apparaissent sémantiquement liées. Cela est illustré par la représentation de blocs textuels pouvant se succéder, s'imbriquer ou se chevaucher. Plusieurs éléments de méthodologie concourent à la description de ces phénomènes : (i) le principe de s'intéresser aux différents modes d'expression des émotions (dans la mesure où cela conduit nécessairement à une couverture beaucoup plus large des segments textuels potentiellement porteurs d'une information émotionnelle) ; (ii) celui de s'intéresser à la catégorie de l'émotion ressentie par une entité (dans la mesure où l'on peut relier entre eux des segments textuels porteurs d'une référence à une même entité ressentant la même catégorie d'émotion) ; (iii) celui de s'intéresser aux relations de cause et de conséquence qui unissent l'expression d'une émotion à d'autres unités linguistiques (dans la mesure où de fait elles font sortir bien souvent de la phrase d'accueil de l'émotion). La mise en évidence de la fonction structurante des émotions dans un texte présentée ici revêt un aspect novateur du point de vue linguistique. En outre, ce phénomène est mobilisable dans des expérimentations psycholinguistiques en vérifiant par exemple son impact sur la compréhension d'un texte chez l'enfant : la forte imbrication de diverses émotions dans une portion de texte peut par exemple être vue comme un facteur de gêne dans la compréhension de cette portion de texte pour un enfant, en tout cas en dessous d'un certain âge.

3) Interactions entre deux dimensions sémantiques des textes : la temporalité linguistique et les émotions

Le schéma d'annotation que j'ai introduit chapitre 3 comporte une partie, sous forme d'ébauche, visant une exploration systématique de notions liées à l'analyse de la temporalité au sens large[87]. Cette partie prévoit notamment le repérage de marqueurs exprimant la modalité – dont la définition, les catégories et les valeurs sont reprises de (Gosselin, 2005, 2010) – et d'adverbiaux temporels, d'après la modélisation proposée dans (Battistelli et al., 2008) et reprise dans (Teissedre, 2012) (*cf.* chapitres 4 & 5). Si elle n'a été à ce jour pas été exploitée pour l'annotation de textes[88], deux travaux menés dans le cadre de stages financés par le projet TextToKids ont donné lieu à quelques observations préliminaires des interactions entre dimensions TAME et émotions en s'appuyant sur des notions de la partie « TAME » du schéma d'annotation que j'ai proposée. Ces observations ont porté dans un premier temps sur le rôle des adverbiaux temporels dans l'expression des émotions (section 3.1) puis sur les liens entre modalité et émotions (section 3.2). Les observations rapportées ici sont issues du travail de stage de M. Mahamdi (2022) que j'ai co-encadré avec D. Battistelli et dont une partie consacrée à l'étude des adverbiaux temporels constitue elle-même un approfondissement de l'étude menée dans le stage de D. Oltéan (2021).

3.1) Adverbiaux temporels et émotions

Si les liens entre adverbiaux temporels et expression des émotions n'avaient jamais à ce jour été formalisés, plusieurs travaux évoquaient des interactions entre expression de notions temporelles et émotions, sans forcément nommer explicitement les adverbiaux temporels. C. Roméro (2017) mentionne par exemple le rôle des adverbes temporels dans le processus d'intensification d'une émotion, tandis que R. Micheli (2014) cite parmi les critères aptes à construire une situation étayant une émotion la distance temps et espace (*cf.* chapitre 2).

M. Mahamdi (2022) – et avant elle D. Oltéan (2021) – s'est appuyée sur la notion d'adverbial temporel, telle qu'elle est définie et modélisée dans (Battistelli et al., 2008) puis (Teissedre, 2012) (*cf.* chapitre 3) pour étudier de manière systématique le rôle des adverbiaux temporels dans l'expression des émotions. L'étude préliminaire

87 C'est-à-dire au sens où l'entendent par exemple (Battistelli, 2009; Gosselin, 2005, 2010).

88 Des travaux actuellement menés dans le cadre de la thèse de C. Bruneau s'attellent au raffinement du schéma d'annotation de la temporalité et à son application.

menée par D. Oltéan a permis de mettre en évidence le fait que les adverbiaux principalement impliqués dans l'expression des émotions sont ceux qui expriment une durée ou une fréquence. (Oltéan, 2021) distingue ainsi quatre types d'adverbiaux temporels : 1) les « duratifs-itératifs », pouvant être interprétés comme duratif ou itératif selon le contexte, par ex. *toujours, pendant trois jours, depuis deux ans* ; 2) les « dates-durées » pouvant être interprétés comme exprimant une date ponctuelle ou une durée selon le contexte, par ex. *le mois prochain* ; les « purement duratifs » exprimant systématiquement une durée, par ex. *de 1914 à 1918, en l'espace de trois ans ;* 4) les « purement itératifs », ayant systématiquement un sens itératif, par ex. *jamais, souvent, quatre fois.* L'étude de M. Mahamdi s'est donc concentrée sur les adverbiaux duratifs et itératifs en reprenant les quatre types introduits par D. Oltéan. Ce travail est conduit sur l'ensemble du corpus annoté en émotions présenté chapitre 4. Les adverbiaux temporels y ont été repérés automatiquement grâce à un ensemble de transducteurs[89] Unitex développés dans (Teissedre, 2012). Ces transducteurs avaient été complétés dans (Oltéan, 2021) pour permettre un repérage plus exhaustif des adverbiaux duratifs et itératifs, puis affinés dans (Mahamdi, 2022).

Pour observer les liens entre adverbiaux temporels et émotions, l'échelle d'analyse choisie est celle de la phrase graphique. Deux cas possibles se distinguent : celui où l'adverbial se situe en dehors de l'unité émotionnelle, comme dans l'exemple [57] où l'adverbial duratifs *depuis quelques jours* est situé au début de la phrase contenant l'unité émotionnelle qui désigne la JOIE ; et celui où l'adverbial se situe au sein de l'unité émotionnelle, comme dans l'exemple [58] où l'unité émotionnelle qui suggère la TRISTESSE comporte l'adverbial itératif *jamais*.

[57] Depuis quelques jours, Mary Grams, une Canadienne de 84 ans, est heureuse.[90]
[58] Pfff, j'y arriverai jamais, se lamente Elisa[91].

Après avoir appliqué les transducteurs[92] pour le repérage des adverbiaux temporels, toutes les phrases contenant à la fois une unité SitEmo et un adverbial ont été a extraites automatiquement puis stockées dans un tableau, ainsi que toutes les informations d'annotation qui leur sont associées. Chaque adverbial temporel a ensuite été

89 Les transducteurs correspondent à des systèmes symboliques (c'est-à-dire fonctionnant sur des règles) capables de repérer des segments textuels puis de les typer.
90 Exemple extrait de l'article « Carotte miraculeuse », *Albert*, n°21, 1er septembre 2017.
91 Exemple extrait de *La famille trop d'filles – Élisa*, S. Morgenstern.
92 Les transducteurs repèrent les quatre types d'adverbiaux temporels avec une F-mesure de 0,89 sur le corpus FictionM2, de 0,96 sur le corpus WP7Fiction et de 0,89 sur le corpus WP7Encyclo (Mahamdi, 2022).

analysé manuellement afin de déterminer son implication (ou non) dans l'expression de l'émotion. Le processus d'analyse du rôle des adverbiaux temporels, pouvant soit être le déclencheur de l'émotion, soit jouer sur son intensité, soit n'avoir aucune influence sur l'émotion, est illustré par la figure 18, reprise de (Mahamdi, 2022, p. 22).

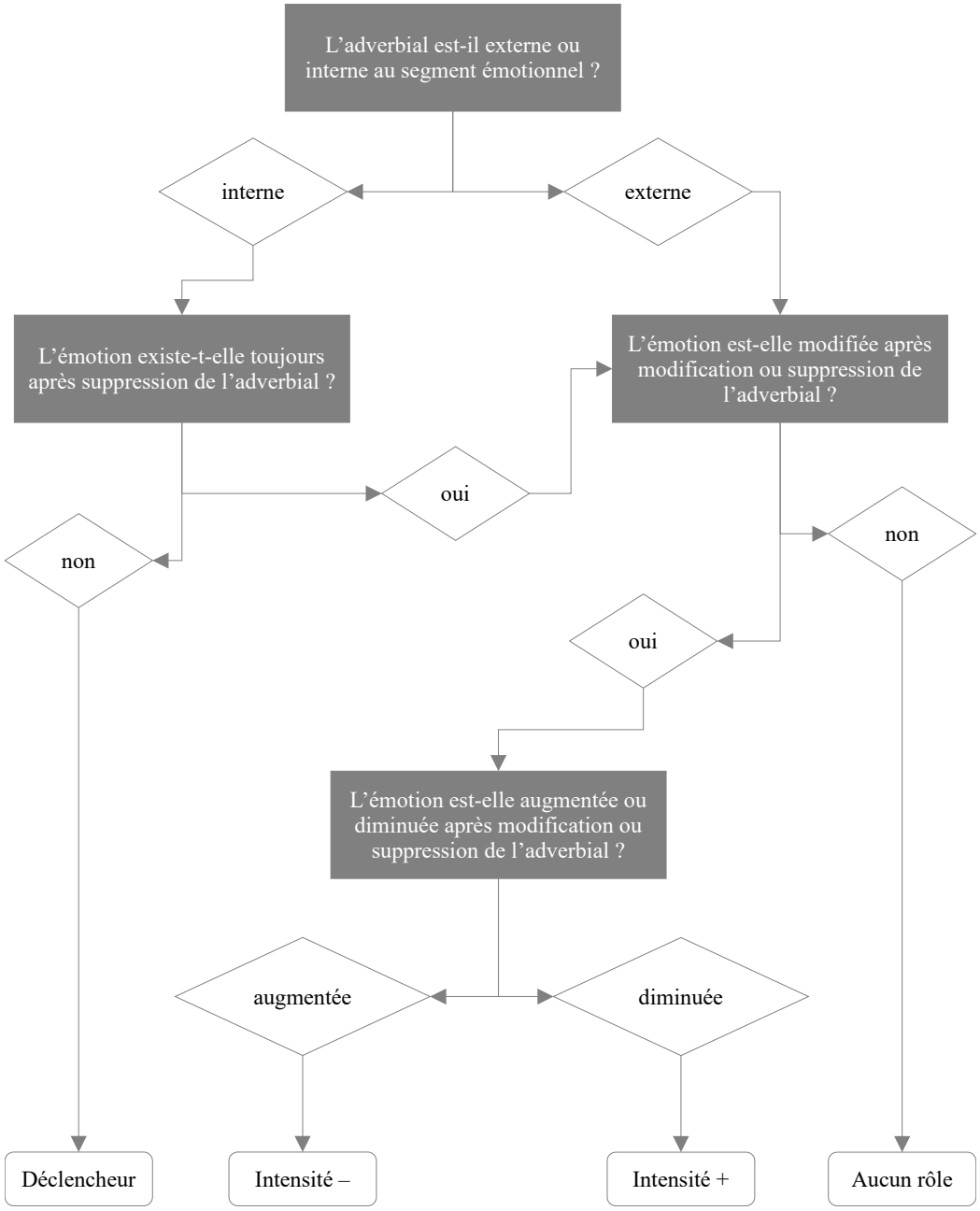


Figure 18: Arbre de décision pour l'analyse du rôle des adverbiaux temporels dans l'expression des émotions (Mahamdi, 2022)

Ainsi, sur les 9 581 adverbiaux temporels repérés dans le corpus, 308 jouent un rôle d'intensifieurs des émotions (soit 3,2 % des adverbiaux repérés par les transducteurs). Parmi eux, 83,8 % sont externes à l'unité émotionnelle et 94,5 % interviennent pour augmenter l'intensité de l'émotion exprimée. 40 adverbiaux (soit 0,4 % des adverbiaux repérés) constituent un déclencheur de l'émotion.

Les unités émotionnelles intensifiées par un adverbial sont principalement des émotions suggérées (37,9 % des unités émotionnelles intensifiées par un adverbial), comportementales (29,2 %) ou désignées (27,5 %). Elles expriment le plus souvent de la COLÈRE (34 % des unités intensifiées), de la PEUR (23 %) ou de la TRISTESSE (14 %), soit trois des émotions les plus fréquentes du corpus (*cf.* chapitre 4). Les adverbiaux temporels de type duratif-itératif et de type purement itératif sont ceux apparaissant le plus souvent comme intensifieurs d'une émotion (respectivement 54,2 % et 36,0 % des adverbiaux intensifieurs). Les cinq adverbiaux intensifieurs les plus fréquents sont *souvent* (25 occurrences, purement itératif), *toujours* (21 occ., duratif-itératif), *jamais* (18 occ., purement itératif), *régulièrement* (13 occ., purement itératif) et *aujourd'hui* (7 occ., date-durée).

Les unités émotionnelles ayant un adverbial comme déclencheur sont, conformément à ce qui était attendu, soit des émotions suggérées (67,5 % des unités), soit des émotions montrées (27,5 %)[93]. Elles expriment principalement la SURPRISE (35 % des unités ayant un adverbial en déclencheur), la JOIE (15 %) et la TRISTESSE (15 %). Les adverbiaux de type duratif-itératif et purement duratif sont également ceux apparaissant le plus fréquemment en déclencheur d'une émotion (tous deux 47,5 % des adverbiaux déclencheurs d'une émotion). Les quatre adverbiaux les plus courants comme déclencheur d'une émotion sont *jamais* (13 occurrences, purement itératif), *toujours* (10 occ., duratif-itératif), *pour la première fois* (2 occ., purement itératif) et *encore* (2 occ., duratif-itératif).

Dans le corpus, l'itération apparaît donc comme la notion temporelle la plus liée à l'expression des émotions, utilisée de manière privilégiée pour intensifier une émotion voire pour la déclencher. Les émotions le plus souvent impliquées dans ces interactions sont les émotions suggérées. Ceci nous a amené à souligner enfin, à propos de la récurrence des adverbiaux *jamais* et *toujours* qu'il semble y avoir une corrélation entre expression d'une quantité temporelle extrême et expression de l'émotion.

93 Il existerait une occurrence d'émotion comportementale ayant un adverbial en déclencheur et une émotion désignée. Ces occurrences ne sont pas relevées dans (Mahamdi, 2022).

M. Mahamdi indique par ailleurs que certains adverbiaux duratifs intensifient l'émotion en indiquant qu'au moment de l'énonciation, l'émotion est toujours en cours. Elle suppose ainsi que « à quantité égale, un adverbial indiquant l'accomplissement du procès donne moins d'intensité à l'émotion qu'un adverbial indiquant l'inaccomplissement » (p. 29) et illustre cette hypothèse par les exemples [59] et [60]. Une proximité plus grande de la situation décrite quand l'émotion est présentée comme toujours en cours permettrait donc d'intensifier l'émotion. Nous retrouvons ici le critère avancé dans (Micheli, 2014) pour l'étayage des émotions.

[59] Il a souffert de dépression pendant trois ans.
[60] Il souffre de dépression depuis trois ans.

À partir des analyses menées sur le corpus, une typologie des adverbiaux intensifieurs des émotions et des adverbiaux déclencheurs des émotions sont proposées. Ces typologies s'appuient sur les notions suivantes :

- la façon d'exprimer l'intensité, *implicite* ou *explicite*. Dans le premier cas, il est nécessaire de faire appel à ses connaissances du monde pour juger du caractère intensificateur (ou non) de la quantité temporelle exprimée (par ex. *depuis une semaine*). Dans le second, des marqueurs linguistiques présents dans l'adverbial orientent l'interprétation (par ex. *pendant ces* ***trop longues*** *années, pendant* ***toutes*** *ces années*) ;
- la quantité temporelle de l'adverbial :
  - *mesurable* à l'aide de valeurs chiffrées (absolue, par ex. *pendant une semaine* ou relative, par ex. *depuis 1964*) ou *non-mesurable* (par ex. *souvent, tout le temps*) ;
  - *simple* (par ex. *pendant une semaine*) ou *complexe*, c'est-à-dire qu'elle apparaît associée à un ou plusieurs procédés d'intensification (par ex. *pendant* ***près de*** *40 ans, pendant* ***trop*** *longtemps*) ;
- le *niveau de la langue* dont relèvent ces procédés d'intensification, tels que le niveau grammatical (par ex. *près de*) ou le niveau lexical (par ex. *longues*) ;
- le caractère systématique du rôle d'intensifieur de l'adverbial ;
- le caractère incertain de l'identification du rôle d'intensifieur de l'adverbial.

Les rôles des adverbiaux temporels dans l'expression des émotions qui émergeaient de travaux linguistiques sont donc ici confirmés. Les adverbiaux interviennent bien pour intensifier l'émotion exprimée (comme suggéré dans (Romero, 2017)) voire pour exprimer l'émotion elle-même (comme ce qui transparaissait dans (Micheli, 2014)).

3.2) Modalité et émotions

Les liens entre modalité et émotions sont manifestés dans plusieurs travaux de linguistique, sans que les terme « modalité » ou « émotions » ne soient forcément explicitement mentionnés (*cf.* chapitre 2). Parmi les marqueurs de modalité cités dans (Gosselin, 2010), certains articulent ainsi clairement une valeur modale et une valeur émotionnelle (par ex. *heureux* pour la modalité appréciative et l'émotion de joie ; *honteux* pour la modalité axiologique et l'émotion d'embarras). Par ailleurs, trois critères intervenant dans (Micheli, 2014) pour définir les émotions étayées (le degré de probabilité des conséquences d'une situation, le degré de maîtrise de cette situation et la signification normative de cette situation) font écho aux modalités épistémique – voire déontique – et axiologique.

Dans le cadre du stage rapporté dans (Mahamdi, 2022), nous avons proposé une piste d'exploration systématique entre marqueurs de modalité et marqueurs d'émotions au sein du corpus annoté du chapitre 4. Comme pour l'unité dédiée à l'annotation de la modalité prévue dans le schéma (*cf.* chapitre 3), la notion de modalité est abordée ici au sens de (Gosselin, 2005, 2010), tant pour sa définition que pour la liste des catégories modales distinguées et de leurs valeurs associées. Les catégories modales retenues ici sont donc : l'aléthique (impossible, contingent, possible, nécessaire), l'épistémique (exclu, contestable, probable, certain), le déontique (interdit, facultatif, permis, obligatoire), le boulique (rejeté, indifférent, désiré), l'appréciatif (indésirable, indifférent, désirable) et l'axiologique (blâmable, indifférent, louable). L'objectif principal du stage de M. Mahamdi était de proposer une méthodologie pour systématiser l'analyse des interactions entre modalité et émotions. Au vu de la difficulté de l'étude de la modalité, notion complexe par la diversité des champs sémantiques qu'elle recouvre et des marqueurs qui la réalisent, deux choix méthodologiques importants ont été réalisés.

Le premier porte sur la fenêtre retenue pour l'analyse du lien entre marqueurs de modalité et marqueurs émotionnels. Trois fenêtres ont été envisagées : celle de la phrase graphique où, comme pour l'étude des adverbiaux temporels, seuls les marqueurs de

modalité apparaissant dans une phrase comportant au moins une unité SitEmo sont observés ; celle de l'unité émotionnelle, où les segments analysés ne correspondent qu'aux marqueurs annotés par une SitEmo ; et celle du déclencheur de l'unité SitEmo, où les segments analysés sont uniquement les tokens renseignés comme déclencheurs au sein de l'unité SitEmo. Dans un premier temps seule l'échelle d'analyse au niveau du déclencheur a été retenue car permettant une analyse plus ciblée.

Le second choix méthodologique porte sur la façon d'aborder l'analyse des marqueurs de modalité. Le procédé retenu consiste à s'appuyer sur un repérage et un typage automatique des marqueurs contenus dans les déclencheurs de SitEmo, puis d'essayer de dégager des tendances du point de vue de l'usage des marqueurs de modalité dans l'expression des émotions par une analyse manuelle. La première étape a donc été de constituer un lexique de marqueurs de modalité afin de permettre une fouille automatique de l'expression de la modalité dans le corpus. Pour cela, chaque marqueur de modalité analysé ou cité dans (Gosselin, 2010, Chapitre 7) a été répertorié, avec sa ou ses catégories modales associées. Deux listes ont ainsi été établies : celle des marqueurs dits de « forme isolée », c'est-à-dire correspondant à un token unique et pouvant donc être aisément repéré dans les déclencheurs ; et celle des marqueurs dit de type « construction », correspondant à des construction impliquant plusieurs tokens, plus difficiles à repérer dans le corpus car nécessitant des prétraitements (analyse syntaxique, *etc.*). Par soucis de praticité, seule la liste des marqueurs de forme isolée a été employée à ce jour. 228 marqueurs de modalité isolés issus de (Gosselin, 2010), ci-après nommés marqueurs de référence, ont ainsi été dégagés. L'intersection entre ces marqueurs de référence et les tokens des déclencheurs des 6 804 SitEmos du corpus a été effectuée automatiquement[94].

Tous les tokens de déclencheurs ne présentant aucune intersection avec un marqueur de référence sont ensuite analysés manuellement afin de chercher de nouveaux marqueurs de modalité, non évoqués dans (Gosselin, 2010), mais répondant aux critères qu'il utilise pour définir la modalité et analyser ses marqueurs. Ce sont ainsi 6 750 tokens qui ont été examinés à la recherche de nouveaux marqueurs. Pour déterminer si le token est un marqueur de modalité ou non, cinq critères ont été retenus :

- la dérivation. Le token correspond à une forme dérivée d'un marqueur de référence ou de la racine du marqueur de référence, par ex. *admiratif* est ajouté car il appartient à la même famille dérivationnelle que *admirable* ;

94 La liste de formes isolées été augmentée automatiquement en formes fléchies à l'aide du Lefff (Sagot, 2010) afin de permettre une couverture plus large : le repérage automatique a en effet été opéré sans lemmatisation.

- la synonymie. Le token est un synonyme d'un marqueur de référence, par ex. *joyeux* est ajouté à la liste des marqueurs de modalité car il entretient une relation de synonymie avec le marqueur de référence *heureux* ;
- le champ lexical. Le token participe de l'expression d'un même champ lexical qu'un marqueur de référence, par ex. le terme *bataille* est ajouté car il participe du même champ lexical que le marqueur de référence *guerre* ;
- la définition lexicographique. La définition du token remplit un certain nombre de critères définitoires, tirés de (Gosselin, 2010) et en lien avec un marqueur de référence, par ex. le terme *bien-être* est ajouté car sa définition lexicographique fait écho à celle du marqueur de référence *satisfait* ;
- un arbre de décision. Un arbre de décision reposant sur des critères et des tests tirés de (Gosselin, 2010) guide l'analyse de tokens identifiés comme marqueur potentiel de modalité mais ne remplissant pas l'un des quatre critères ci-dessus. Cette analyse a par exemple permis d'ajouter les marqueurs *banal* ou *blague*.

Chaque nouveau marqueur ajouté est associé à une catégorie modale (soit la même que celle du marqueur de référence, soit celle découlant de l'analyse conduite à l'aide de l'arbre de décision). 577 marqueurs ont ainsi été ajoutés : 29,6 % à l'aide de l'arbre, 25,3 % par proximité de champ lexical, 19,9 % d'après leur définition lexicographique, 15,4 % par relation de synonymie et 9,7 % par relation de dérivation[95]. Les proportions représentées par les différentes catégories modales dans les ensembles de marqueurs de référence et de marqueurs ajoutés[96] sont présentées tableau 65.

Catégorie modale	**Marqueurs de référence**	**Ajouts**
aléthique	19,7%	10,2%
épistémique	25,4%	13,9%
déontique	8,8%	3,3%
boulique	14,9%	7,6%
appréciatif	32,9%	69,2%
axiologique	19,3%	32,8%

Tableau 65: Proportion des catégories modales pour chaque ensemble de marqueurs

95 307 tokens, identifiés intuitivement comme marqueur potentiel de modalité mais encore non analysés à l'aide de l'arbre de décision n'ont pas encore été associés à une catégorie modale.

96 Sur les 228 marqueurs de référence, 43 articulent l'expression de plusieurs catégories modales. Ce chiffre montent à 196 sur les 577 marqueurs ajoutés.

51 % des marqueurs de référence se trouvent dans un déclencheur de SitEmo (12 % des déclencheurs contiennent l'un de ces marqueurs). Les cinq marqueurs les plus récurrents sont *mal, bien, guerre, fête* et *heureux*. Lorsqu'un déclencheur contient un marqueur de modalité de référence, il s'agit dans 48,5 % des cas d'une émotion suggérée et dans 28,1 % des cas d'une émotion désignée. Les émotions comportementales représentent 11,1 % des cas et les montrées 12,4 %.

Ce travail a permis de souligner que les catégories modales les plus impliquées dans l'expression des émotions à cette échelle d'analyse sont l'appréciatif et l'axiologique, avec des marqueurs comme *heureux* pour l'appréciatif et *bien* ou *mal* pour la combinaison de l'appréciatif et de l'axiologique. La JOIE est l'émotion majoritairement exprimée par un déclencheur en intersection avec un marqueur de modalité. La TRISTESSE représente 16 % des occurrences, la PEUR 15 % et la COLÈRE 14 %. Les liens unissant ces catégories émotionnelles et l'expression de la modalité peuvent être résumés comme suit : « ce qui est désirable et/ou louable rend [...] heureux, et ce qui est indésirable et/ou blâmable est souvent attristant, effrayant ou frustrant » (Mahamdi, 2022, p. 58).

Cette étude exploratoire a permis de constituer des ressources de marqueurs de modalité et d'établir une méthodologie pour l'analyse des interactions entre modalité et émotions. Quelques tendances ont pu être dégagées, notamment la prépondérance des modalités appréciatives et axiologiques dans l'expression des émotions de type JOIE, PEUR, COLÈRE et TRISTESSE, en particulier pour les émotions suggérées et désignées. L'analyse menée mériterait d'être poursuivie, d'une part en approfondissant l'analyse qualitative des intersections entre marqueurs d'émotion et marqueurs de modalité, et d'autre part en élargissant la fenêtre d'analyse à la SitEmo, voire à la phrase graphique.

Conclusion

L'application du schéma d'annotation, résultant en la constitution d'un corpus annoté varié en genres, a permis de conduire un ensemble d'analyses visant à mieux cerner l'expression des émotions dans les textes.

Les modes d'expression des émotions apparaissent ainsi comme un élément clé de la caractérisation des émotions, leur usage variant en fonction de la catégorie émotionnelle exprimée (par ex. la SURPRISE est majoritairement montrée), du genre textuel

considéré (par ex. le mode comportemental est particulièrement présent dans le journalistique et le montré dans le fictionnel) ou de l'auteur des textes (par ex. usage prononcé des émotions comportementales dans le texte de M. Aymé).

Grâce à l'usage de l'ensemble des notions du schéma, la fonction structurante des émotions à l'échelle textuelle peut être mise en évidence, ainsi que les liens forts unissant émotions et causalité. Le rôle fondamental des émotions dans la compréhension de textes est alors à nouveau mis en lumière.

Enfin, l'annotation systématique des unités émotionnelles permet l'exploration de liens entre émotions et dimensions TAME. Les adverbiaux temporels itératifs apparaissent ainsi particulièrement impliqués dans l'intensification des émotions, voire dans leur expression en constituant le déclencheur pour des émotions suggérées notamment. Les modalités appréciatives, et axiologiques dans une moindre mesure, participent pour leur part à l'expression des émotions (particulièrement de JOIE), en intervenant comme déclencheur des unités émotionnelles (surtout suggérées et désignées). Si ces observations restent préliminaires, elles soulignent bien l'interaction de l'expression des émotions et de la temporalité linguistique au sens large. Les notions relatives à l'analyse des dimensions TAME pouvant présenter des points de difficulté pour les plus jeunes (*cf.* chapitre 1), cela ouvre également de nouvelles pistes à explorer pour l'étude de la compréhension chez les enfants.

Conclusion générale

1) Bilan

L'objectif de la présente thèse était de proposer des outils théoriques opératoires pour l'analyse linguistique des émotions, mobilisables pour évaluer le caractère plus ou moins accessible – c'est-à-dire compréhensible – d'un texte pour un enfant.

L'exploration de cette problématique est fondée sur une modélisation des émotions reposant sur les notions de *mode d'expression*, de *type d'émotion* et de *catégorie émotionnelle*, identifiées comme marqueurs de complexité des textes dans la littérature psycholinguistique. Un classifieur à même d'analyser les émotions à l'aide de ces trois notions a été développé puis intégré dans la chaîne de traitement automatique des textes du projet TextToKids sous la forme du processeur EMOTYC. L'application sur corpus de cette modélisation a quant à elle permis de mieux cerner la façon dont les émotions influencent la complexité d'un texte, avec par exemple l'illustration du rôle prépondérant des émotions dans la cohérence d'un texte par la mise au jour de leur fonction structurante à l'échelle textuelle. De fortes interactions entre marqueurs émotionnels et des marqueurs TAME de complexité des textes (plus précisément certains adverbiaux temporels et certains marqueurs de modalité) également pointés dans la littérature psycholinguistique ont par ailleurs été soulignées.

Le présent travail offre cinq contributions principales.

Une modélisation de la notion d'émotion : le schéma d'annotation. La modélisation pluridisciplinaire des émotions que je propose dans cette thèse articule de multiples notions. Certaines, identifiées dans la littérature psycholinguistique, rendent compte d'étapes développementales de la compréhension des textes et des émotions ; d'autres, issues de la linguistique, assurent un typage fin d'une grande hétérogénéité de marqueurs émotionnels ; et d'autres encore sont compatibles avec une automatisation de l'analyse des émotions. Cette modélisation s'actualise sous la forme d'un schéma d'annotation, qui rend son application aisée pour l'analyse de textes. Ce schéma d'annotation est le seul à s'intéresser explicitement à la façon dont les émotions sont exprimées. Quatre modes d'expression des émotions, issus de la conciliation de travaux psycholinguistiques présentés dans (Blanc, 2010) et de travaux linguistiques détaillés dans (Micheli, 2014), y sont définis : le mode désigné, renvoyant à l'usage du

lexique émotionnel ; le mode comportemental, renvoyant à la description de comportements émotionnels ; le mode montré, renvoyant à la monstration des émotions de l'énonciateur par le biais de caractéristiques diverses des énoncés ; et le mode suggéré, renvoyant à la description de situations associées de manière conventionnelle à des ressentis émotionnels. Ce schéma repose sur quatre unités linguistiques pour étudier l'expression des émotions : l'unité émotionnelle à proprement parler, l'unité référant à l'expérienceur de l'émotion, l'unité exprimant sa cause et celle dénotant sa conséquence. Une analyse fine et précise de la dimension émotionnelle des textes est garantie par des systèmes de traits spécifiant les unités, par exemple la catégorie émotionnelle, le type d'émotion et le mode d'expression pour les unités émotionnelles.

Un corpus annoté en émotions. Le schéma d'annotation a été appliqué manuellement par six expertes en linguistique *via* la plateforme Glozz (Widlöcher & Mathet, 2012) sur un corpus de 1 594 textes pour enfants, de genres et sources variés. Plus de 4 000 phrases émotionnelles ont ainsi été repérées, parmi lesquelles 6 804 unités émotionnelles ont été délimitées, de l'échelle du caractère (par ex. marqueur typographique montrant une émotion) à celle de la proposition (par ex. situation suggérant une émotion).

Un système d'analyse automatique des émotions. Des classifieurs ont été entraînés à l'aide de techniques d'apprentissage automatique profond (ajustement de CamemBERT) à partir du corpus annoté. Un classifieur à même de prédire, pour une phrase donnée, son caractère émotionnel ainsi que les catégories émotionnelles, les types d'émotion et les modes d'expression qu'elle contient a ainsi été développé. Ce classifieur, capable d'analyser des phrases selon certaines notions de mon schéma, est intégré dans la chaîne de traitement automatique des textes du projet TextToKids par le biais de EMOTYC pour extraire des descripteurs émotionnels de complexité. Ce processeur est testable en ligne à l'adresse suivante : http://vheborto-corliapi.inist.fr:8984/. Il vient compléter l'analyse des descripteurs de complexité émotionnels offerte par le processeur ÉMOTIONS, présent dès la version initiale de la chaîne TextToKids, qui s'appuie uniquement sur des lexiques figés. Le classifieur sous-jacent à EMOTYC retrouve efficacement de nombreuses informations émotionnelles, par exemple le caractère émotionnel d'une phrase (F1 0,74), le mode désigné (F1 0,80), le type d'émotion « de base » (F1 0,71) ou encore la surprise (F1 0,71) et la peur (F1 0,67).

Un ensemble de marqueurs de complexité des textes. Ma problématique de thèse s'insère dans le cadre plus large du projet ANR TextToKids dédié à l'étude de la complexité des textes destinés aux enfants jeunes lecteurs. Dans ce cadre, j'ai identifié 64 marqueurs de complexité des textes, dont la majorité (soit 39) sont dégagés de la littérature psycholinguistique (tandis que 17 le sont de la lisibilité ou de la simplification de textes et 8 sont exploratoires, c'est-à-dire intuitivement jugés pertinents pour l'analyse de la complexité). Ces marqueurs issus de travaux psycholinguistiques reflètent des étapes développementales. Ils sont alors associés à des âges d'acquisition, offrant ainsi une analyse fine de la complexité des textes. Par ailleurs, il est à noter que parmi les 64 marqueurs, 25 sont d'ordre sémantique, niveau linguistique usuellement peu pris en compte dans les travaux de lisibilité ou de simplification. De ce fait, la chaîne TextToKids qui permet d'extraire automatiquement des descripteurs traduisant ces marqueurs se démarque dans le paysage scientifique actuel. Parmi ces marqueurs sémantiques figurent les émotions qui ont fait l'objet dans cette thèse du développement d'un modèle d'analyse lui aussi original, notamment de par le repérage qu'il fait d'émotions exprimées par d'autres moyens que le lexique émotionnel.

Un approfondissement de la connaissance linguistique de l'expression des émotions. Les observations linguistiques majeures effectuées à l'aide du schéma d'annotation portent sur l'usage des modes d'expression des émotions – influencé aussi bien par la catégorie émotionnelle exprimée que par le genre de texte considéré –, sur les interactions entre dimension émotionnelle et dimensions TAME des textes, ainsi que sur le rôle structurant des émotions à l'échelle textuelle. Si l'intérêt que présentent ces observations pour l'analyse de la complexité a été précédemment mentionné, il est important de souligner qu'elles revêtent également un intérêt purement linguistique. L'approche proposée ici offre une exploration de la dimension émotionnelle dans la lignée de travaux tels que (Micheli, 2014).

Ces contributions ont donné lieu à des publications dont la liste est indiquée dans le chapitre « Publications personnelles » du manuscrit.

2) Perspectives

Du travail mené dans la présente thèse se dégagent des pistes d'approfondissement pour l'étude de l'expression des émotions mais aussi de la complexité des textes. Des perspectives d'application à d'autres domaines ou cas d'usage du schéma d'annotation et du classifieur en émotions élaborés peuvent également être considérées.

2.1) Approfondissement de l'étude de l'expression des émotions

Plusieurs pistes d'analyse de l'expression des émotions des personnages des textes restent à encore ce jour inexploitées.

En premier lieu, l'exploration des données annotées en émotions mériterait d'être approfondie, notamment du point de vue des segments textuels captés par les unités émotionnelles, en particulier de leurs déclencheurs. En appliquant des techniques simples telles que la lemmatisation ou la racinisation (aussi appelée stemmatisation), des groupements de segments textuels ou déclencheurs similaires émergeraient, plus adéquats que ceux présentés dans la section 4 du chapitre 4. Ces groupements donneraient lieu à des observations plus généralisables sur les marqueurs linguistiques des émotions (analysables par exemple selon les catégories émotionnelles ou les modes d'expression). Ils permettraient aussi l'élaboration de ressources lexicales, comme des lexiques émotionnels ou des lexiques de marqueurs relevant de différents modes d'expression.

Il serait par ailleurs intéressant d'appliquer le schéma d'annotation sur des corpus de textes différents de celui employé ici, par exemple sur des textes issus de l'oral, des tweets ou commentaires Internet, des textes pour adultes ou même des textes dans d'autres langues pour voir si les observations relatives à l'expression des émotions effectuées sur le corpus de ma thèse se confirment sur d'autres données.

Ce travail de thèse souligne également la nécessité d'améliorer le guide d'annotation (Étienne & Battistelli, 2021) qui accompagne le schéma, en y précisant et affinant encore les critères régissant l'annotation manuelle en émotions. Une première modification consisterait à intégrer le raffinement de l'annotation des causes et des conséquences de l'émotion introduit dans (Battistelli, Étienne, et al., 2022). Une autre piste d'amélioration se dégage des résultats de l'expérience perceptuelle présentée à la section 6 du chapitre 6, qui mettent en évidence quelques irrégularités dans les anno-

tations humaines, aisément corrigeables avec l'édition de critères plus rigoureux dans le guide.

De manière générale, les résultats de cette expérience perceptuelle gagneraient à être exploités plus en amont afin de mieux cerner les forces et faiblesses de l'annotation humaine et des prédictions du classifieur. Cela permettrait, idéalement, de mieux circonscrire les cas complexes irréductibles où plusieurs analyses (en catégories émotionnelles, en mode d'expression) sont de fait possibles.

Les expérimentations informatiques menées à partir du corpus annoté offrent pour leur part plusieurs pistes de continuation. Une amélioration des performances du classifieur en catégories émotionnelles, types d'émotion et modes d'expression serait peut-être constatée en injectant des descripteurs experts (par ex. termes d'un lexique émotionnel) dans la représentation des phrases fournies en entrée au modèle. Des systèmes capables de prédire d'autres notions du schéma, par exemple les déclencheurs des unités émotionnelles ou les expérienceurs, pourraient aussi être développés. Une analyse approfondie du consensus sur l'annotation de ces notions devrait alors être mise en place par le biais d'un calcul d'accord inter-annotateurs. Une autre tâche de détection des émotions, particulièrement utile pour l'analyse de la complexité linguistique, consisterait enfin à descendre en dessous du niveau de la phrase et à développer des modèles à même d'identifier et de typer les marqueurs linguistiques des émotions, comme cela a pu être proposé dans (Kim & Klinger, 2018). Cette perspective est en phrase avec la tendance actuelle visant la prédiction non plus seulement de *classes* mais de *structures* par des modèles dits génératifs (Bevilacqua et al., 2021).

Notons enfin que la partie du schéma d'annotation dédiée à l'analyse des dimensions TAME, esquissée chapitre 3, section 3, doit être précisée pour garantir une étude plus systématique de ces dimensions et de leurs interactions avec les émotions des personnages. Une thèse de doctorat, conduite par C. Bruneau sous la direction de D. Battistelli et de J. Schuh au laboratoire MoDyCo, s'attelle actuellement à cette mission.

2.2) Approfondissement de l'étude de la notion de complexité

L'étude de la notion de complexité n'a été qu'effleurée ici et de nombreuses pistes d'approfondissement sont envisageables.

Beaucoup d'observations reposant sur l'usage de la chaîne d'extraction des descripteurs de complexité sont encore à mener. L'analyse des corrélations entre âges édi-

teurs et descripteurs linguistiques présentée chapitre 2, section 3 mériterait d'être affinée, en mobilisant un corpus de textes plus étendu ou en appliquant la chaîne sur des corpus d'autres genres (encyclopédique et journalistique au lieu de fictionnel). De manière plus générale, il paraît intéressant de sonder les corrélations observables entre descripteurs linguistiques et genres de textes, sans nécessairement prendre en compte des distinctions d'âges. Appliquée sur le matériel expérimental du WP7, la chaîne mettrait sans doute en lumière des liens entre les descripteurs de complexité extraits par la chaîne et les résultats des expériences conduites auprès des enfants (la « réalité terrain »).

Les possibilités offertes par le classifieur en émotions développé dans cette thèse n'ont par ailleurs été que peu prospectées à ce jour. L'analyse produite par le classifieur constituerait sans nul doute une base précieuse, exploitable dans l'élaboration d'un matériel expérimental contrôlé et minutieux pour sonder l'impact de critères linguistiques précis dans les processus de compréhension. Les résultats du classifieur serviraient aussi certainement le développement de modèles d'analyse automatique de la complexité des textes. Les prédictions du classifieur, considérées comme des descripteurs de textes, pourraient en effet être sollicitées comme descripteurs experts lors de l'apprentissage d'un modèle de prédiction de la complexité des textes (voir les expériences de ce type menées dans (Blandin et al., 2020)) ou lors d'une démarche d'explicabilité des résultats d'un tel modèle. L'analyse réalisée par le classifieur gagnerait enfin à être prise en compte dans une démarche de simplification (automatique ou manuelle) des textes et dans la constitution de recommandations pour la rédaction de textes à destination des enfants. Il est à noter que l'ensemble des perspectives d'usage de l'analyse conduite par le classifieur s'applique également à celle opérée par la chaîne d'extraction de descripteurs développée dans TextToKids.

2.3) Application à d'autres domaines

Les pistes d'approfondissement présentées sections 2.1 et 2.2 concernent directement les problématiques de ma thèse et du projet TextToKids, à savoir la caractérisation (psycho-)linguistique de l'expression des émotions et l'analyse de la complexité des textes. Le schéma d'annotation des émotions et le classifieur en émotions introduits offrent cependant des perspectives intéressantes pour d'autres domaines.

Le schéma, et en particulier le classifieur, bénéficieraient certainement à des champs de recherche proches de celui de la détection des émotions, telles que la fouille d'opi-

nions ou l'analyse de sentiments. La notion de mode d'expression paraît particulièrement pertinente pour ces domaines qui, bien souvent, limitent leur analyse au lexique subjectif.

Les différents savoirs linguistiques sur l'expression des émotions mis au jour par les annotations, ainsi que les ressources pouvant en être extraites (*cf.* section 2.1) semblent quant à eux offrir des perspectives pour la génération automatique de textes comportant des contenus émotionnels. Intégrés directement dans la construction des modèles ou mobilisés comme critères pour l'évaluation des productions, les savoirs et ressources issus de ma thèse pourraient en effet améliorer le « naturel » de textes émotionnels générés automatiquement (par ex. la surprise étant principalement montrée, un texte voulant dénoter cette émotion devrait plutôt utiliser le mode montré).

Enfin, l'application du schéma à d'autres corpus évoquée section 2.1 servirait également la mise en évidence de mécanismes dépassant l'étude du fonctionnement linguistique des émotions. Le schéma d'annotation a par exemple déjà été employé dans le projet ANR Flyer pour caractériser des contenus extrémistes (Dragos et al., 2022). Cette étude mériterait d'être approfondie pour affiner l'exploration des liens entre ces types de contenus et l'expression des émotions. De manière plus générale, une fois encore de par l'usage de la notion de mode d'expression, le schéma semble proposer des pistes intéressantes pour la caractérisation – voire la détection – de contenus haineux (racistes, sexistes, *etc.*), tâche dont les enjeux et applications sont nombreux aujourd'hui. La thèse actuellement menée par C. Bruneau utilise quant à elle le schéma d'annotation des émotions pour mieux cerner les ressorts linguistiques du processus de transmission de valeurs au sein d'un corpus de manuels d'histoire. Enfin, dans la lignée des études conduites à l'aide du LIWC rapportées dans (Tausczik & Pennebaker, 2010), les annotations produites à l'aide du schéma pourraient intervenir dans des protocoles d'études en psychologie visant à établir des liens entre l'usage du langage émotionnel et l'état psychologique du scripteur.

Annexe I Livret individuel

LIVRET INDIVIDUEL



Pour mieux te connaître :

Tu es en classe de :

- ☐ CE2
- ☐ CM2

Ton âge : ___ ans

Ta date de naissance : ___ /_____/_____

Tu es :

- ☐ une fille
- ☐ un garçon



CONSIGNES

Dans ce livret, tu vas trouver neuf textes que nous avons sélectionnés et pour lesquels nous sommes curieux d'avoir ton avis. Trois séances de lecture seront organisées. A chaque séance, tu auras trois textes à lire. Aujourd'hui a lieu ta première séance de lecture.

CE N'EST PAS UN EXERCICE, CE N'EST PAS UNE ÉVALUATION !!!

En même temps que tu liras un texte, tu entendras ce texte lu par une autre personne. Cette procédure sera répétée une seconde fois tout de suite après. Tu auras donc lu (et entendu) deux fois le même texte. Ensuite, nous te proposerons à chaque fois un ensemble de questions. Pour chacune d'entre elles, ta mission sera de dire si la réponse est VRAI ou FAUX par rapport à ce que tu as lu et compris du texte. Pour la dernière question, tu auras juste à cocher une case.

ATTENTION, pour répondre aux questions, tu ne pourras pas revenir au texte pour prendre ta décision, il faudra que tu décides d'après ce dont tu te souviens du texte.

Ton avis est très important pour nous, donc ne regarde pas sur ton voisin ou ta voisine, puisque c'est ton avis qui nous intéresse. Avant de commencer, écoute bien ce que va expliquer ton maître ou ta maîtresse.

MERCI POUR TA PARTICIPATION !

SÉANCE DE LECTURE 1

LA RÉDACTION

À l'école, on a fait une rédaction sur Noël. Monsieur Languepin a écrit au tableau : "Voici Noël ! Décrivez vos impressions".

On a tous fait des yeux ronds. Qu'est-ce que c'est, des impressions ?

– Vous n'avez qu'à parler du sapin, des vitrines, a expliqué monsieur Languepin. Décrivez les jouets dont vous rêvez, montrez comment Noël…

On n'écoutait déjà plus. On se racontait ce qu'on allait demander et on avait plein d'idées.

- Un peu de silence, a dit notre maître, on n'écrit pas avec sa langue.

Moi, j'ai décidé que j'allais faire ma poésie de Noël. Mais j'avais le trac. Peut-être que je ne sais pas du tout faire de la poésie ? Peut-être que tout le monde va se moquer de moi ? J'ai écrit ça :

> Noël, c'est comme une grande lumière dans une forêt très noire.
>
> Noël, c'est comme un oiseau qui chante après la pluie.
>
> Noël, c'est comme les guerres quand elles s'arrêtent.
>
> Pourquoi ce n'est pas Noël toute la vie ?

Ça m'a donné du mal, ma poésie. J'avais les joues rouges de chaleur. Mais je trouvais que c'était beau ce que j'avais écrit. À la récré, tout le monde se demandait :

- Et toi, qu'est-ce que tu as mis ?

Moi, je n'ai rien dit. J'avais honte. Je suis le seul à avoir écrit une poésie.

Le lendemain, monsieur Languepin a sorti nos cahiers de classe de son cartable. J'ai vu le mien dessus. Mon cœur s'est mis à battre très fort.

- J'ai lu vos rédactions, a dit le maître. Vous faites beaucoup trop de fautes ! Parfois, on ne comprend même pas ce que vous avez écrit. Mais il y en a qui m'ont fait de jolies choses. Tristan, par exemple…

Je suis devenu tout chaud comme si j'avais de la fièvre et monsieur Languepin a lu ma poésie. Il l'a lue très bien, avec le ton. Tout le monde a dit que c'était super beau.

QUESTIONS

Q1. Le maître demande à ses élèves de faire une rédaction avant l'arrivée des fêtes de Noël.

☐ **VRAI** ☐ **FAUX**

Q2. Quand les élèves commencent le travail de rédaction demandé par leur maître, ils se sentent très inquiets.

☐ **VRAI** ☐ **FAUX**

Q3. Quand Tristan se donne du mal avec sa rédaction, il se sent en colère.

☐ **VRAI** ☐ **FAUX**

Q4. Tristan et ses camarades ont réalisé l'exercice de rédaction avant la récréation.

☐ **VRAI** ☐ **FAUX**

Q5. Quand à la récré Tristan ne dit pas à ses camarades qu'il a écrit une poésie, il se sent honteux.

☐ **VRAI** ☐ **FAUX**

Q6. Le maître a rendu les rédactions aux élèves une semaine plus tard.

☐ **VRAI** ☐ **FAUX**

Q7. Le maître n'a pas pu corriger toutes les rédactions des élèves qu'il avait rangées dans son cartable.

☐ **VRAI** ☐ **FAUX**

Q8. Le maître a lu la poésie de Noël de Tristan dès la fin de l'exercice.

☐ **VRAI** ☐ **FAUX**

Q9. Quand le maître lit à la classe la poésie que Tristan a écrit, Tristan se sent fier.

☐ **VRAI** ☐ **FAUX**

Q10. La rédaction sur Noël inspire Tristan qui va écrire sa première poésie.

☐ **VRAI** ☐ **FAUX**

D'après toi, ce texte est tiré :

- ☐ d'une publicité
- ☐ d'un livre de recettes de cuisine
- ☐ d'une encyclopédie en ligne (comme vikidia ou wikipedia)
- ☐ d'un roman
- ☐ d'un journal d'informations

LES ORIGINES DE TINTIN

*"*Les Aventures de Tintin*"* est une série de bandes dessinées créée en 1929 par le dessinateur belge Hergé (de son vrai nom Georges Remi).

Tintin est un journaliste travaillant pour "Le Petit Vingtième". Avec son fox-terrier Milou, un chien blanc non dénué d'humour et d'intelligence, il est envoyé dans les zones sensibles du globe.

Le capitaine Haddock, dont le prénom est « Archibald », est le meilleur ami de Tintin.

Ce marin, contrairement à l'apparence relativement neutre de Tintin (semblable à son caractère), a un physique très reconnaissable : souvent une pipe à la bouche, doté d'une barbe noire, vêtu d'un pull bleu avec une ancre dessus et coiffé d'une casquette de marin.

Il est l'exact inverse de Tintin, car plein de défauts : colérique, alcoolique, têtu et pantouflard ; ses faiblesses font de lui un personnage très "humain" et haut en couleurs. Comme Milou, ses défauts sont compensés par de grandes qualités : il est ainsi loyal, courageux et doué d'une réelle bonté.

Il est surtout connu pour la variété de ses nombreuses insultes, qu'il utilise lorsqu'il est en colère (autrement dit très souvent). Les plus célèbres sont "Moule à gaufres !", "Bachi-bouzouk !", "Tonnerre de Brest !" et surtout "Mille millions de mille sabords !".

Hergé raconte ainsi la naissance de cette idée d'insultes imagées :

> "Ça remonte à l'époque où venait d'être conclu un accord politique entre quatre puissances que l'on appelait "le Pacte à quatre". Et j'avais assisté à une querelle entre une marchande et une de ses clientes qui avait sans doute discuté d'un prix ou critiqué la qualité de sa marchandise. Toujours est-il que la marchande, probablement à bout d'arguments, lui avait lancé derrière sa charrette "Espèce de pacte-à-quatre que vous êtes !...". Ça m'avait paru d'une efficacité surprenante !"

Le capitaine Haddock apparaît pour la première fois dans "Le Crabe aux pinces d'or*"* (1940-1941), Hergé ne se doute pas de l'immense importance que va connaître ce personnage, qui reviendra dans tous les albums suivants et y joue la plupart du temps un rôle au moins aussi primordial que Tintin, devenant de ce fait le troisième personnage principal de la série.

QUESTIONS

Q1. La première bande dessinée de Tintin est sortie il y a une dizaine d'années.
☐ **VRAI** ☐ **FAUX**

Q2. Dans la bande dessinée qui raconte les aventures de Tintin, Tintin est un journaliste qui a un chien blanc prénommé Milou.
☐ **VRAI** ☐ **FAUX**

Q3. Quand le Capitaine Haddock dit des insultes, il est en colère.
☐ **VRAI** ☐ **FAUX**

Q4. Hergé a imaginé les insultes du capitaine Haddock à l'époque du pacte à quatre.
☐ **VRAI** ☐ **FAUX**

Q5. Quand la marchande dit « espèce de pacte-à-quatre que vous êtes ! » à la cliente, elle se sent fière.
☐ **VRAI** ☐ **FAUX**

Q6. Quand Hergé entend la marchande dire « espèce de pacte-à-quatre que vous êtes ! » à une de ses clientes, il ressent de la surprise.
☐ **VRAI** ☐ **FAUX**

Q7. Quand le personnage du Capitaine Haddock connaît un franc succès, Hergé n'est pas surpris.
☐ **VRAI** ☐ **FAUX**

Q8. Le capitaine Haddock était présent dès le premier album de Tintin.
☐ **VRAI** ☐ **FAUX**

Q9. Après sa première apparition dans la bande dessinée, le Capitaine Haddock sera toujours présent dans les histoires de Tintin.
☐ **VRAI** ☐ **FAUX**

Q10. Les aventures de Tintin ont été imaginées par un dessinateur français.
☐ **VRAI** ☐ **FAUX**

D'après toi, ce texte est tiré :

- ☐ d'un livre de recettes de cuisine
- ☐ d'un journal d'informations
- ☐ d'une publicité
- ☐ d'une encyclopédie en ligne (comme vikidia ou wikipedia)
- ☐ d'un roman

L'IMMEUBLE D'INES ET D'ADAM

Dans l'appartement d'Inès, 4 ans, et Adam, 2 ans, il y a des dizaines de photos de famille sur les murs. Des coussins moelleux sont disposés sur le canapé, et les enfants regardent des dessins animés à la télé. Rien d'anormal ! Enfin presque … Cet appartement situé à Coulommiers, une ville à l'est de Paris, n'est pas vraiment le leur. Ils y ont emménagé en catastrophe avec leurs parents et leurs deux chats, la veille de Noël. L'immeuble où ils vivaient juste avant menaçait de s'effondrer… La famille était dans une situation de mal-logement.

"Il y a deux ans et demi, des fissures ont commencé à apparaître sur les murs de notre immeuble. Petit à petit, elles ont grossi. A la fin, on pouvait passer nos bras à l'intérieur !", raconte Jennifer, la maman d'Inès et Adam. A cause de ces trous, il faisait toujours très froid et humide chez eux. Adam tombait souvent malade : "On l'emmenait trois fois par mois chez le pédiatre", affirme Jalil, son papa.

Inès aimait beaucoup sa chambre "violette et blanche avec des rideaux roses". Mais elle se souvient des murs qui n'arrêtaient pas de faire "crac". Ça lui faisait très peur. "La maison allait tomber sur nous …", murmure-t-elle en se blottissant sur le canapé. Pour protéger leurs enfants, Jennifer et Jalil ont fini par les faire dormir chez leur grand-mère.

Les habitants de l'immeuble ont demandé des dizaines, voire peut-être des centaines de fois, aux propriétaires de faire des travaux, sans succès. Cet automne, Jennifer a appelé l'association Droit au logement, qui aide les personnes dans ce genre de situation. Une bénévole a alors tout mis en œuvre pour les sortir de là.

La maire de Coulommiers leur a trouvé un appartement de secours, dans lequel ils peuvent vivre gratuitement en attendant de trouver une solution. Même si leur combat pour trouver un nouveau logement est loin d'être terminé, toute la famille est déjà un peu soulagée. Inès la première :

"Maintenant, la nuit, on a bien chaud", se réjouit-elle.

QUESTIONS

Q1. La famille s'est installée dans le logement de secours seulement la veille de Noël.
☐ **VRAI** ☐ **FAUX**

Q2. Les habitants ont signalé que l'immeuble s'était dégradé en quelques mois seulement.
☐ **VRAI** ☐ **FAUX**

Q3. Le petit frère d'Inès tombait souvent malade quand ils vivaient dans leur ancien logement.
☐ **VRAI** ☐ **FAUX**

Q4. Quand les murs de l'appartement faisaient crac, Inès avait peur.
☐ **VRAI** ☐ **FAUX**

Q5. Quand ils décident de faire dormir leurs enfants chez leur grand-mère, les parents d'Inès et Adam sont joyeux.
☐ **VRAI** ☐ **FAUX**

Q6. La maman d'Inès a fait appel à l'association Droit au logement pendant l'été.
☐ **VRAI** ☐ **FAUX**

Q7. Quand la bénévole de l'association Droit au logement est saisie de leur dossier, elle se sent impuissante.
☐ **VRAI** ☐ **FAUX**

Q8. Quand le maire a trouvé un appartement de secours, la famille s'est sentie soulagée.
☐ **VRAI** ☐ **FAUX**

Q9. Les familles qui sont en situation de mal-logement peuvent faire appel à l'association Droit au logement pour obtenir de l'aide.
☐ **VRAI** ☐ **FAUX**

Q10. Le mal-logement est sans conséquence sur la santé des habitants.
☐ **VRAI** ☐ **FAUX**

D'après toi, ce texte est tiré :

- ☐ d'une publicité
- ☐ d'un journal d'informations
- ☐ d'une encyclopédie en ligne (comme vikidia ou wikipedia)
- ☐ d'un roman
- ☐ d'un livre de recettes de cuisine

SÉANCE DE LECTURE 2

LE MYSTÈRE DES DINOSAURES

Une équipe de scientifiques vient de révéler que certains dinosaures ne ressemblaient pas du tout à l'image que nous nous en faisons. En plus du T-rex et du diplodocus, on trouve aussi le psittacosaurus, qui n'a rien des grands méchants dinosaures mis en scène dans la plupart des films.

Si on vous dit dinosaure, à quoi pensez-vous ? Beaucoup d'entre nous imaginent une sorte de lézard géant, avec des dents pointues et des griffes acérées, comme dans Jurassic Park. Mais des scientifiques viennent de découvrir que ce n'était pas toujours le cas. Une équipe de chercheurs de l'université de Bristol, en Grande-Bretagne, a travaillé sur le psittacosaurus, un dinosaure qui vivait il y a plus de 100 millions d'années.

À partir d'un fossile particulièrement bien conservé découvert en Chine, ils ont pu reconstituer très fidèlement un modèle de psittacosaurus grandeur nature. Et il se trouve qu'il n'était pas aussi impressionnant que Godzilla ! En réalité, la plupart des dinosaures (et le psittacosaurus en fait partie) n'étaient pas couverts d'écailles mais de plumes. Ils sont donc les ancêtres des oiseaux que nous connaissons. Sur les fossiles, on voit clairement les traces de ces plumes.

D'autre part, le psittacosaurus est loin d'être immense : il faisait la taille d'une grosse dinde. Ce dinosaure possédait aussi un énorme bec, qui ressemblait à ceux des perroquets. C'est pourquoi il a été surnommé le "dinosaure-perroquet". De chaque côté de sa tête, il avait de gros yeux ronds et des cornes au niveau des tempes. Sur sa queue, il possédait aussi des plumes, hérissées comme les poils d'une brosse à dents. Personne jusqu'ici n'a réussi à déterminer à quoi ces plumes pouvaient bien lui servir.

Enfin, les scientifiques ont également réussi un exploit, celui de retrouver de quelle couleur était réellement ce dinosaure. À l'aide d'un petit bout de peau prélevé sur le fossile, ils ont pu déterminer que le psittacosaurus était marron, avec de petites tâches jaunes sur les joues et des bandes foncées sur le devant des pattes.

À partir de ces éléments, ils ont déduit que le psittacosaurus vivait plutôt dans la forêt. Cette étude va maintenant pouvoir être menée sur d'autres fossiles et, peut-être, révéler un jour les secrets d'autres espèces de dinosaures.

QUESTIONS

Q1. Les dinosaures mis en scène dans les films ne ressemblent pas à ce que la science a pu découvrir.

☐ **VRAI** ☐ **FAUX**

Q2. Quand le dinosaure est représenté avec des dents pointues et des griffes acérées, c'est pour faire peur aux spectateurs dans les films.

☐ **VRAI** ☐ **FAUX**

Q4. Le psittacosaurus avait déjà disparu il y a plus de 100 millions d'années.

☐ **VRAI** ☐ **FAUX**

Q3. Quand les scientifiques ont découvert un fossile de psittacosaurus en Chine, ils ont été très déçus.

☐ **VRAI** ☐ **FAUX**

Q5. Quand les scientifiques ont réalisé que le psittacosaurus n'était pas plus grand qu'une grosse dinde, ils étaient furieux.

☐ **VRAI** ☐ **FAUX**

Q6. Les psittacosaurus ont été surnommés "dinosaures-perroquets" pour la diversité des couleurs de leur plumage.

☐ **VRAI** ☐ **FAUX**

Q7. On sait depuis longtemps à quoi servent les plumes sur la queue du psittacosaurus.

☐ **VRAI** ☐ **FAUX**

Q8. Quand les scientifiques ont découvert que le dinosaure-perroquet était de couleur marron, ils étaient fiers.

☐ **VRAI** ☐ **FAUX**

Q9. Les scientifiques ont pu découvrir le milieu de vie du psittacosaurus après avoir analysé un petit morceau de peau prélevé sur le fossile.

☐ **VRAI** ☐ **FAUX**

Q10. Au moment où le texte a été écrit, les secrets des dinosaures n'ont pas encore été tous révélés.

☐ **VRAI** ☐ **FAUX**

D'après toi, ce texte est tiré :

- ☐ d'un roman
- ☐ d'un journal d'informations
- ☐ d'une encyclopédie en ligne (comme vikidia ou wikipedia)
- ☐ d'une publicité
- ☐ d'un livre de recettes de cuisine

VERTE

Elle aurait pu faire l'effort de m'appeler Violette. Mais non, il a fallu qu'elle choisisse Verte. Quelquefois j'ai envie de l'attaquer en justice. Mais quelquefois je l'aime et j'ai envie de lui offrir des vacances de rêve à Honolulu. Rien n'est plus fatigant qu'une mère. Étant entendu que je ne sais pas ce que c'est qu'un père.

J'ai toujours vécu avec ma mère. Pendant des années, je n'ai pas eu à m'en plaindre, au contraire. Elle était un peu étrange, certes. Elle ne ressemblait pas aux mères de mes copines. En un sens, tant mieux : elle avait une allure folle, elle disait des gros mots et elle m'emmenait au cinéma pour un oui, pour un non. Mais sa qualité de sorcière présentait aussi quelques désavantages. Elle passait un temps fou dans sa cuisine à marmonner devant sa Cocotte-Minute en regardant bouillir de dégoûtantes purées brunâtres. L'appartement empestait pendant des jours. Et les catastrophes s'abattaient sur l'immeuble. Fuites d'eau à tous les étages, décès foudroyants de chiens du voisinage, éruptions de boutons sur des familles entières. Il fallait ensuite affronter pendant des semaines les remarques furieuses des habitants de l'immeuble.

Pour m'épargner les ennuis, j'ai réussi très tôt à l'éloigner des abords de l'école. Si je l'avais laissée faire, elle se serait sûrement mêlée de ma scolarité. D'autant qu'en général elle ne pensait pas beaucoup de bien de la directrice, ni de mes instituteurs. Aussi j'aimais autant qu'elle reste à bonne distance. À cette condition, je passais pour une fille à peu près normale.

Dans le fond, nous nous entendions bien. Quelquefois même nous nous amusions franchement. Quand j'étais petite, elle mettait le monde à mes pieds. Elle faisait venir les oiseaux pour qu'ils mangent dans ma main. Elle changeait en un tournemain la couleur de mes robes. Elle s’arrangeait toujours pour qu'il y ait des tas d'enfants lorsque j'allais jouer au square. Je gagnais toujours les jeux de cache-cache qu'elle organisait. Au milieu de l'après-midi, elle nous distribuait des glaces à la menthe et des bonbons en forme de tête de mort. Nous étions si heureuses toutes les deux que je n'avais besoin de personne d'autre dans la vie. Elle me suffisait.

QUESTIONS

Q1. Quand Verte pense au fait que sa mère l'a appelée « Verte », elle ressent de la fierté.

☐ **VRAI** ☐ **FAUX**

Q2. Quand Verte réalise que sa mère est différente de celles de ses copines, elle ressent de la fierté.

☐ **VRAI** ☐ **FAUX**

Q3. La maman de Verte l'emmenait souvent au cinéma.

☐ **VRAI** ☐ **FAUX**

Q4. La mère de Verte restait rarement plus de quelques minutes dans sa cuisine.

☐ **VRAI** ☐ **FAUX**

Q5. Quand Verte voit les purées que cuisine sa mère, elle se sent triste.

☐ **VRAI** ☐ **FAUX**

Q6. Les effets de la sorcellerie de la maman de Verte se dissipaient rapidement.

☐ **VRAI** ☐ **FAUX**

Q7. Quand des catastrophes s'abattent sur l'immeuble, les habitants sont furieux contre la mère de Verte.

☐ **VRAI** ☐ **FAUX**

Q8. Il a fallu du temps à Verte pour réussir à éloigner sa mère des abords de l'école.

☐ **VRAI** ☐ **FAUX**

Q9. Verte aurait préféré que sa mère soit plus impliquée dans sa scolarité.

☐ **VRAI** ☐ **FAUX**

Q10. Verte fait le récit de son enfance et dresse le portrait d'une mère hors norme.

☐ **VRAI** ☐ **FAUX**

D'après toi, ce texte est tiré :

- ☐ d'une encyclopédie en ligne (comme vikidia ou wikipedia)
- ☐ d'un journal d'informations
- ☐ d'un roman
- ☐ d'une publicité
- ☐ d'un livre de recettes de cuisine

MOLIÈRE ET LE THÉÂTRE

Molière (de son vrai nom Jean-Baptiste Poquelin) est un dramaturge et comédien français, né à Paris le 15 janvier 1622, mort à Paris le 17 février 1673.

Jean-Baptiste Poquelin est le fils de Jean Poquelin, un bourgeois aisé de Paris, tapissier, qui avait acheté la charge de valet de chambre du roi.

Dès l'enfance, le jeune Jean-Baptiste assiste à des représentations théâtrales, surtout des farces et des parades de comédiens ambulants. Sa mère meurt quand il a dix ans. Il fait de solides études au Collège de Clermont (aujourd'hui appelé lycée Louis-le-Grand) à Paris où il étudie les mathématiques, la physique, la philosophie, la danse et l'escrime. Il apprend également le latin, ce qui lui permet de lire dans le texte original les comédies de Plaute, de Térence dont il s'inspirera souvent.

Il rencontre, en 1643, une famille de comédiens : les Béjart. Il renonce à succéder à son père comme tapissier du roi. Il décide de devenir comédien malgré le sort qui leur était réservé (ils étaient excommuniés). En juin 1643, s'associant aux Béjart, il fonde avec eux la troupe de “L'Illustre Théâtre”. Il choisit le nom de scène de Molière. Sa troupe, environ une dizaine de comédiens, joue des tragédies à la mode. Elle se heurte aux troupes installées à Paris, comme celles de l'Hôtel de Bourgogne ou du Marais. Les dettes s'accumulent : un moment, Molière est même emprisonné au Châtelet de Paris.

Il n'abandonne pas pour autant le théâtre. L'Illustre Théâtre, s'associant alors à la troupe du comédien Dufresne (protégé par le duc d'Épernon, gouverneur de Guyenne), part jouer en province. De 1645 à 1650, ils se produisent à Agen, Albi, Carcassonne, Nantes, Narbonne et Toulouse entre autres. Les comédiens jouent des tragédies, des comédies d'autres auteurs, mais aussi les premières comédies écrites par Molière. Ils mènent une vie relativement facile. En 1650, Molière devient le directeur de la troupe, ce qui l’oblige à se confronter à des comédiens souvent agités par des querelles et des jalousies. Il doit également résoudre les problèmes d'intendance (location des salles, voyages, paiement de taxes diverses). Il doit aussi affronter l'opposition des autorités locales qui souvent sont hostiles aux spectacles où l'on rit des vieux qui épousent des jeunes filles.

QUESTIONS

Q1. Molière est un comédien français qui s'appelait en réalité Jean-Baptiste Poquelin.

☐ **VRAI** ☐ **FAUX**

Q2. Molière a découvert le théâtre quand il était encore un enfant.

☐ **VRAI** ☐ **FAUX**

Q3. Molière n'a jamais eu la possibilité de faire des études avant de choisir son métier.

☐ **VRAI** ☐ **FAUX**

Q4. Molière a renoncé à succéder à son père après avoir rencontré une famille de comédiens.

☐ **VRAI** ☐ **FAUX**

Q5. Quand Molière renonce à devenir tapissier du roi, il est triste.

☐ **VRAI** ☐ **FAUX**

Q6. Quand Molière se produit avec sa troupe de ville en ville, en province, il est heureux.

☐ **VRAI** ☐ **FAUX**

Q7. Molière devient le directeur de la troupe avant de partir en tournée en province.

☐ **VRAI** ☐ **FAUX**

Q8. Quand Molière prend la direction de la troupe, il est confronté aux jalousies et querelles des acteurs.

☐ **VRAI** ☐ **FAUX**

Q9. Quand Molière est emprisonné pour ses dettes, il sombre dans la dépression.

☐ **VRAI** ☐ **FAUX**

Q10. Molière abandonne le théâtre après avoir été emprisonné.

☐ **VRAI** ☐ **FAUX**

D'après toi, ce texte est tiré :

- ☐ d'un roman
- ☐ d'un livre de recettes de cuisine
- ☐ d'une encyclopédie en ligne (comme vikidia ou wikipedia)
- ☐ d'une publicité
- ☐ d'un journal d'informations

SÉANCE DE LECTURE 3

LES GILETS JAUNES

Depuis le 17 novembre, de nombreux Français bloquent des routes et manifestent contre le gouvernement et la politique du président Emmanuel Macron. Comme signe de leur mécontentement, ils ont enfilé un gilet jaune fluo : c'est pourquoi ce mouvement de protestation est appelé "mouvement des gilets jaunes".

Qui sont les gilets jaunes et comment ce mouvement est-il apparu ?

À l'origine, le mouvement des gilets jaunes s'est formé suite à une forte augmentation du prix des carburants au mois de septembre. D'après les manifestants, l'État était en partie responsable de cette hausse des prix car il avait annoncé l'augmentation des taxes sur le carburant, en particulier sur le diesel, afin qu'il coûte à terme le même prix que l'essence.

Ainsi, le 17 novembre, des rassemblements de gilets jaunes se sont organisés un peu partout en France, notamment grâce aux réseaux sociaux. Des groupes se sont formés pour bloquer des routes, des ronds-points ou des péages d'autoroute durant toute la journée. Puis, de jour en jour, les mouvements de blocage ont continué, car les manifestants estimaient que le président Emmanuel Macron n'avait pas entendu leur colère.

Que revendique le mouvement des gilets jaunes ?

Au départ, les gilets jaunes dénonçaient la hausse des taxes sur le carburant programmée pour janvier 2019 : il était prévu que le diesel augmente de 6,5 centimes par litre et l'essence de 2,9 centimes par litre.

Mais petit à petit, les revendications se sont élargies et la lutte des gilets jaunes est devenue une lutte pour le pouvoir d'achat en général. De nombreux manifestants estiment en effet qu'ils ne gagnent pas suffisamment d'argent pour pouvoir vivre dignement : se loger, se chauffer ou manger à leur faim par exemple.

Ainsi, les gilets jaunes ont dressé une très longue liste de revendications qu'ils ont adressée au gouvernement. Parmi elles, on retrouve par exemple la hausse du SMIC, le salaire minimum accordé aux travailleurs. Des retraités demandaient également la baisse de la CSG, un impôt qui sert à financer la sécurité sociale et qui a augmenté en janvier dernier. D'autres réclamaient le rétablissement de l'ISF (Impôt de solidarité sur la fortune), un impôt payé par les personnes très riches et supprimé début janvier.

Enfin, d'autres réclament simplement plus de lutte contre la pauvreté, contre le mal logement, ou encore plus d'efforts sur l'écologie.

QUESTIONS

Q1. Le mouvement des gilets jaunes a utilisé les réseaux sociaux pour mobiliser les manifestants dans toutes les régions de France.

☐ **VRAI** ☐ **FAUX**

Q2. Quand les manifestants ont enfilé un gilet jaune, ils souhaitaient exprimer leur joie.

☐ **VRAI** ☐ **FAUX**

Q3. Quand les manifestants estiment que le président Macron n'entend pas leurs revendications, ils sont tristes.

☐ **VRAI** ☐ **FAUX**

Q4. Quand les gilets jaunes décident de se rassembler pour bloquer les routes et les ronds-points, ils ressentent de la colère.

☐ **VRAI** ☐ **FAUX**

Q5. Quand le gouvernement annonce la hausse des taxes sur le carburant, les gilets jaunes se sentent inquiets pour leur pouvoir d'achat.

☐ **VRAI** ☐ **FAUX**

Q6. Le mouvement des gilets jaunes a fait son apparition en novembre, avec une première journée de blocage des routes.

☐ **VRAI** ☐ **FAUX**

Q7. Les gilets jaunes se sont mobilisés avant la mise en œuvre de la hausse des taxes du carburant annoncée pour janvier 2019.

☐ **VRAI** ☐ **FAUX**

Q8. Les gilets jaunes se sont d'abord mobilisés pour le pouvoir d'achat en général avant de se concentrer sur la hausse des taxes du carburant.

☐ **VRAI** ☐ **FAUX**

Q9. L'impôt de solidarité sur la fortune a été supprimé après le mouvement des gilets jaunes.

☐ **VRAI** ☐ **FAUX**

Q10. Les revendications de ce mouvement concernent uniquement la pauvreté et le mal-logement.

☐ **VRAI** ☐ **FAUX**

D'après toi, ce texte est tiré :

- ☐ d'un livre de recettes de cuisine
- ☐ d'un journal d'informations
- ☐ d'une publicité
- ☐ d'une encyclopédie en ligne (comme vikidia ou wikipedia)
- ☐ d'un roman

L'EMPOISONNEMENT

Juché sur la cime d'un arbre mort, le corbeau contemplait pensivement la sombre bâtisse devant lui.

Il songeait aux occupants des lieux. À Brunoît, le roi déchu, à Névé, sa femme, et à Mara, leur fille… Trois innocents condamnés à vivre enfermés ici, à Castelbrune, à tout jamais. Dans un instant, il bouleverserait définitivement l'existence de l'un d'entre eux.

L'oiseau connaissait par cœur leur terrible histoire, elle lui avait été contée à maintes reprises.

Tout avait commencé quatorze ans plus tôt, alors qu'au palais le roi, passionné par l'étude des plantes médicinales, consultait un grimoire dans son laboratoire.

Nour, sa sœur aînée, avait surgi, affolée :

– Viens vite, Angus est brûlant de fièvre et ne peut se lever.

Nour, veuve depuis fort longtemps, chérissait son fils au plus haut point, redoutant sans cesse qu'il lui arrive malheur. Le roi s'était rendu au chevet d'Angus, sur lequel veillait son dévoué précepteur. Il l'avait examiné avant de déclarer en souriant :

– Ce n'est qu'une mauvaise grippe. Je vais de ce pas préparer un remède qui fera chuter sa température.

Peu après, le malade avait avalé d'un trait la tisane odorante concoctée par Brunoît. Hélas, il avait aussitôt été pris de violentes convulsions et, en quelques minutes à peine, sous le regard impuissant des adultes, il avait perdu la vie. Le soir même, le médecin de la cour avait livré ses conclusions :

– Angus souffrait en effet d'une simple grippe, mais ce n'est pas la cause du décès…

Puis il avait toussoté, mal à l'aise, avant de reprendre :

– Il a succombé suite à l'absorption simultanée de sa tisane et d'une substance toxique, vraisemblablement du cyanure.

De là à accuser le roi d'assassinat, il n'y avait qu'un pas que la mère d'Angus avait franchi en se ruant sur son frère.

– La tisane… Du cyanure… C'est toi qui l'as tué ! Tu es… tu es un monstre !

Le roi, stupéfait, avait protesté avec vigueur :

– Comment peux-tu imaginer une chose pareille ? Il te faut raison garder, tu sais bien que pour rien au monde je n'aurais fait de mal à Angus.

Mais la fureur de Nour, convaincue d'être en présence du coupable, ne s'était nullement apaisée.

QUESTIONS

Q1. Selon le corbeau, le roi est bien le responsable de la mort d'Angus.

☐ **VRAI** ☐ **FAUX**

Q2. L'empoisonnement d'Angus a eu lieu alors que le corbeau contemple la sombre bâtisse.

☐ **VRAI** ☐ **FAUX**

Q3. Quand Nour constate que son fils Angus a de la fièvre, elle se sent affolée et court chercher son frère le roi.

☐ **VRAI** ☐ **FAUX**

Q4. Quand le roi déclare qu'il va préparer un remède pour Angus, il se sent fier.

☐ **VRAI** ☐ **FAUX**

Q5. Le remède a été administré à Angus sur plusieurs jours.

☐ **VRAI** ☐ **FAUX**

Q6. Angus a été pris de convulsions dès qu'il a bu la tisane.

☐ **VRAI** ☐ **FAUX**

Q7. Quand le médecin de la cour annonce que Angus est mort empoisonné, il se sent fier d'avoir élucidé les causes du décès.

☐ **VRAI** ☐ **FAUX**

Q8. Le médecin de la cour a annoncé la cause du décès d'Angus le jour même.

☐ **VRAI** ☐ **FAUX**

Q9. Quand Nour accuse le roi d'avoir empoisonné son fils, elle ressent de la peur.

☐ **VRAI** ☐ **FAUX**

Q10. La passion du roi pour les plantes médicinales ne lui a pas permis de sauver Angus.

☐ **VRAI** ☐ **FAUX**

D'après toi, ce texte est tiré :

- ☐ d'une publicité
- ☐ d'un roman
- ☐ d'une encyclopédie en ligne (comme vikidia ou wikipedia)
- ☐ d'un journal d'informations

☐ d’un livre de recettes de cuisine

HALLOWEEN

Halloween est une fête qui se déroule dans la nuit du 31 octobre au 1er novembre, c'est-à-dire la veille de la Toussaint. Elle est fêtée principalement dans le monde anglo-saxon.

Traditionnellement, les enfants portent des costumes effrayants. Ils se déguisent en créatures souvent détestées et craintes telles que les fantômes, les vampires ou encore les sorcières, pour sonner aux portes du voisinage en demandant des bonbons ou des pâtisseries. Les petits anglophones utilisent la formule “Trick or treat !” qui signifie “Des bonbons ou un sort !” pour demander ces friandises.

Les Français adoptent Halloween vers la fin des années 90 suite à des campagnes publicitaires et commerciales de marques américaines. En 1997, France Telecom lance la grande opération d'« Olaween » (un téléphone portable de couleur orange) avec notamment 8500 citrouilles décorant le Trocadéro à Paris, finissant ainsi de populariser cette fête. Dès 1998, Halloween est adoptée par les commerçants, la fête ayant lieu juste au moment de la “période creuse” avant les fêtes de Noël. Cependant l’intérêt pour cette fête ne dure pas plus d’une dizaine d’années.

Cette fête est restée critiquée en France parce qu'elle consiste avant tout en une gigantesque opération commerciale. Les véritables organisateurs de cette fête, ce sont finalement les magasins. Ils profitent d'Halloween pour séduire de nouveaux clients, notamment parmi les jeunes, et vendre davantage de marchandises : des vêtements, de la nourriture, des jeux, des livres, des fournitures pour l'école...

Aujourd'hui, Halloween n'est plus une fête importante en France.

L'histoire de la Jack-o'-lantern

D'après la légende, Jack était un ivrogne et un avare. À sa mort, il n'a pas été accepté au Paradis à cause de ses péchés. Mais le diable lui a aussi refusé l'entrée aux enfers car, de son vivant, Jack lui avait joué de mauvais tours.

Le diable l'a donc condamné à errer dans les ténèbres pour l'éternité, avec un navet pour lampion. C'est devenu la Jack-o'-lantern.

Le navet s'est ensuite transformé en citrouille, plus facile à creuser. On y dessine un visage effrayant, et on la place sur les rebords des fenêtres pour chasser les esprits maléfiques.

QUESTIONS

Q1. C'est France Télécom qui a inventé la fête d'Halloween.

☐ **VRAI** ☐ **FAUX**

Q2. La fête d'Halloween a lieu une fois par an, la veille de la Toussaint.

☐ **VRAI** ☐ **FAUX**

Q3. Quand les enfants utilisent la formule des bonbons ou un sort, les voisins se sentent effrayés.

☐ **VRAI** ☐ **FAUX**

Q4. Les français ont commencé à fêter Halloween dès la fin des années 2000.

☐ **VRAI** ☐ **FAUX**

Q5. Halloween permet aux commerçants de tirer profit d'une période creuse, juste après les fêtes de noël.

☐ **VRAI** ☐ **FAUX**

Q6. Quand les français réalisent que Halloween consiste avant tout en une gigantesque opération commerciale, ils se sentent agacés.

☐ **VRAI** ☐ **FAUX**

Q7. Quand le diable refuse à Jack l'entrée aux enfers, le diable se sent triste.

☐ **VRAI** ☐ **FAUX**

Q8. Le diable a condamné Jack à errer dans les ténèbres pour toujours.

☐ **VRAI** ☐ **FAUX**

Q9. Quand on dessine un visage sur la citrouille, c'est pour effrayer les esprits maléfiques.

☐ **VRAI** ☐ **FAUX**

Q10. Halloween est une fête traditionnelle qui vient du monde anglo-saxon et de la légende de Jack-o'-lantern.

☐ **VRAI** ☐ **FAUX**

D'après toi, ce texte est tiré :

- ☐ d'un journal d'informations
- ☐ d'un roman
- ☐ d'une encyclopédie en ligne (comme vikidia ou wikipedia)
- ☐ d'un livre de recettes de cuisine
- ☐ d'une publicité

Annexe II Arbre de décision pour l'identification des modes d'expression

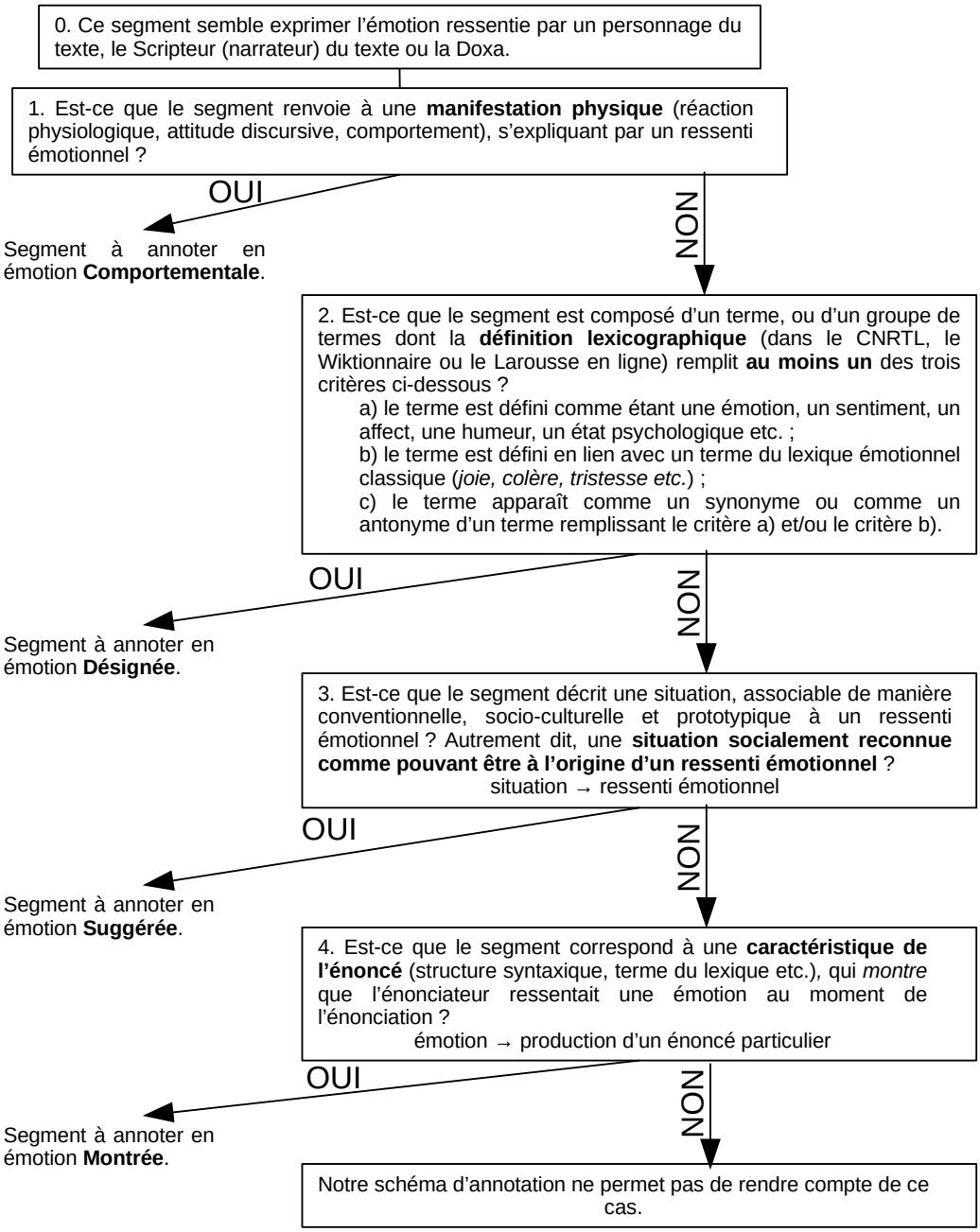


Figure 19: Arbre de décision pour l'identification des modes d'expression

Annexe III Tableaux complémentaires pour le calcul d'AIA

Les tableaux 66 à 68 présentent le détail par sous-corpus des annotations produites par les annotateurs A1, A2 et A7 sur lesquelles s'appuient les mesures d'accord inter-annotateurs (AIA). Ces informations sont présentées par « étiquettes », en nombre d'unités délimitées et de phrases portant l'étiquette.

		Nb d'unités		Nb de phrases	
Corpus	**Étiquette**	**A1 ou A2**	**A7**	**A1 ou A2**	**A7**
Albert50	SitEmo	310	455	201	282
	comportementale	158	195	124	145
	désignée	46	71	45	65
	montrée	32	36	22	26
	suggérée	74	147	68	116
	base	304	453	195	275
	complexe	10	19	10	16
	admiration	8	9	8	9
	colère	202	288	137	187
	culpabilité	0	0	0	0
	dégoût	1	1	1	1
	embarras	2	8	2	6
	fierté	0	2	0	2
	jalousie	0	0	0	0
	joie	6	23	6	19
	peur	53	77	31	53
	surprise	36	36	24	25
	tristesse	6	24	6	22
PtitLibé50	SitEmo	614	498	331	298
	comportementale	147	123	110	87
	désignée	168	113	135	96
	montrée	57	69	34	53
	suggérée	242	189	175	144
	base	592	497	313	286
	complexe	131	45	70	37
	admiration	43	10	25	10
	colère	101	120	65	76
	culpabilité	3	0	2	0
	dégoût	2	0	2	0
	embarras	60	15	33	10
	fierté	25	20	22	18
	jalousie	0	0	0	0
	joie	69	58	48	49
	peur	202	129	128	88
	surprise	77	49	43	37
	tristesse	141	137	88	84

Tableau 66: Modes et catégories émotionnelles annotées, en nombre d'unité et de phrases, selon l'annotateur – corpus Albert50 et PtitLibé50

Corpus	Étiquette	Nb d'unités		Nb de phrases	
		A1 ou A2	**A7**	**A1 ou A2**	**A7**
WP7Encyclo	SitEmo	121	99	63	59
	comportementale	24	11	18	9
	désignée	22	20	19	18
	montrée	37	28	21	16
	suggérée	38	40	32	31
	base	115	99	58	57
	complexe	22	7	12	4
	admiration	15	4	9	2
	colère	42	32	21	19
	culpabilité	0	0	0	0
	dégoût	0	1	0	1
	embarras	0	0	0	0
	fierté	4	1	3	1
	jalousie	3	2	2	2
	joie	6	7	5	6
	peur	26	24	14	15
	surprise	19	10	12	9
	tristesse	22	25	16	18
WP7Fiction	SitEmo	207	145	127	110
	comportementale	44	31	37	29
	désignée	49	37	44	35
	montrée	62	48	41	39
	suggérée	52	21	46	19
	base	214	132	114	100
	complexe	30	17	21	15
	admiration	3	0	2	0
	colère	44	36	22	26
	culpabilité	1	1	1	1
	dégoût	9	10	6	7
	embarras	12	14	11	12
	fierté	14	2	9	2
	jalousie	0	0	0	0
	joie	44	22	30	19
	peur	48	20	32	17
	surprise	30	23	21	21
	tristesse	39	13	25	13

Tableau 67: Modes et catégories émotionnelles annotées, en nombre d'unité et de phrases, selon l'annotateur – corpus WP7Encyclo et WP7Fiction

Corpus	Étiquette	Nb d'unités		Nb de phrases	
		A1 ou A2	**A7**	**A1 ou A2**	**A7**
WP7Journa	SitEmo	119	142	60	79
	comportementale	36	58	25	41
	désignée	23	26	20	22
	montrée	20	17	14	12
	suggérée	40	36	34	30
	base	98	129	50	69
	complexe	29	16	18	12
	admiration	12	9	9	8
	colère	43	62	22	38
	culpabilité	0	0	0	0
	dégoût	0	0	0	0
	embarras	11	7	5	4
	fierté	6	0	5	0
	jalousie	0	0	0	0
	joie	11	14	7	10
	peur	19	21	12	14
	surprise	18	7	10	3
	tristesse	7	19	5	14

Tableau 68: Modes et catégories émotionnelles annotées, en nombre d'unité et de phrases, selon l'annotateur – corpus WP7Journa

Les tableaux 69 à 71 présente le détail des Kappa de Cohen obtenus entre A1 et A7, pour chaque étiquette. Ces calculs sont opérés sur une simplification de la tâche d'annotation : les étiquettes sont évaluées à l'échelle de la phrase uniquement. Les cases grisées correspondent à des étiquettes absentes des annotations chez les deux annotatrices. Le détail par sous-corpus n'est pas présenté pour A2 car un seul sous-corpus annoté par A2 (Albert50) a également été annoté par A7. Le tableau 71 qui présente les mesures par genre de texte (journalistique, fictionnel, encyclopédique) ne donne à voir que les Kappa de Cohen obtenu pour le genre journalistique. Les valeurs pour les genres encyclopédique et fictionnel correspondent respectivement à celles des corpus WP7Encyclo et WP7Fiction, seuls sous-corpus correspondant à ces genres.

Corpus	Étiquette	A1 *vs.* A7
PtitLibé50	émotionnelle	0,643
	comportementale	0,673
	désignée	0,765
	montrée	0,606
	suggérée	0,538
	base	0,623
	complexe	0,444
	admiration	0,219
	colère	0,736
	culpabilité	0
	dégoût	0
	embarras	0,364
	fierté	0,644
	jalousie	
	joie	0,43
	peur	0,674
	surprise	0,637
	tristesse	0,56
WP7Encyclo	émotionnelle	0,811
	comportementale	0,645
	désignée	0,851
	montrée	0,673
	suggérée	0,642
	base	0,817
	complexe	0,484
	admiration	0,353
	colère	0,833
	culpabilité	
	dégoût	0
	embarras	
	fierté	0,496
	jalousie	1
	joie	0,72
	peur	0,74
	surprise	0,749
	tristesse	0,679

Tableau 69: Kappa de Cohen entre A1 et A7 - corpus PtitLibé50 et WP7Encyclo

Corpus	Étiquette	A1 *vs.* A7
WP7Fiction	émotionnelle	0,636
	comportementale	0,731
	désignée	0,756
	montrée	0,516
	suggérée	0,413
	base	0,653
	complexe	0,53
	admiration	0
	colère	0,685
	culpabilité	1
	dégoût	0,922
	embarras	0,685
	fierté	0,174
	jalousie	
	joie	0,714
	peur	0,496
	surprise	0,491
	tristesse	0,556
WP7Journa	émotionnelle	0,544
	comportementale	0,485
	désignée	0,892
	montrée	0,751
	suggérée	0,466
	base	0,488
	complexe	0,637
	admiration	0,691
	colère	0,644
	culpabilité	
	dégoût	
	embarras	0,886
	fierté	0
	jalousie	
	joie	0,568
	peur	0,668
	surprise	0,447
	tristesse	0,286

Tableau 70: Kappa de Cohen entre A1 et A7 - corpus WP7Fiction et WP7Journa

Genre	**Étiquette**	**A1 *vs.* A7**
Journa	émotionnelle	0,63
	comportementale	0,633
	désignée	0,784
	montrée	0,64
	suggérée	0,528
	base	0,603
	complexe	0,487
	admiration	0,373
	colère	0,715
	culpabilité	0
	dégoût	0
	embarras	0,453
	fierté	0,571
	jalousie	
	joie	0,45
	peur	0,673
	surprise	0,61
	tristesse	0,532

Tableau 71: Kappa de Cohen entre A1 et A7 – genre journalistique

Annexe IV Tableaux d'entités des unités Expérienceurs

Les tableaux 72 à 75 présentent, pour chaque sous-corpus annoté, le top 10 des segments renseignés pour le trait *Entité* des unités Expérienceurs.

PtitLibe		Albert	
Entité	**Nb.**	**Entité**	**Nb.**
doxa	425	scripteur	551
Scripteur	359	doxa	48
les gilets jaunes	27	Donald Trump	19
Lecteur	19	la population	16
Gabriel	17	les policiers	14
les animaux	16	de nombreux Français	14
des gens	14	les habitants	10
les Français	14	Certains	7
les enfants	14	le président américain Donald Trump	7
un camarade de classe	12	certains	7

Tableau 72: Top 10 des entités – Corpus journalistiques (1/2)

CovidAdultes		CovidJeunesse	
Entité	**Nb.**	**Entité**	**Nb.**
doxa	7	Myriam, 10 ans	4
L'OMS	2	scripteur	3
les syndicats Force ouvrière (FO) et la Confédération française démocratique du travail (CFDT)	1	Orleynie, 11 ans, et Oxyne 15 ans	3
nous	1	Martin, 10 ans	2
une partie des français	1	Sa maman	2
trois médecins, représentants d'un collectif de soignants	1	doxa	2
ses pairs	1	Une infirmière qui travaille en réanimation	2
les gouvernements	1	Jiashu, 12 ans	2
la population	1	Abel, 9 ans	2
Les pays avec des structures sanitaires vulnérables	1	Aurélien	1

Tableau 73: Top 10 des entités – Corpus journalistiques (2/2)

FictionProjetM2		**WP7Fiction**	
Entité	**Nb.**	**Entité**	**Nb.**
Flavia	118	l'âne	13
Elisa	114	Verte	12
Orlando	63	Émilie	8
Anna	56	Chien Pourri	8
narrateur	36	Tristan	8
Eléonore	30	Le roi	8
Solenn Croazic	29	le loup	6
Gabriel	27	Nour, sa sœur aînée	5
Jasper	24	Narrateur	5
Dana	19	Mr et Mrs Dursley	5

Tableau 74: Top 10 des entités – Corpus fictionnels

WP7Encyclo	
Entité	**Nb.**
Scripteur	27
Doxa	7
son héritier, le frondeur Paul de Gondi	4
Molière (de son vrai nom Jean-Baptiste Poquelin)	3
L'ara de Spix	3
le diable	3
Le capitaine Haddock, dont le prénom est « Archibald »	3
Les comédiens	2
autorités locales qui souvent sont hostiles aux spectacles où l'on rit des vieux qui épousent des jeunes filles	2
les princes	2

Tableau 75: Top 10 des entités – Corpus encyclopédique

Annexe V Répartition des phrases du corpus global au sein du corpus d'apprentissage

Le tableau 76 illustre la répartition des phrases du corpus annoté global, selon la classe présente dans la phrase. Le *train* comporte par exemple 73,5 % des phrases du corpus global associées à la classe « émotionnelle ».

Classe	**Prop. des phrases du corpus global dans le**		
	train	***dev***	***test***
émotionnelle	73,5 %	8,2 %	18,3 %
comportemental	72,9 %	8,0 %	19,2 %
désigné	69,2 %	9,7 %	21,1 %
montré	73,4 %	6,5 %	20,1 %
suggéré	73,2 %	8,4 %	18,4 %
base	72,7 %	8,4 %	18,8 %
complexe	67,6 %	10,0 %	22,4 %
admiration	58,3 %	14,7 %	27,0 %
autre	76,6 %	7,0 %	16,4 %
colère	76,4 %	7,6 %	15,9 %
culpabilité	73,7 %	5,3 %	21,1 %
dégoût	67,3 %	13,5 %	19,2 %
embarras	69,1 %	10,5 %	20,4 %
fierté	70,3 %	5,9 %	23,8 %
jalousie	71,4 %	14,3 %	14,3 %
joie	70,2 %	7,1 %	22,7 %
peur	71,1 %	8,7 %	20,2 %
surprise	72,3 %	10,4 %	17,2 %
tristesse	71,3 %	8,2 %	20,5 %

Tableau 76: Répartition des phrases du corpus global entre le train, le dev et le test, selon les classes

Annexe VI Procédure de transposition des annotations à l'échelle de la phrase

L'entraînement de classifieurs en émotions à partir du corpus annoté décrit chapitre 6 a nécessité plusieurs transformation du corpus. L'une d'elle a consisté en la transposition au niveau de la phrase des informations d'annotation portées par les unités annotées du corpus (niveau du marqueur linguistique). Cette transposition a été opérée par un script Python, dont les étapes de fonctionnement sont résumées ici.

Résultant de l'application manuelle du schéma introduit chapitre 5 sur un ensemble de textes *via* la plateforme Glozz (Widlöcher & Mathet, 2012), le corpus annoté est au format Glozz. Il est donc composé de fichiers de type AC, qui stockent les contenus textuels des textes du corpus, et de fichiers de type AA, qui stockent les informations d'annotation au format XML. Un système d'indices opère le lien entre le contenu textuel et les unités annotées. Chaque unité du fichier AA est caractérisée par un indice de début noté a, indiquant sur quel caractère du texte commence l'unité, et par un indice de fin noté b, indiquant sur quel caractère du texte termine l'unité[97]. Les marqueurs linguistiques captés par une unité d'annotation correspondent ainsi à la portion du texte du fichier AC qui commence au caractère d'indice a et qui finit au caractère d'indice b.

Pour chaque texte, autrement dit pour chaque couple formé d'un fichier AC et d'un fichier AA, la première étape du script consiste à récupérer le contenu textuel et les annotations en s'appuyant sur le système d'indices pour retrouver les marqueurs linguistiques captés par chaque unité. Puisque la plateforme Glozz encode les paragraphes comme des unités d'annotation, la liste des paragraphes qui composent le texte fait partie des informations ainsi retrouvées.

Pour chaque paragraphe, le script emploie ensuite une expression régulière pour chercher de manière itérative l'indice de début noté i et l'indice de fin noté j de chaque phrase du paragraphe. Les textes des phrases sont obtenus en s'appuyant sur les indices ainsi trouvés.

Le script confronte ensuite les indices i et j de chaque phrase aux indices a et b des unités annotées pour déterminer celles contenues dans la phrase. Une phrase d'indices

97 Dans le cas présent, un *indice* est un numéro associé à chaque caractère du texte stocké dans le fichier AC. Le premier caractère a pour indice 0, le second 1, le troisième 2, *etc.*

(i,j), où i vaut 0 et j vaut 42, inclura par exemple les unités d'indices (10,16) ou (30,42) mais pas l'unité d'indices (56, 60). Si les annotateurs ont respecté les consignes du guide (Étienne & Battistelli, 2021), aucune unité ne doit se trouver à cheval entre deux phrases. Toutes les informations d'annotation associées à chaque phrase du corpus sont ainsi enregistrées. Ces informations sont ensuite filtrées pour ne garder que celles utiles aux tâches de classification A à E : seules les listes des unités SitEmo et des unités Autre contenues dans la phrase sont conservées.

Annexe VII Performances détaillées des *baselines*

Cette annexe présente les performances détaillées des modèles de *baseline* obtenus sur les cinq tâches de classification définies chapitre 7. Les résultats par classe obtenus par ces modèles sont également exposés.

1) Tâche A

La tâche A vise la prédiction de la classe « émotionnelle ». Le tableau 77 présente les performances de tous les modèles SVM et XGBoost testés pour cette tâche. On constate que les modèles SVM sont globalement meilleurs que les modèles XGBoost.

		Dev			*Test*		
Modèle	**Représentation**	**R**	**P**	**F**	**R**	**P**	**F**
SVM	sac_tokens_10000	0,411	0,723	0,524	0,384	0,705	0,497
	sac_tokens_15000	0,463	0,666	0,546	0,478	0,656	0,553
	sac_tokens_20000	0,477	0,668	0,557	0,481	0,659	0,556
	embeddings	0,463	0,677	0,550	0,497	0,674	0,572
XBGoost	sac_tokens_10000	0,221	0,758	0,343	0,223	0,700	0,338
	plongements	0,201	0,550	0,294	0,188	0,511	0,275

Tableau 77: Performances des modèles SVM et XGBoost pour la tâche A

2) Tâche B

La tâche B vise la prédiction des quatre classes de modes. Le tableau 78 présente les performances de tous les modèles SVM et XGBoost testés pour cette tâche. Les performances sont des macro-moyennes (moyennes des R, des P et des F1 obtenus pour chaque classe). On constate que les modèles SVM sont globalement meilleurs que les modèles XGBoost.

		Dev			*Test*		
Modèle	**Représentation**	**macro-R**	**macro-P**	**macro-F1**	**macro-R**	**macro-P**	**macro-F1**
SVM	sac_tokens_20000	0,317	0,759	0,408	0,267	0,721	0,368
	plongements	0,208	0,625	0,299	0,195	0,671	0,291
XGBoost	sac_tokens_20000	0,296	0,702	0,384	0,218	0,730	0,314
	plongements	0,033	0,230	0,057	0,057	0,311	0,095

Tableau 78: Performances des modèles SVM et XGBoost pour la tâche B

3) Tâche C

La tâche C vise la prédiction des deux classes de types. Le tableau 79 présente les performances de tous les modèles SVM et XGBoost testés pour cette tâche. Les performances sont des macro-moyennes (moyennes des R, des P et des F1 obtenus pour chaque classe). On constate que pour la représentation de phrase de type plongement, les performances du modèle SVM sont meilleures que celles du modèle XGBoost. C'est l'inverse pour la représentation de type sac de tokens (pour la F1 et la P, pas le R) sur le *dev*. Sur le *test*, le modèle SVM reste globalement meilleur (F1 et R).

		Dev			*Test*		
Modèle	**Représentation**	**macro-R**	**macro-P**	**macro-F1**	**macro-R**	**macro-P**	**macro-F1**
SVM	sac_tokens_20000	0,216	0,342	0,265	0,211	0,343	0,261
	plongements	0,220	0,695	0,293	0,213	0,831	0,273
XGBoost	sac_tokens_20000	0,165	0,856	0,274	0,120	0,659	0,200
	plongements	0,079	0,424	0,130	0,072	0,382	0,119

Tableau 79: Performances des modèles SVM et XGBoost pour la tâche C

4) Tâche D

La tâche D vise la prédiction des douze classes de catégories émotionnelles. Le tableau 80 présente les performances de tous les modèles SVM et XGBoost testés pour cette tâche. Les performances sont des macro-moyennes (moyennes des R, des P et des F1 obtenus pour chaque classe). On constate que pour la représentation plongement, les modèles SVM sont meilleurs que les modèles XGBoost. C'est l'inverse pour la représentation sacs de tokens.

		Dev			*Test*		
Modèle	**Représentation**	**macro-R**	**macro-P**	**macro-F1**	**macro-R**	**macro-P**	**macro-F1**
SVM	sac_tokens_20000	0,144	0,494	0,209	0,125	0,487	0,186
	plongements	0,126	0,441	0,190	0,122	0,480	0,181
XGBoost	sac_tokens_20000	0,179	0,557	0,263	0,192	0,565	0,272
	plongements	0,036	0,203	0,060	0,041	0,210	0,068

Tableau 80: Performances des modèles SVM et XGBoost pour la tâche D

5) Tâche E

La tâche E vise la prédiction de dix-neuf classes. Le tableau 81 présente les performances de tous les modèles SVM et XGBoost testés pour cette tâche. Les performances sont des macro-moyennes (moyennes des R, des P et des F1 obtenus pour chaque classe). On constate que pour la représentation plongement, les modèles SVM

sont meilleurs que les modèles XGBoost. C'est globalement l'inverse pour la représentation sacs de tokens.

		Dev			***Test***		
Modèle	**Représentation**	**macro-R**	**macro-P**	**macro-F1**	**macro-R**	**macro-P**	**macro-F1**
SVM	sac_tokens_20000	0,205	0,543	0,275	0,183	0,530	0,252
	plongements	0,171	0,519	0,242	0,167	0,567	0,235
XGBoost	sac_tokens_20000	0,204	0,629	0,294	0,191	0,617	0,277
	plongements	0,048	0,250	0,079	0,055	0,265	0,090

Tableau 81: Performances des modèles SVM et XGBoost pour la tâche E

6) Performances par classe des modèles de *baseline*

Les performances des modèles de *baseline* sont ici présentés pour chaque classe. Le tableau 82 résume les performances du modèle aléatoire, le tableau 83 celles des modèles SVM et le tableau 84 celles des modèles XGBoost.

	Dev			***Test***		
Classe	**R**	**P**	**F1**	**R**	**P**	**F1**
émotionnelle	0,202	0,045	0,074	0,202	0,176	0,188
comportemental	0,046	0,036	0,040	0,046	0,043	0,044
désigné	0,053	0,052	0,053	0,053	0,057	0,055
montré	0,036	0,023	0,028	0,036	0,035	0,036
suggéré	0,071	0,058	0,064	0,071	0,063	0,067
base	0,154	0,126	0,138	0,154	0,140	0,146
complexe	0,020	0,021	0,020	0,020	0,023	0,021
admiration	0,006	0,011	0,008	0,006	0,010	0,008
autre	0,050	0,032	0,039	0,050	0,037	0,043
colère	0,046	0,032	0,038	0,046	0,034	0,039
culpabilité	0,001	0,000	0,000	0,001	0,001	0,001
dégoût	0,002	0,003	0,002	0,002	0,002	0,002
embarras	0,006	0,006	0,006	0,006	0,006	0,006
fierté	0,007	0,004	0,005	0,007	0,009	0,008
jalousie	0,000	0,000	0,000	0,000	0,000	0,000
joie	0,032	0,023	0,027	0,032	0,036	0,034
peur	0,038	0,033	0,035	0,038	0,038	0,038
surprise	0,030	0,031	0,031	0,030	0,025	0,028
tristesse	0,025	0,020	0,022	0,025	0,025	0,025

Tableau 82: Performances par classe du modèle aléatoire

	Représentation sac de tokens (N=20 000)						***Représentation plongement***					
	Dev			***Test***			***Dev***			***Test***		
Classe	**R**	**P**	**F**	**R**	**P**	**F**	**R**	**P**	**F**	**R**	**P**	**F**
émotionnelle	0,477	0,668	0,557	0,481	0,659	0,556	0,463	0,677	0,550	0,497	0,674	0,572
comportemental	0,374	0,712	0,490	0,265	0,778	0,395	0,253	0,641	0,362	0,185	0,579	0,280
désigné	0,262	0,884	0,404	0,328	0,867	0,476	0,214	0,756	0,333	0,281	0,742	0,407
montré	0,587	0,740	0,655	0,446	0,713	0,549	0,317	0,571	0,408	0,282	0,764	0,412
suggéré	0,044	0,700	0,082	0,028	0,526	0,054	0,050	0,533	0,091	0,034	0,600	0,065
base	0,431	0,685	0,529	0,422	0,685	0,523	0,405	0,723	0,519	0,402	0,662	0,500
complexe	0	0	0	0	0	0	0,035	0,667	0,067	0,024	1	0,046
admiration	0	0	0	0	0	0	0	0	0	0	0	0
autre	0,348	0,721	0,470	0,351	0,664	0,459	0,292	0,897	0,441	0,216	0,608	0,319
colère	0,378	0,756	0,504	0,277	0,929	0,426	0,356	0,696	0,471	0,324	0,604	0,422
culpabilité	0	0	0	0	0	0	0	0	0	0	0	0
dégoût	0	0	0	0	0	0	0	0	0	0	0	0
embarras	0,118	1	0,211	0,091	1	0,167	0,059	0,500	0,105	0,061	0,667	0,111
fierté	0	0	0	0	0	0	0	0	0	0,042	1	0,080
jalousie	0	0	0	0	0	0	0	0	0	0	0	0
joie	0,127	0,727	0,216	0,114	0,719	0,197	0,190	0,857	0,312	0,158	0,627	0,253
peur	0,308	0,778	0,441	0,311	0,733	0,437	0,297	0,771	0,429	0,354	0,743	0,479
surprise	0,372	0,941	0,533	0,282	0,800	0,417	0,233	0,741	0,354	0,218	0,861	0,348
tristesse	0,073	1	0,136	0,072	1	0,135	0,091	0,833	0,164	0,094	0,650	0,165

Tableau 83: Performances par classe des modèles SVM

	Représentation sac de tokens (N=20 000)						***Représentation plongement***					
	Dev			***Test***			***Dev***			***Test***		
Classe	**R**	**P**	**F1**	**R**	**P**	**F1**	**R**	**P**	**F1**	**R**	**P**	**F1**
émotionnelle	0,221	0,758	0,343	0,223	0,700	0,338	0,201	0,550	0,294	0,188	0,511	0,275
comportemental	0,263	0,703	0,382	0,147	0,814	0,249	0,061	0,462	0,107	0,076	0,360	0,125
désigné	0,228	0,943	0,367	0,240	0,884	0,377	0,041	0,375	0,075	0,066	0,467	0,116
montré	0,651	0,774	0,707	0,436	0,691	0,535	0,032	0,083	0,046	0,077	0,319	0,124
suggéré	0,044	0,389	0,079	0,051	0,529	0,093	0	0	0	0,009	0,097	0,016
base	0,207	0,837	0,332	0,185	0,818	0,302	0,141	0,598	0,228	0,112	0,497	0,182
complexe	0,123	0,875	0,215	0,055	0,500	0,099	0,018	0,250	0,033	0,031	0,267	0,056
admiration	0,097	0,750	0,171	0,035	0,250	0,062	0	0	0	0,018	0,167	0,032
autre	0,315	0,651	0,424	0,274	0,582	0,373	0,079	0,438	0,133	0,063	0,302	0,104
colère	0,311	0,700	0,431	0,223	0,875	0,356	0,111	0,556	0,185	0,128	0,480	0,202
culpabilité	0	0	0	0	0	0	0	0	0	0	0	0
dégoût	0	0	0	0,500	0,714	0,588	0	0	0	0	0	0
embarras	0,294	0,714	0,417	0,182	0,500	0,267	0,059	0,500	0,105	0,030	0,200	0,053
fierté	0,083	0,333	0,133	0,146	0,875	0,250	0	0	0	0	0	0
jalousie	0	0	0	0	0	0	0	0	0	0	0	0
joie	0,206	1	0,342	0,119	0,686	0,203	0,079	0,313	0,127	0,074	0,250	0,115
peur	0,308	0,933	0,463	0,325	0,812	0,465	0,088	0,533	0,151	0,108	0,622	0,185
surprise	0,407	0,897	0,560	0,401	0,803	0,535	0,012	0,100	0,021	0,042	0,261	0,073
tristesse	0,127	0,700	0,215	0,094	0,684	0,166	0	0	0	0,029	0,235	0,052

Tableau 84: Performances par classe des modèles XGBoost

Annexe VIII Détails de la comparaison du *dev* et du *test*

Cette annexe présente le détail de la comparaison effectuée entre le *dev* et le *test* afin de trouver des pistes d'explication à la baisse des performances des classifieurs observée sur le *test*.

Piste 1 : composition du *test*. La première piste explorée pour cerner l'origine des différences de performances entre *dev* et *test* est d'observer la composition du *test*. Deux sous-corpus ont été distingués au sein du *test* selon leur composition :

- le *test vus*, composé de sources présentes dans le train ou le dev ;
- *le test non vus*, composé de sources absentes du train et du dev.

Les performances du modèle E ont été calculées sur chacun de ces sous-corpus. Le tableau 85 présente les mesures de R, de P et de F1 du modèle E, sur le *dev*, le *test*, le *test vus* et le *test non vus*, organisées par tâche. Pour chaque tâche, le taux de rappel le plus élevé est toujours obtenu sur le *test non vus*. En revanche, la précision et la F-mesure sont systématiquement inférieures sur ce corpus.

Tâche	**Mesure**	***dev***	***test***	***test vus***	***test non vus***
A	R	0,785	0,765	0,755	0,919
	P	0,748	0,745	0,817	0,358
	F	0,766	0,755	0,785	0,516
B	macro-R	0,693	0,625	0,617	0,760
	macro-P	0,672	0,668	0,729	0,314
	macro-F	0,681	0,645	0,668	0,433
C	macro-R	0,574	0,564	0,550	0,780
	macro-P	0,783	0,675	0,728	0,375
	macro-F	0,640	0,610	0,623	0,506
D	macro-R	0,418	0,397	0,390	0,487
	macro-P	0,501	0,467	0,507	0,184
	macro-F	0,441	0,422	0,433	0,253
E	macro-R	0,512	0,482	0,474	0,598
	macro-P	0,579	0,546	0,593	0,240
	macro-F	0,530	0,506	0,521	0,331

Tableau 85: Performances du modèle E groupées par tâche, selon le sous-corpus

Le tableau 86 présente uniquement les F-mesures obtenues par le modèle E, sur le *dev*, le *test*, le *test vus* et le *test non vus*, pour chaque classe. Mise à part pour la classe « fierté », les F1 du *test non vus* sont toujours inférieures à celles du *dev* et du *test vus*.

Classe	***dev***	***test***	***test vus***	***test non vus***
comportementale	0,740	0,628	0,655	0,350
désignée	0,813	0,815	0,832	0,725
montrée	0,741	0,701	0,725	0,438
suggérée	0,429	0,438	0,458	0,218
base	0,756	0,716	0,735	0,568
complexe	0,524	0,505	0,510	0,444
admiration	0,308	0,360	0,372	0
autre	0,708	0,661	0,713	0,054
colère	0,788	0,672	0,678	0,600
culpabilité	0	0	0	abs.
dégoût	0	0	0	abs.
embarras	0,643	0,462	0,480	0
fierté	0,286	0,421	0,394	0,600
jalousie	0	0	0	abs.
joie	0,591	0,618	0,631	0,485
peur	0,670	0,689	0,727	0,535
surprise	0,747	0,719	0,731	0,400
tristesse	0,554	0,458	0,473	0,364

Tableau 86: Performances du modèle E par classe, selon le sous-corpus

Piste 2 : distance entre les distributions de classes selon les couples (*c,m*). Une seconde piste pour expliquer la spécificité du *test* repose sur l'observation des distributions de l'ensemble de *test*, *test non vus*, *test vus*, *dev* et *train*. Les distributions comparées sont celles obtenues à partir de la répartition des phrases selon les couples (*c,m*) où *c* est une catégorie émotionnelle dénoté par la phrase et *m* le mode qui l'exprime.

Distances entre les distributions des différents ensembles. L'ensemble ayant la distribution la plus éloignée du *train* est *test non vus* (B 2,41). Cet ensemble est donc encore plus éloigné du *train* que la distribution uniforme. Entre le *dev* (B 1,59) et le *test vus* (B 1,36), c'est ce dernier qui est le plus proche du *train*. L'éloignement par rapport au *train* pourrait expliquer ici la chute de performances observées sur le *test non vus*. Les performances du modèle E sont cependant globalement meilleure sur le *dev* que sur le *test vus*, ce dernier étant pourtant plus proche du *train*.

Distances entre distributions attendues et distributions prédites. Le tableau 87 présente les distances de Bhattacharyya calculées entre différentes distributions attendues (celles des ensembles de référence) et prédites (celles correspondant aux prédictions du modèle E). La distribution prédite sur le *test* est plus proche de la distribution attendue pour cet ensemble (B 1,38) qu'elle ne l'est de celle attendue pour le *train* (B 1,49). Les distributions prédites du *dev* et du *test non vus* sont en revanche plus proches de la distribution attendue pour le *train* (respectivement B 1,62 et B 1,86), qu'elles ne le sont des distributions attendues pour ces ensembles (resp. B 1,73 et B 2,66), les écarts étant bien plus prononcés pour le *test non vus*. Les écarts prononcés observés entre la distribution prédite pour le *test non vus* et d'une part, celle attendue pour cet ensemble, d'autre part celle attendue pour le *train*, semblent illustrer la chute de performance observée sur le *test non vus*. Il est en revanche plus difficile d'interpréter les distances pour le *dev* et le *test vus*. Les performances sont en effet légèrement moins bonnes sur ce dernier que sur le *dev*. Pourtant, les distributions attendues et prédites sur le *test vus* sont les plus proches. Par ailleurs, les plus petites distances qui séparent la distribution du *train* des distributions prédites pour le *dev* et pour le *test non vus* suggèrent que le modèle E s'appuie sur la distribution du train pour faire ses prédictions.

Distribution attendue	**Distribution prédite**	**Distance de Bhattacharyya**
train	*dev*	1,62
train	*test vus*	1,49
train	*test non vus*	1,86
dev	*dev*	1,73
test vus	*test vus*	1,38
test non vus	*test non vus*	2,66

Tableau 87: Distance de Bhattacharyya entre différentes distributions attendues et prédites

Annexe IX Compléments aux matrices de confusion

Cette annexe expose des compléments accompagnant l'analyse des résultats menée à l'aide des matrices de confusion au chapitre 6.

1) Tables de croisement des classes

Les tableaux 88 à 90 présente, en nombre de phrases, les croisement entre classes au sein du *test*. L'étiquette « aucune » (abrégée « auc. ») permet de rendre compte des phrases n'articulant qu'une seule classe de mode, de type ou de catégorie émotionnelle.

	comp.	**dés.**	**mont.**	**sugg.**	**auc.**
comp.		49	23	59	130
dés.	49		28	76	189
mont.	23	28		61	97
sugg.	59	76	61		179

Tableau 88: Croisement des classes de mode au sein du test

	base	comp.	auc.
base		71	708
comp.	71		56

Tableau 89: Croisement des classes de type au sein du test

	adm.	**aut.**	**col.**	**culp.**	**deg.**	**emb.**	**fier.**	**jal.**	**joie**	**peur**	**surp.**	**tris.**	**auc.**
adm.		3	1	0	0	0	14	0	21	0	15	4	23
aut.	3		10	0	1	5	5	0	21	14	5	12	149
col.	1	10		0	0	3	0	1	2	8	8	15	146
culp.	0	0	0		1	2	0	0	0	1	0	0	1
deg.	0	1	0	1		1	0	0	0	2	1	3	6
emb.	0	5	3	2	1		0	0	1	6	1	7	13
fier.	14	5	0	0	0	0		0	34	0	11	1	11
jal.	0	0	1	0	0	0	0		0	0	0	0	0
joie	21	21	2	0	0	1	34	0		4	20	7	124
peur	0	14	8	1	2	6	0	0	4		12	35	145
surp.	15	5	8	0	1	1	11	0	20	12		4	92
tris.	4	12	15	0	3	7	1	0	7	35	4		64

Tableau 90: Croisement des classes de catégorie émotionnelle au sein du test

2) Matrices de confusion : *test vus* et *test non vus*

Les tableaux 91 à 94 présentent les matrices de confusion pour les tâches A à D, d'après les prédictions du modèle E, sur le *test vus* et le *test non vus*.

test vus

	émo.	non-émo.
émo.	696	226
non-émo.	156	3086

test non vus

	émo.	non-émo.
émo.	57	5
non-émo.	102	1252

Tableau 91: Matrices de confusion tâche A, test vus et test non vus

test vus

	comp.	dés.	mont.	sugg.	auc.
comp.	135	42	19	58	47
dés.	28	223	29	67	28
mont.	15	25	124	32	53
sugg.	41	74	42	139	128
auc.	37	17	25	66	3255

test non vus

	comp.	dés.	mont.	sugg.	auc.
comp.	7	3	0	5	1
dés.	2	37	4	11	2
mont.	0	0	7	4	0
sugg.	1	5	2	6	4
auc.	23	22	13	22	1281

Tableau 92: Matrices de confusion tâche B, test vus et test non vus

test vus

	base	comp.	auc.
base	497	44	199
comp.	33	50	50
auc.	126	13	3253

test non vus

	base	comp.	auc.
base	50	4	6
comp.	2	4	0
auc.	69	4	1283

Tableau 93: Matrices de confusion tâche C, test vus et test non vus

test vus

	adm.	aut.	col.	culp.	deg.	emb.	fier.	jal.	joie	peur	surp.	tris.	auc.
adm.	16	5	0	0	0	0	6	0	6	0	10	1	23
aut.	3	154	9	0	0	2	3	0	18	12	3	11	35
col.	1	11	116	0	0	5	0	0	3	7	6	6	43
culp.	0	0	0	0	0	1	0	0	0	0	0	1	2
deg.	0	4	0	0	0	2	0	0	1	3	0	6	1
emb.	0	6	2	0	0	12	0	0	0	2	4	5	9
fier.	4	4	0	0	0	1	13	0	8	0	5	0	20
jal.	0	0	1	0	0	0	0	0	0	0	0	0	0
joie	9	24	2	0	0	1	15	0	101	4	18	4	51
peur	1	17	9	0	0	2	0	0	3	128	7	19	39
surp.	6	6	7	0	0	1	5	0	9	7	98	2	29
tris.	4	14	9	0	0	5	0	0	4	26	4	52	40
auc.	8	43	36	0	0	1	2	0	16	26	18	18	3089

test non vus

	adm.	aut.	col.	culp.	deg.	emb.	fier.	jal.	joie	peur	surp.	tris.	auc.
adm.	0	0	0	0	0	0	1	0	1	0	0	0	0
aut.	0	1	0	0	0	0	0	0	0	0	0	0	0
col.	0	1	9	0	0	0	0	0	0	3	0	0	1
culp.	0	0	0	0	0	0	0	0	0	0	0	0	0
deg.	0	0	0	0	0	0	0	0	0	0	0	0	0
emb.	0	0	0	0	0	0	0	0	0	0	0	1	0
fier.	0	0	0	0	0	0	3	0	0	0	0	0	0
jal.	0	0	0	0	0	0	0	0	0	0	0	0	0
joie	0	2	0	0	0	0	0	0	8	0	2	2	1
peur	0	0	2	0	0	1	0	0	0	23	1	1	2
surp.	0	0	0	0	0	0	0	0	0	1	2	0	0
tris.	0	1	0	0	0	0	0	0	1	0	0	6	1
auc.	1	31	9	0	0	0	3	0	13	34	3	17	1251

Tableau 94: Matrices de confusion tâche D, test vus et test non vus

Annexe X Procédure d'analyse des interactions entre mode et catégorie émotionnelle (prédictions du modèle E)

Cette annexe détaille la procédure mise en œuvre pour étudier l'impact des interactions entre mode et catégorie émotionnelle sur les prédictions du modèle E sur le *test*.

Observation 1 : Impact du mode sur la prédiction de la catégorie. Pour évaluer l'impact d'un mode m sur la prédiction des catégories émotionnelles, seules les informations émotionnelles issues d'une unité émotionnelle exprimée selon le mode m sont prises en compte.

Chaque phrase du *test* est ainsi associée à quatre vecteurs de dimension 11 – un vecteur pour chaque mode – où chaque position du vecteur correspond à une classe émotionnelle suivant la liste : « admiration », « colère », « culpabilité », « dégoût », « embarras », « fierté », « jalousie », « joie », « peur », « surprise » et « tristesse ». Chaque vecteur est une liste d'étiquettes attendues pour la phrase. Une phrase comportant de la COLÈRE comportementale, de la PEUR désignée et suggérée, ainsi que de la TRISTESSE désignée et montrée sera par exemple associées aux vecteurs d'étiquettes suivants :

- étiquettes attendues pour étudier le mode comportemental [0,1,0,0,0,0,0,0,0,0,0] ;
- étiquettes attendues pour étudier le mode désigné [0,0,0,0,0,0,0,0,1,0,1] ;
- étiquettes attendues pour étudier le mode montré [0,0,0,0,0,0,0,0,0,0,1] ;
- étiquettes attendues pour étudier le mode suggéré [0,0,0,0,0,0,0,0,0,0,1].

Un cinquième vecteur est associé à chaque phrase pour représenter les classes de catégories émotionnelles prédites le modèle E.

À partir de l'ensemble de ces vecteurs, les mesures de R, de P et de F1 peuvent être calculées pour chaque classe de catégorie émotionnelle, selon le mode d'expression.

Observation 2 : Impact de la catégorie émotionnelle sur la prédiction du mode. Le principe permettant d'observer ce phénomène est le même que celui mis en œuvre pour observer l'impact du mode sur la catégorie émotionnelle, mais inversé. Les informations émotionnelles sont donc cette fois observées catégorie émotionnelle par catégorie émotionnelle. Toutes les phrases du *test* sont ainsi associées à onze vecteurs d'étiquettes attendues – un vecteur pour chaque catégorie émotionnelle – de dimension 4, où chacune des positions du vecteur correspond à un mode d'expression, selon la liste de classe : « comportementale », « désignée », « montrée » et « suggérée ». Une phrase comportant de la COLÈRE comportementale, de la PEUR désignée et suggérée, ainsi que de la TRISTESSE désignée et montrée sera par exemple associée aux vecteurs d'étiquettes suivants :

- étiquettes attendues pour étudier la catégorie ADMIRATION [0,0,0,0] ;
- étiquettes attendues pour étudier la catégorie COLÈRE [1,0,0,0] ;
- étiquettes attendues pour étudier la catégorie PEUR [0,1,0,1] ;
- étiquettes attendues pour étudier la catégorie TRISTESSE [0,1,1,0] ;
- etc.

Un douzième vecteur est associé à chaque phrase pour représenter les classes de mode prédites par le modèle E.

À partir de l'ensemble de ces vecteurs, les mesures de R, de P et de F1 peuvent être calculées pour chaque classe de mode, selon la catégorie émotionnelle.

Annexe XI Tableaux de résultats complémentaires de l'évaluation perceptuelle

Le tableau 95 indique, pour chaque tâche évaluée, le nombre d'étiquettes pour lesquelles au moins deux évaluateurs étaient « d'accord », « plutôt d'accord », « pas d'accord » ou « plutôt pas d'accord » avec la valeur d'annotation proposée, ainsi que celles pour lesquelles aucune convergence d'opinion ne peut être dégagée.

	Nb d'étiquettes	
	cat. émo.	**mode**
d'accord	182	204
plutôt d'accord	4	0
pas d'accord	12	5
plutôt pas d'accord	0	0
aucune convergence d'opinion	26	9
Total	224	218

Tableau 95: Différents avis des évaluateurs, par nombre d'occurrences

Le tableau 96 indique le nombre d'étiquettes avec accord ou désaccord, pour chaque catégorie émotionnelle évaluée par les experts, selon la source de la valeur proposée.

Opinion des évaluateurs	**Valeur proposée**	**Source de la valeur proposée**	**Nb étiquettes**
D'accord	aucune	humain & machine	47
	admiration	humain & machine	2
		machine	1
		humain	5
	autre	humain & machine	3
		machine	8
		humain	3
	colère	humain & machine	10
		machine	6
		humain	4
	culpabilité	humain	4
	dégoût	humain	7
	embarras	humain & machine	3
		machine	1
		humain	7
	fierté	humain & machine	2
		machine	1
		humain	2
	jalousie	humain	1
	joie	humain & machine	10
		machine	5
		humain	6
	peur	humain & machine	10
		machine	10
		humain	7
	surprise	humain & machine	10
		machine	5
		humain	7
	tristesse	humain & machine	8
		machine	2
		humain	5
Pas d'accord	aucune	humain & machine	3
	autre	humain & machine	1
		machine	2
	dégoût	humain	1
	embarras	machine	2
		humain	1
	joie	machine	1
		humain	1
	peur	humain & machine	1
		machine	1
		humain	1
	surprise	machine	2
	tristesse	machine	4
		humain	1

Tableau 96: Nombre d'étiquettes avec accord ou désaccord, pour chaque valeur de catégorie émotionnelle, selon la source de la valeur proposée

Le tableau 97 indique le nombre d'étiquettes avec accord ou désaccord, pour chaque mode d'expression évalué par les experts, selon la source de la valeur proposée.

Opinion des évaluateurs	**Valeur proposée**	**Source de la valeur proposée**	**Nb étiquettes**
D'accord	aucun	humain & machine	47
	comportemental	humain & machine	18
		machine	9
		humain	13
	désigné	humain & machine	23
		machine	6
		humain	10
	montré	humain & machine	24
		machine	3
		humain	3
	suggéré	humain & machine	17
		machine	19
		humain	15
Pas d'accord	aucun	humain & machine	3
	comportemental	machine	2
	désigné	machine	1
	montré	machine	1
		humain	3
	suggéré	humain	1

Tableau 97: Nombre d'étiquettes avec accord ou désaccord, pour chaque valeur de mode, selon la source de la valeur proposée

Annexe XII Consignes du protocole d'évaluation perceptuelle

Expression des émotions dans les textes – expérience de validation par des experts Consignes

La présente expérience vise à faire valider par des experts humains des annotations émotionnelles produites soit par des annotateurs humains, soit par un système d'analyse automatique.

A) Principes généraux de l'évaluation

L'objectif de l'expérience est de dire, pour chaque phrase qui vous sera présentée, si vous êtes d'accord avec les valeurs d'annotation proposées.

Deux types d'informations sont évalués :

1) **la catégorie émotionnelle :** quelle est l'émotion exprimée par la phrase ? Dans le cadre de l'expérience, nous considérons 11 catégories émotionnelles possibles, plus une catégorie AUTRE correspondant à des émotions sur lesquelles notre étude ne se focalise pas. Chaque catégorie est à entendre comme une macro-catégorie, regroupant des noms d'émotion plus fins, comme illustré de manière non-exhaustive ci-dessous :
   - ADMIRATION : *émerveillement*
   - COLÈRE : *agacement, contestation, désaccord, désapprobation, énervement, fureur, rage, indignation, insatisfaction, irritation, mécontentement, réprobation, révolte, etc.*
   - CULPABILITÉ :
   - DÉGOÛT : *lassitude, répulsion, etc.*
   - EMBARRAS : *gêne, honte, humiliation, timidité, etc.*
   - FIERTÉ : orgueil, etc.
   - JALOUSIE :
   - JOIE : *amusement, enthousiasme, exaltation, plaisir, etc.*
   - PEUR : *angoisse, appréhension, effroi, horreur, inquiétude, méfiance, stress, timidité, etc.*
   - SURPRISE : *étonnement, stupeur, etc.*
   - TRISTESSE : *blues, chagrin, déception, désespoir, peine, souffrance, etc.*
   - AUTRE : *amour, courage, curiosité, désir, détermination, envie, espoir, haine, impuissance, mépris, soulagement, etc.*

2) **le mode d'expression :** si une émotion est présente dans la phrase, par quels moyens linguistiques est-elle exprimée ? Dans le cadre de l'expérience, nous retenons 4 types de modes d'expression :
   - **comportemental :** une émotion comportementale est inférée à partir de la description d'une manifestation physiologique d'une émotion ou d'un comportement émotionnel.
     Par exemple dans la phrase « Paul pleure. », la TRISTESSE de Paul peut être inférée à partir de « pleure ». Dans la phrase « Luc claqua sa main sur la table. », la COLÈRE de Luc peut être inférée à partir de « claqua sa main sur la table » ;
   - **désigné :** une émotion est désignée quand elle est directement exprimée par un terme du lexique émotionnel.
     Par exemple dans la phrase « Marc est heureux. », la JOIE de Marc est désignée par le terme « heureux » ;
   - **suggéré :** une émotion suggérée est inférée à partir de la description d'une situation conventionnellement associée à un ressenti émotionnel.
     Par exemple dans la phrase « Léo a gagné la course. », la JOIE de Léo peut être inférée à partir de « a gagné la course » ;
   - **montré :** une émotion montrée est inférée à partir de caractéristiques linguistiques d'un énoncé qui montrent que le locuteur ressentait une émotion au moment de l'énonciation. Ces caractéristiques sont très variées et peuvent par exemple être lexicales (*ex.* interjections), syntaxiques (*ex.* énoncés averbaux) ou typographiques (*ex.* point d'exclamation). Pour être associé à une catégorie émotionnelle précise, un marqueur qui montre une émotion doit nécessairement se trouver en présence d'autres marqueurs d'émotion, qu'ils soient comportementaux, désignés ou suggérés.
     Par exemple, dans la phrase « Zoé s'exclama : "Oh ! Quel bonheur ! Enfin en vacances !" », la JOIE de Zoé peut être inférée à partir de l'interjection « oh », des phrases averbales « quel bonheur » et « enfin en vacances » et des points d'exclamation.

B) Description du matériel expérimental

Le matériel en quelques mots

Le matériel expérimental est constitué d'un corpus de 150 phrases différentes, présentées dans un tableur comportant 2 feuilles :

1. la feuille « Eval_cat_emo » contient, pour chacune des phrases du corpus, des informations d'annotation relatives à la catégorie émotionnelle. L'évaluation de ces annotations s'effectue directement sur cette feuille ;
2. la feuille « Eval_mode » contient, pour chacune des phrases du corpus, des informations d'annotation relatives au mode d'expression. L'évaluation de ces annotations s'effectue directement sur cette feuille.

Dans ces feuilles, chaque ligne correspond à une phrase à évaluer.

Déroulé de l'expérience

L'évaluation doit se faire phrase par phrase, par des allers-retours entre la feuille « Eval_cat_emo » et la feuille « Eval_mode ». Pour une phrase donnée, il est impératif de commencer par analyser les annotations proposées pour la catégorie émotionnelle, puis d'analyser celles proposées pour le mode d'expression.

L'ensemble des 150 phrases du corpus ne doit pas nécessairement être traité en une seule session.

Présentation de la feuille « Eval_cat_emo » – évaluation des annotations de catégorie émotionnelle

Une phrase peut dénoter plusieurs catégories émotionnelles. Pour une même phrase, plusieurs annotations de catégorie émotionnelle pourront ainsi être proposées. Dans le tableur, cela se réalise par une duplication de la phrase. Un schéma illustrant l'analyse d'un tel cas est présenté en Annexe.

Par exemple, la phrase Ph *est indiquée dans les annotations comme exprimant de la* JOIE *et de la* FIERTÉ. Ph *apparaîtra donc sur deux lignes séparées dans le tableur, l'une indiquant la valeur « Joie » et l'autre la valeur « Fierté ».*

De même, l'évaluateur pourra, s'il le souhaite, proposer plusieurs catégories émotionnelles pour une même phrase.

Pour chaque phrase à évaluer, les informations suivantes sont réparties en colonne sur la feuille « Eval_cat_emo » :

- **« Num » :** numéro, de 1 à 150, associé à la phrase (sorte d'identifiant) ;
- **« Phrase_avant » :** élément de contexte pour la phrase à analyser – la phrase qui précède, dans le texte source, la phrase à analyser. Dans le cas d'une phrase à analyser située en début de texte, cette case peut être vide ;
- **« Phrase » :** la phrase à analyser, *i.e.* la phrase à laquelle sont associées les informations d'annotation proposées dans le tableur ;
- **« Phrase_apres » :** élément de contexte pour la phrase à analyser – la phrase qui suit, dans le texte source, la phrase à analyser. Dans le cas d'une phrase à analyser située en fin de texte, cette case peut être vide ;
- **« Valeur_cat_emo » :** la valeur proposée pour l'annotation de la catégorie émotionnelle. C'est cette valeur qui doit être évaluée. Les valeurs possibles ici sont : « Admiration », « Colere », « Culpabilite », « Degout », « Embarras », « Fierté », « Jalousie », « Joie », « Peur », « Surprise », « Tristesse », « Autre » et « Aucune ».
  → La valeur « Aucune » indique qu'aucune catégorie émotionnelle n'est proposée pour la phrase. La phrase est donc considérée comme non émotionnelle ;
- **« Avis_cat_emo » :** avis de l'évaluateur concernant la valeur d'annotation proposée dans la colonne « Valeur_cat_emo ». Les valeurs possibles sont :
  - « D'accord » : accord total avec l'annotation proposée ;

  - « Plutôt d'accord » : accord partiel avec l'annotation proposée – l'analyse proposée n'est pas celle que l'on aurait spontanément produite ;
  - « Plutôt pas d'accord » : désaccord partiel avec l'annotation proposée – on considère que la phrase exprime plutôt une autre catégorie émotionnelle ;
  - « Pas d'accord » : désaccord total avec l'annotation proposée ;
- **« Proposition_cat_emo » :** proposition(s) de catégorie(s) émotionnelle(s) à associer à la phrase. Cette case doit impérativement être renseignée en cas de désaccord total ou partiel avec l'annotation proposée dans la colonne « Valeur_cat_emo ». Les valeurs possibles ici sont : « Admiration », « Colere », « Culpabilite », « Degout », « Embarras », « Fierté », « Jalousie », « Joie », « Peur », « Surprise », « Tristesse », « Autre » et « Pas d'émotion ».

  → La valeur « Pas d'émotion » est employée par l'évaluateur pour indiquer qu'aucune émotion n'est exprimée dans la phrase (phrase non émotionnelle).

  → L'évaluateur peut proposer plusieurs valeurs de catégories émotionnelles à la fois (*ex.* « Joie » et « Fierte ») ;
- **« Remarque » :** toute remarque éventuelle sur la phrase ou son analyse.

Présentation de la feuille « Eval_mode » – évaluation des annotations du mode d'expression

Une phrase peut mobiliser plusieurs modes pour exprimer une même émotion ou des émotions différentes. Pour une même phrase, plusieurs annotations de mode pourront ainsi être proposées. Dans le tableur, cela se réalise par une duplication de la phrase. Un schéma illustrant l'analyse d'un tel cas est présenté en Annexe.

Par exemple, la phrase Ph *est indiquée dans les annotations comme exprimant de l'émotion désignée et de l'émotion comportementale.* Ph *apparaîtra donc sur deux lignes séparées dans le tableur, l'une indiquant la valeur « désignée » et l'autre la valeur « comportementale ».*

De même, l'évaluateur pourra, s'il le souhaite, proposer plusieurs modes pour une même phrase.

Nous soulignons également que la question de l'identification du mode d'expression n'est à traiter que pour les 11 catégories émotionnelles retenues dans notre étude. Cette question ne s'applique donc pas à la catégorie AUTRE.

Pour chaque phrase à évaluer, les informations suivantes sont réparties en colonne sur la feuille « Eval_mode » :

- **« Num » :** numéro, de 1 à 150, associé à la phrase (sorte d'identifiant) ;
- **« Phrase_avant » :** élément de contexte pour la phrase à analyser – la phrase qui précède, dans le texte source, la phrase à analyser. Dans le cas d'une phrase à analyser située en début de texte, cette case peut être vide ;
- **« Phrase » :** la phrase à analyser, *i.e.* la phrase à laquelle sont associées les informations d'annotation proposées dans le tableur ;

- **« Phrase_apres » :** élément de contexte pour la phrase à analyser – la phrase qui suit, dans le texte source, la phrase à analyser. Dans le cas d'une phrase à analyser située en fin de texte, cette case peut être vide ;
- **« Valeur_mode » :** la valeur proposée pour l'annotation du mode d'expression. C'est cette valeur qui doit être évaluée. Les valeurs possibles ici sont : « Comportementale », « Designee », « Montree », « Suggeree », « Aucun » et « Aucun ».
  → La valeur « Aucun » indique qu'aucun mode d'expression n'est proposée pour la phrase. Cela se produit soit quand la phrase est non émotionnelle, soit quand la phrase exprime uniquement la catégorie émotionnelle AUTRE ;
- **« Avis_mode » :** avis de l'évaluateur concernant la valeur d'annotation proposée dans la colonne « Valeur_mode ». Les valeurs possibles sont :
  - « D'accord » : accord total avec l'annotation proposée ;
  - « Plutôt d'accord » : accord partiel avec l'annotation proposée – l'analyse proposée n'est pas celle que l'on aurait spontanément produite ;
  - « Plutôt pas d'accord » : désaccord partiel avec l'annotation proposée – on considère que la phrase exprime plutôt une autre catégorie émotionnelle ;
  - « Pas d'accord » : désaccord total avec l'annotation proposée ;
- **« Proposition_mode » :** proposition(s) de mode(s) d'expression à associer à la phrase. Cette case doit impérativement être renseignée en cas de désaccord total ou partiel avec l'annotation proposée dans la colonne « Valeur_mode ». Les valeurs possibles ici sont : « Comportementale », « Designee », « Montree », « Suggeree », « Non concernée » et « Pas d'émotion ».
  → La valeur « Non concernée » est employée par l'évaluateur pour indiquer que la phrase à analyser exprime seulement la catégorie émotionnelle AUTRE. La question du mode d'expression de l'émotion n'est donc pas à traiter ici.
  → La valeur « Pas d'émotion » est employée par l'évaluateur pour indiquer qu'aucune émotion n'est exprimée dans la phrase (phrase non émotionnelle).
  → L'évaluateur peut proposer plusieurs valeurs de modes d'expression à la fois (*ex.* « Designee » et « Montree ») ;
- **« Remarque » :** toute remarque éventuelle sur la phrase ou son analyse.

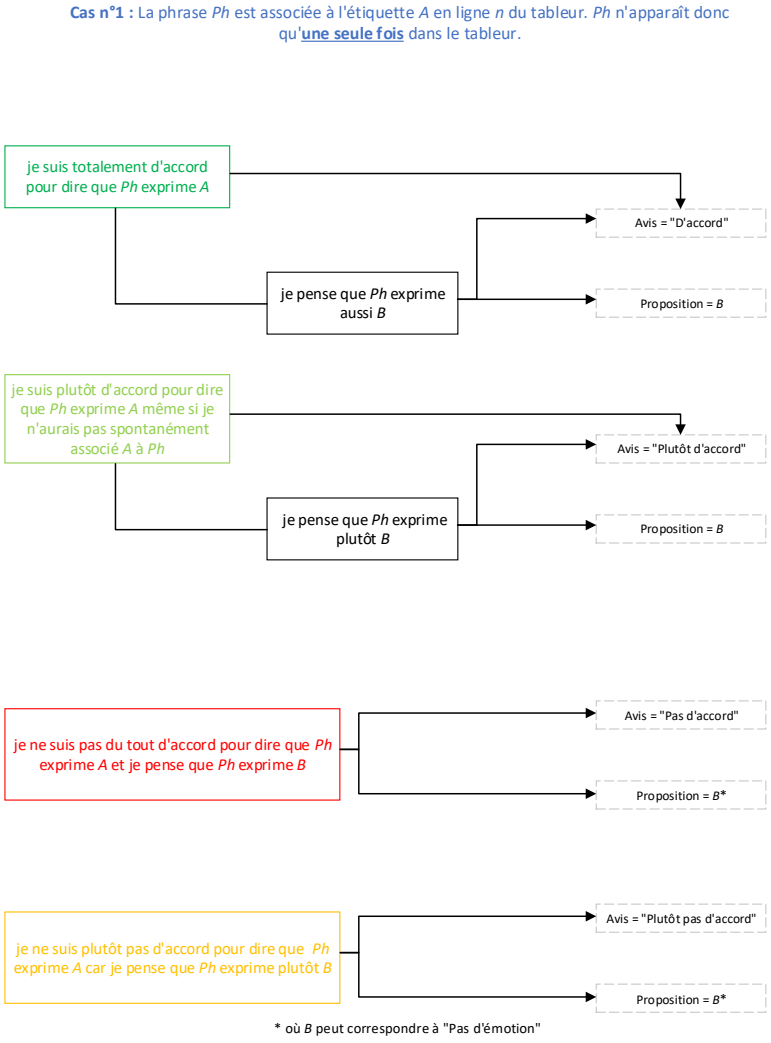
C) Annexe

Cette annexe propose une représentation schématique des procédures de décision déjà décrites Section B. Cette représentation est commune au travail d'analyse à effectuer pour l'évaluation des annotations de catégorie émotionnelle sur la feuille « Eval_cat_emo » et pour l'évaluation des annotations de mode d'expression sur la feuille « Eval_mode ».
La représentation vise à guider l'annotateur dans l'affectation d'une valeur d'« Avis » (colonnes « Avis_cat_emo » et « Avis_mode ») et de « Proposition » (colonnes « Proposition_cat_emo » et « Proposition_mode »). Dans cette représentation, les étiquettes *A*, *B* et *C* représente donc une valeur

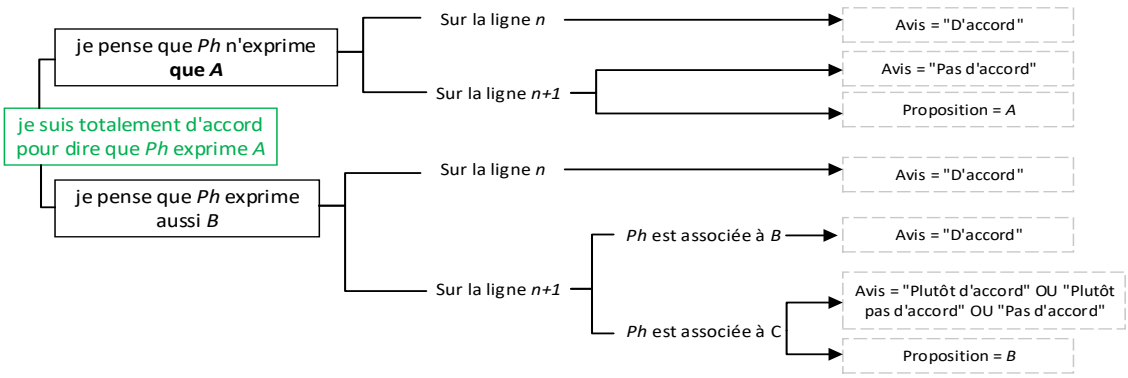
de catégorie émotionnelle ou une valeur de mode d'expression, parmi les valeurs possibles listées Section B.

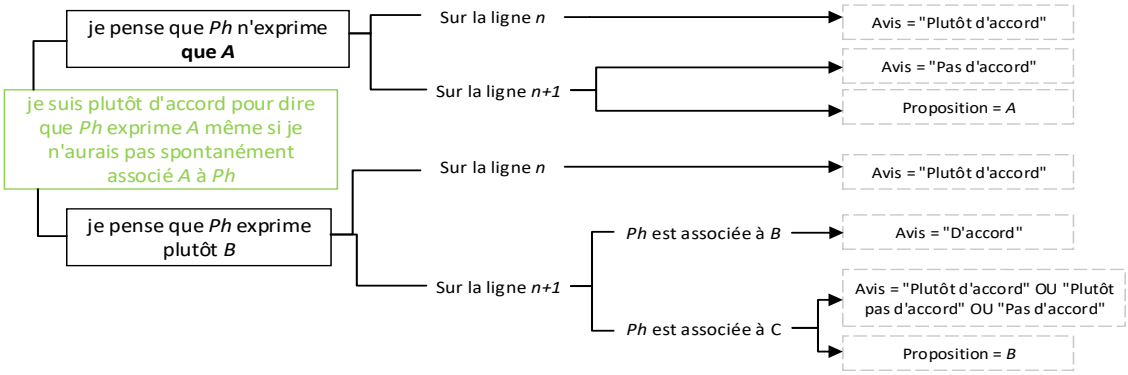
La représentation schématique distingue 2 cas :

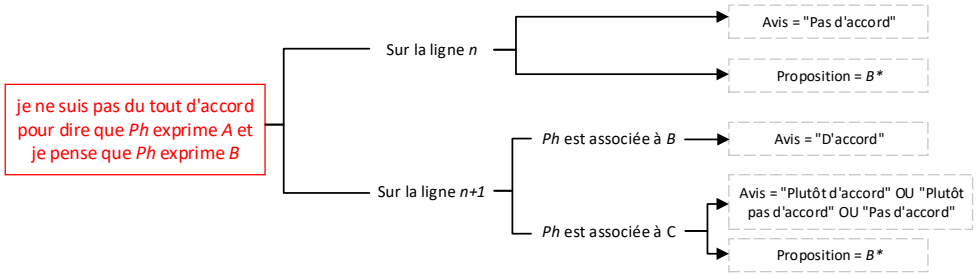
1. celui où, pour une feuille du tableur donnée, la phrase n'est pas dupliquée ;
2. celui où, pour une feuille du tableur donnée, la phrase est dupliquée.

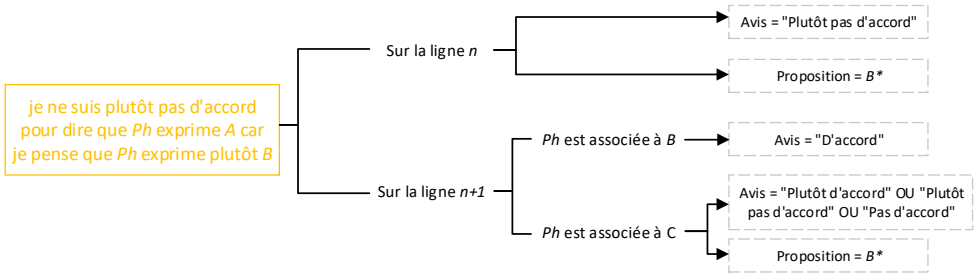


Cas n°2 : La phrase *Ph* est associée à l'étiquette *A* en ligne *n* et à une aute étiquette en ligne *n+1* du tableur. *Ph* apparaît donc **plusieurs fois** dans le tableur.









* où *B* peut correspondre à "Pas d'émotion"

Annexe XIII Compléments pour la comparaison du modèle E' à ÉMOTIONS

1) Comparaison avec le module TextBlob

Les prédictions effectuées par le modèle E' sont booléennes : la prédiction d'un 1 indique la présence de la classe, la prédiction d'un 0 indique son absence. La mise en relation de ces prédictions avec les scores du module TextBlob nécessite donc de convertir les scores en valeurs booléennes. La traduction se fait par la constitution d'un vecteur d'étiquettes à 3 dimensions, où chaque dimension représente l'une des classes servant de point de comparaison entre le modèle E' et TextBlob selon l'ordre « émotionnelle », « positive » et « négative ». De cette manière :

- une phrase non émotionnelle (neutre) est associée au vecteur [0,0,0] ;
- une phrase positive est associée au vecteur [1,1,0] ;
- une phrase négative est associée au vecteur [1,0,1].

La conversion des scores issus du module TextBlob en valeurs booléennes nécessite de fixer un seuil pour assigner la classe « positive » ou « négative » (ou absence de classe « émotionnelle ») aux phrases. Plusieurs méthodes de conversion et plusieurs valeurs de seuil ont été testées sur le *dev* pour identifier la méthode et le seuil permettant d'obtenir les meilleurs résultats.

Méthode 1 : un seul seuil permet de départager les phrases négatives et positives. Pour un seuil de valeur x et un score de phrase noté *score*, le raisonnement est le suivant :

- si $score < x$, la phrase est classée comme « émotionnelle » et « négative » ;
- si $score > x$, la phrase est classée comme « émotionnelle » et « positive » ;
- $score = x$, la phrase est considérée comme non-émotionnelle.

Méthode 2 : un seuil positif et un seuil négatif, ayant la même valeur absolue, sont mobilisés pour départager les phrases négatives et positives. Pour un seuil négatif de valeur $-x$, un seuil positif de valeur x et un score de phrase noté *score*, le raisonnement est le suivant :

- si $score < -x$, la phrase est classée comme « émotionnelle » et « négative » ;
- si $score > x$, la phrase est classée comme « émotionnelle » et « positive » ;
- si $-x \leq score \leq x$, la phrase est considérée comme non-émotionnelle.

Afin de déterminer laquelle des deux méthodes de conversion donne les meilleurs résultats, elles ont toutes deux été mises en œuvre sur le corpus de développement. Il était par ailleurs nécessaire de déterminer, quelque soit la méthode choisie, la valeur de seuil permettant d'obtenir les performances les plus élevées. Pour cela, chacune des deux méthodes a été employée avec plusieurs valeurs de seuil. Les valeurs testées étaient comprises entre −1 et 1 (les bornes des scores prédits par TextBlob), échelonnées selon un pas de 0,2 choisi arbitrairement. Le tableau 98 présente, pour chaque combinaison de méthode et de valeur de seuil, les F-mesures obtenues sur le *dev* pour chacune des trois classes à reconnaître et les macros précision, rappel et F-mesure de l'analyse réalisée par le module TextBlob.

Si le macro-rappel est toujours très bon (plus de 50 % des items sont retrouvés), la macro-précision est toujours très basse (au mieux, un peu moins de 25 % des items correctement classés). La sélection de la meilleure méthode de conversion s'effectue à partir de la macro F-mesure. La méthode de conversion retenue pour comparer les performances du module TextBlob au modèle E' est la méthode 1 (un seul seuil pour départager les phrases positives et négatives), avec une valeur de seuil à 0,4.

Le tableau 99 montre les performances du module TextBlob et du modèle E' pour la reconnaissance des classes « émotionnelle », « positive » et « négative » sur l'ensemble du corpus de test (*test vus* et *test non vus* confondus). Les performances de E' sont toujours bien supérieures à celles du module TextBlob.

Méthode	**Seuil**	**macro P**	**macro R**	**macro F**	**F-mesure pour la classe...** émo.	pos.	nég.
méthode 1	-1	0,063	0,666	0,112	0,273	0,062	0
	-0,8	0,130	0,667	0,114	0,272	0,062	0,009
	-0,6	0,154	0,675	0,128	0,273	0,063	0,048
	-0,4	0,142	0,683	0,141	0,273	0,064	0,086
	-0,2	0,121	0,703	0,158	0,273	0,066	0,137
	0	0,123	0,553	0,189	0,280	0,083	0,204
	0,2	0,103	0,729	0,180	0,276	0,097	0,165
	0,4	0,124	0,705	0,190	0,275	0,141	0,155
	0,6	0,138	0,682	0,170	0,273	0,085	0,153
	0,8	0,247	0,673	0,157	0,273	0,043	0,153
	1	0,080	0,665	0,142	0,272	0,000	0,153
méthode 2	1	0,063	0,666	0,112	0,273	0,062	0,000
	0,8	0,130	0,667	0,114	0,272	0,062	0,009
	0,6	0,154	0,675	0,128	0,273	0,063	0,048
	0,4	0,142	0,683	0,141	0,273	0,064	0,086
	0,2	0,121	0,703	0,158	0,273	0,066	0,137
	0	0,123	0,553	0,189	0,280	0,083	0,204

Tableau 98: Évaluation de différentes méthodes de conversion des scores du module TextBlob sur le dev

Outil	**macro P**	**macro R**	**macro F**	**F-mesure pour la classe...** **émo.**	**pos.**	**nég.**
Modèle E'	0,702	0,637	0,664	0,740	0,575	0,678
Module TextBlob	0,148	0,700	0,210	0,299	0,163	0,168

Tableau 99: Évaluation du module TextBlob et du modèle E' pour la reconnaissance des classes "émotionnelle", "positive" et "négative" sur l'ensemble du test

La méthode de conversion retenue pour comparer les performances du module TextBlob au modèle E' utilise un seuil de valeur 0,4 pour départager les phrases positives et négatives, de sorte que pour une phrase de score s :

- si $s < 0{,}4$, la phrase est classée comme « négative » ;
- si $s > 0{,}4$, la phrase est classée comme « positive » ;
- si $s = 0{,}4$, la phrase est considérée comme non émotionnelle.

2) Comparaison au module EmotaixRemanié

Le module EmotaixRemanié analyse une phrase tokenisée et lemmatisée dont il ne retient que la liste des lemmes de « mots pleins ». Un vecteur de 24 dimensions est ensuite associé à la phrase : chaque dimension du vecteur correspond à une catégorie émotionnelle du lexique EMOTAIX révisé. Le processeur compte ensuite le nombre de fois que chaque catégorie apparaît dans la liste de lemmes de mots pleins puis divise ce nombre d'occurrences par le nombre de lemmes de mots pleins de la phrase.

Points de comparaison entre le module EmotaixRemanié et le modèle E'. Plusieurs points de comparaison peuvent ainsi être établis entre les analyses du processeur Emotaix et les prédictions du modèle E' :

- **le mode comportemental.** Puisque la catégorie 'comportement' du lexique EMOTAIX remanié contient des termes exprimant des comportements émotionnels, l'analyse que le module EmotaixRemanié est donc comparable avec la prédiction de la classe « comportementale » ;
- **le mode désigné.** Une fois groupées, les analyses opérées par le module EmotaixRemanié sont comparables avec la prédiction de la classe « désignée ». Doivent toutefois être exclues les analyses portant sur la catégorie 'comportement' et sur l'ensemble des catégories correspondant à la classe « autre », pour laquelle aucun mode d'expression n'est spécifié dans les données annotées sur lesquelles le modèle E' a été appris ;
- **la catégorie émotionnelle AUTRE.** Plusieurs catégories du lexique EMOTAIX remanié correspondent à la classe « autre » du modèle E' (*cf.* tableau 99). Une fois groupées, les analyses faites par le module pour ces catégories sont comparables à la prédiction de la classe « autre » ;
- **onze catégories émotionnelles désignées.** Chacune des classes de catégories émotionnelles « admiration », « culpabilité », « colère », « dégoût », « embarras », « fierté », « jalousie », « joie », « peur », « surprise » et « tristesse » du modèle E' a au moins une catégorie qui lui correspond dans le lexique EMOTAIX remanié (*cf.* tableau 99). Puisque le lexique n'inclut que des termes qui désignent les émotions, la comparaison des prédictions du modèle E' pour ces onze classes et les analyses des catégories correspondantes faites par le module EmotaixRemanié doit s'effectuer uniquement sur les émo-

tions désignées du corpus d'évaluation. Sont donc comparées ici, pour une phrase donnée :

  - les informations de référence, constituées par la liste des catégories émotionnelles désignées dans la phrase, parmi les onze classes d'émotion du modèle E' ;
  - les prédictions de E' pour ces onze classes, sous réserve qu'elles soient conjointes à la prédiction de la classe « désignée » ;
  - les analyses du module EmotaixRemanié pour les catégories correspondant à ces onze classes ;
- le caractère émotionnel. Dans sa globalité, le module effectue des analyses prenant en compte des informations émotionnelles relative au mode comportemental (sans spécification de catégorie émotionnelle), au mode désigné pour les onze catégories émotionnelles prise en compte par le modèle E' et à la catégorie AUTRE. Une fois combinées, toutes ces analyses sont globalement comparables à la prédiction de la classe « émotionnelle » par E'.

Transformation des analyses du module EmotaixRemanié. Pour pouvoir être comparées aux prédictions du modèle E', les analyses du module EmotaixRemanié, exprimées sous la forme d'un vecteur de proportions, doivent être traduites en valeurs booléennes. Le principe de conversion retenue est le suivant : chaque catégorie EMOTAIX ayant une proportion non nulle est considérée comme présente dans la phrase. Autrement dit, dès que la phrase contient au moins un terme de la catégorie ciblée, la catégorie est considérée comme présente dans la phrase. Ce principe s'applique pour les tâches de comparaison 1 à 4.

La tâche de comparaison 5 nécessite une conversion des analyses du module EmotaixRemanié légèrement différente. Pour une phrase donnée, les proportions représentées par chaque catégorie analysée par le processeur sont d'abord additionnées afin d'obtenir la proportion de lemmes émotionnels de la phrase. Cette proportion est ensuite comparée à un seuil fixé à l'avance afin de déterminer si la phrase doit être classée comme « émotionnelle » ou non. Le module EmotaixRemanié analysera par exemple l'énoncé [61] comme étant constitué de quatre lemmes de mots pleins (*Mona*, *heureux*, *Virginie* et *sourire*), parmi lesquels seulement deux catégories sont présentes, 'joie' et 'comportement'. Chacune étant représentée par un lemme, elles auront toutes deux

pour proportion 0,25. La proportion de lemmes émotionnels de la phrase est donc égale à 0,5. Si cette proportion est supérieure au seuil fixé, la phrase est jugée « émotionnelle ».

[61] Mona est heureuse pour Virginie et lui sourit.

Pour trouver le seuil permettant d'obtenir les meilleures performances, plusieurs valeurs de seuil, de 0 à 100 avec un pas arbitraire de 2, sont testées sur le *dev*. D'après les résultats obtenus sur le *dev* (voir ci-dessous), la valeur de seuil retenue pour convertir les analyses du processeur en valeurs booléennes est 0.

Détermination d'un seuil pour la conversion des analyses du module EmotaixRemanié en valeurs booléennes.

Pour être comparées à la prédiction de la classe « émotionnelle » par le modèle E', les analyses du module EmotaixRemanié doivent être converties en valeurs booléennes. La première étape consiste à additionner les proportions calculées par le processeur pour chacune des catégories qu'il analyse, afin d'obtenir une proportion de lemmes émotionnels. Cette proportion est ensuite comparée à un seuil pour déterminer si la classe doit être associée à la classe « émotionnelle » ou non.

Des valeurs de seuil allant de 0 à 100, par un pas de 2 ont été testées sur le *dev*. Les résultats obtenus sont présentés tableau 100.

Seuil	R	P	F
0	0,374	0,511	0,432
2	0,374	0,511	0,432
4	0,320	0,520	0,396
6	0,237	0,531	0,328
8	0,167	0,562	0,257
10	0,121	0,570	0,200
12	0,098	0,597	0,169
14	0,078	0,618	0,138
16	0,053	0,622	0,097
18	0,030	0,722	0,057
20	0,016	0,700	0,031
22	0,014	0,667	0,027
24	0,014	0,750	0,027
26	0,011	0,833	0,023
28	0,011	0,833	0,023
30	0,009	0,800	0,018
32	0,009	0,800	0,018
34	0,002	1	0,005
36	0,002	1	0,005
38	0,002	1	0,005
40	0,002	1	0,005
42	0,002	1	0,005
44	0,002	1	0,005
46	0,002	1	0,005
48	0,002	1	0,005
50	0	0	0

Tableau 100: Évaluation du module EmotaixRemanié sur le dev, selon différentes valeurs de seuil pour la conversion des analyses

À partir de 50, les performances sont toujours égales à 0. Deux valeurs de seuil offrent les meilleurs résultats : 0 et 2. C'est la valeur 0 qui est retenue afin de permettre au modèle d'avoir, éventuellement, un meilleur rappel.

Résultats sur le *test* des différentes comparaisons mises en œuvre

Les tableau 102 et 101 présentent respectivement les résultats obtenus sur le *test* par le modèle E' et le module EmotaixRemanié pour les cinq tâches de comparaison et pour l'identification des catégories émotionnelles désignées, les performances sont des macro-mesures.

Tâche	Outil	R	P	F
identification du mode comp.	Modèle E'	0,609	0,711	0,656
	EmotaixRe.	0,021	0,714	0,041
identification du mode désigné	Modèle E'	0,804	0,792	0,798
	EmotaixRe.	0,640	0,496	0,559
identification de la classe « autre »	Modèle E'	0,654	0,553	0,599
	EmotaixRe.	0,389	0,243	0,299
identification des cat. désignées	Modèle E'	0,450	0,500	0,466
	EmotaixRe.	0,481	0,507	0,425
identification de phrases binaires	Modèle E'	0,748	0,732	0,740
	EmotaixRe.	0,384	0,530	0,445

Tableau 102: Performances du modèle E' et du module EmotaixRemanié sur le test pour cinq tâches de comparaison

		test		
Tâche	**Outil**	**R**	**P**	**F**
admiration désignée	Modèle E'	0,471	0,571	0,516
	EmotaixRe.	0,118	0,025	0,042
colère désignée	Modèle E'	0,743	0,619	0,675
	EmotaixRe.	0,657	0,460	0,541
culpabilité désignée	Modèle E'	0	0	0
	EmotaixRe.	0,750	0,214	0,333
dégoût désignée	Modèle E'	0	0	0
	EmotaixRe.	0,200	1	0,333
embarras désignée	Modèle E'	0,250	0,444	0,320
	EmotaixRe.	0,250	0,800	0,381
fierté désignée	Modèle E'	0,583	1	0,737
	EmotaixRe.	0,167	0,167	0,167
jalousie désignée	Modèle E'	0	0	0
	EmotaixRe.	1,000	0,500	0,667
joie désignée	Modèle E'	0,750	0,789	0,769
	EmotaixRe.	0,600	0,610	0,605
peur désignée	Modèle E'	0,861	0,766	0,811
	EmotaixRe.	0,795	0,664	0,724
surprise désignée	Modèle E'	0,733	0,786	0,759
	EmotaixRe.	0,467	0,875	0,609
tristesse désignée	Modèle E'	0,556	0,521	0,538
	EmotaixRe.	0,289	0,260	0,274

Tableau 101: Performances du modèle E' et du module EmotaixRemanié pour la reconnaissance de catégories émotionnelles désignées, sur le test

Résultats classe par classe de l'identification de onze catégories émotionnelles désignées.

Le tableau 103 résume les résultats classe par classe pour l'identification de onze catégories émotionnelles désignées, sur le *test vus* et le *test non vus*.

		test vus			***test non vus***		
Tâche	**Outil**	**R**	**P**	**F**	**R**	**P**	**F**
admiration désignée	Modèle E'	0,471	0,571	0,516			
	EmotaixRe.	0,118	0,034	0,053			
colère désignée	Modèle E'	0,727	0,667	0,696	1	0,333	0,500
	EmotaixRe.	0,697	0,548	0,613	0	0	0
culpabilité désignée	Modèle E'	0	0	0			
	EmotaixRe.	0,750	0,250	0,375			
dégoût désignée	Modèle E'	0	0	0			
	EmotaixRe.	0,200	1	0,333			
embarras désignée	Modèle E'	0,250	0,500	0,333			
	EmotaixRe.	0,250	0,800	0,381			
fierté désignée	Modèle E'	0,444	1	0,615	1	1	1
	EmotaixRe.	0,222	0,200	0,211	0	0	0
jalousie désignée	Modèle E'	0	0	0			
	EmotaixRe.	1	0,500	0,667			
joie désignée	Modèle E'	0,736	0,830	0,780	0,857	0,600	0,706
	EmotaixRe.	0,566	0,577	0,571	0,857	0,857	0,857
peur désignée	Modèle E'	0,845	0,854	0,850	0,920	0,561	0,697
	EmotaixRe.	0,814	0,731	0,771	0,720	0,474	0,571
surprise désignée	Modèle E'	0,727	0,780	0,753	1	1	1
	EmotaixRe.	0,455	0,870	0,597	1	1	1
tristesse désignée	Modèle E'	0,512	0,583	0,545	1	0,333	0,500
	EmotaixRe.	0,244	0,286	0,263	0,750	0,200	0,316

Tableau 103: Performances du modèle E' et du module EmotaixRemanié pour la reconnaissance de catégories émotionnelles désignées, sur le test vus et le test non vus

Annexe XIV Processus de révision du lexique EMOTAIX

Le lexique EMOTAIX (Piolat & Bannour, 2009) comporte 56 catégories émotionnelles, réparties en « thèmes médians », eux-mêmes regroupés en « thèmes majeurs » (*cf.* figure *Représentation sous forme de roue de l'organisation du lexique de l'émotion et de l'affect* ici : https://centrepsycle-amu.fr/outils-recherche/). Ces 56 catégories émotionnelles ont été sélectionnées en s'appuyant sur d'autres travaux en psychologie sur les catégories émotionnelles afin d'avoir un jeu de catégories respectant la contrainte suivante : le jeu de catégories retenu devait être composé de « catégories strictement oppositives » (Piolat & Bannour, 2009, p. 683) du point de vue de la valence. Autrement dit, les 56 catégories d'EMOTAIX sont en fait divisées en 28 catégories négatives et 28 catégories positives, qui s'opposent entre elles. 3 catégories sans valences ont été ajoutées ('surprise', 'impassibilité' et 'émotions non spécifiées').

L'objectif du remaniement du lexique EMOTAIX était à la fois de maximiser les correspondances entre le lexique et les catégories retenues dans le schéma d'annotation des émotions défini dans le cadre du projet TextToKids (ci-après nommées « catégories TTK »), et de vérifier le contenu du lexique. La révision de EMOTAIX s'est donc faite en deux étapes principales :

1) **mise en correspondance entre les catégories EMOTAIX et les catégories TTK**, comme illustré sur les figures 20 et 21. Cette mise en correspondance s'est faite :

a. d'une part en s'appuyant uniquement sur les relations de synonymie ou de correspondance exacte entre le nom d'une catégorie EMOTAIX et le nom d'une des catégories TTK. Par exemple, les catégories EMOTAIX 'colère', 'rage' et 'irritation' sont des synonymes ou correspondent exactement à la catégorie TTK COLÈRE ;

b. d'autre part en vérifiant, terme par terme, quelle était la catégorie la plus appropriée, parmi les catégories TTK ;

c. si aucune correspondance avec une catégorie TTK n'était possible, une autre catégorie était conservée ou créée, en essayant de réduire le plus possible le nombre final de catégories dans le lexique remanié.

À l'issue de cette première étape, le lexique remanié comportait 24 catégories émotionnelles :

- les onze catégories émotionnelles TTK ;
- une catégorie 'comportement' qui contient des termes dénotant un comportement émotionnel mais exprimant une catégorie émotionnelle non spécifiée. Par exemple « pleurer » est un comportement émotionnel pouvant dénoter aussi bien de la TRISTESSE que de la JOIE ou de la FIERTÉ. Cette catégorie est donc directement en lien avec le mode d'expression comportemental, présent dans le schéma d'annotation de TextToKids ;
- douze autres catégories qui semblaient suffisamment pertinentes pour être conservées dans le lexique remanié mais ne coïncidaient pas avec les catégories TTK (*ex.* 'ressentiment', 'amour', 'audace', *etc.*).

Par ailleurs, neuf catégories EMOTAIX, ainsi que les termes qu'elles contenaient, ont été entièrement mises de côté (*ex.* 'patience', 'bon sens', 'éveil', 'santé mentale', *etc.*).

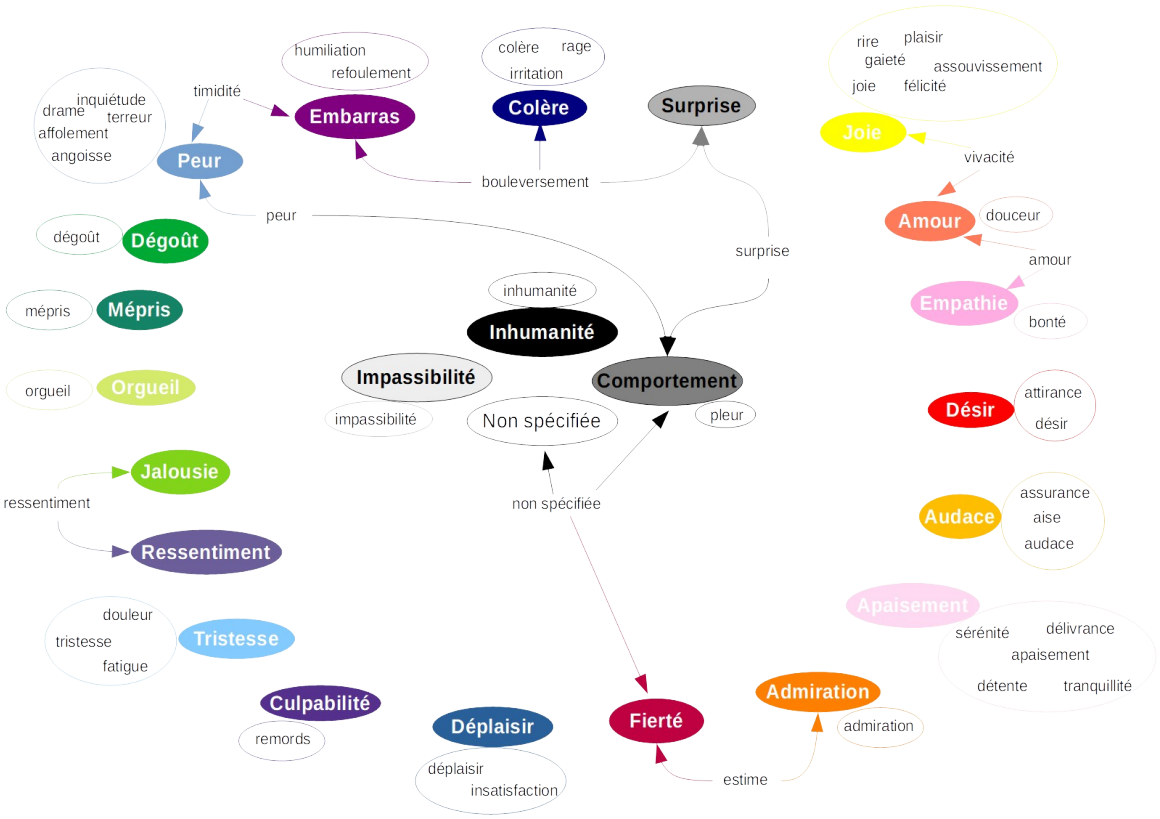


Figure 20: Mise en correspondance entre les catégories EMOTAIX et nos catégories TTK

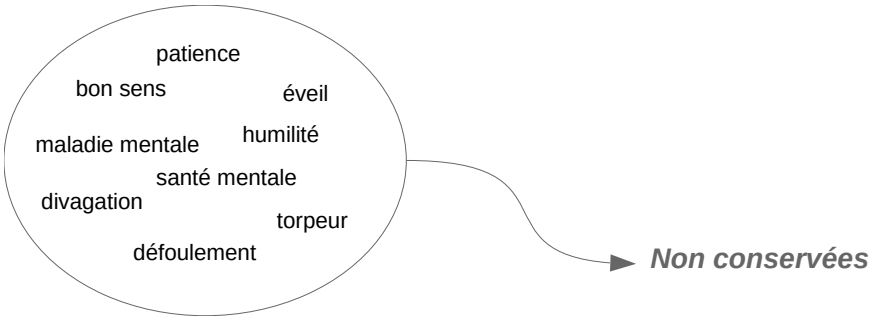


Figure 21: Catégorie EMOTIAX mises de côté

2) **nettoyage des listes de termes**. Pour chacune des catégories du lexique remanié, les termes ont été observés un à un afin de :

a. supprimer les doublons ;

b. vérifier, en s'appuyant sur les critères ci-dessous, que le terme était bien un terme du lexique émotionnel et rangé dans la bonne catégorie.

Afin de déterminer l'appartenance d'un terme au lexique émotionnel, (Piolat & Bannour, 2009, p. 675-676) énoncent des critères explicites, fondés sur la vérification de la définition lexicographique du terme dans le CNRTL :

> « Pour inclure un terme dans notre collection, il devait contenir dans sa définition (section sens propre ou sens figuré) au moins une des caractéristiques sémantiques suivantes : « affection », « caractère », « émotion », « état affectif », « état émotionnel », « état psychologique », « force/faiblesse morale », « humeur », « sentiment », « tempérament », « trait de personnalité », « trouble affectif ».
>
> Nous avons aussi considéré comme étant intégrables dans la collection, les termes dont la définition contenait des mots relevant du lexique émotionnel « classique » (colère, gaîté, tristesse, dégoût, etc.) ou d'états psychologiques agréables ou désagréables (dépression, découragement, euphorie, soulagement, etc.).
>
> Enfin, nous avons intégré dans la collection les termes synonymiques et antonymiques donnés dans cette définition. »

Ce sont ces critères qui sont mobilisés dans le guide accompagnant le schéma d'annotation pour aider les annotateurs à déterminer si une unité émotionnelle est désignée ou non. Ces critères ont tout de même été légèrement modifiés par l'ajout, en plus du CNRTL, de deux dictionnaires (le Wiktionnaire et le Larousse en ligne) comme sources possibles pour la vérification des définitions.

> « Dans le cadre de notre schéma d'annotation, nous considérons qu'un terme appartient au lexique émotionnel si sa définition lexicographique (au sens propre ou au sens figuré, selon le sens mobilisé dans le texte) respecte au moins un des critères ci-dessous, dans au moins un dictionnaire parmi le CNRTL, le Wiktionnaire et le Larousse en ligne :
>
> a) le terme est défini à l'aide d'un des concepts suivants : «affection», «caractère», «émotion», «état affectif», «état émotionnel», «état psychologique», «force/faiblesse morale», «humeur», «sentiment», «tempérament», «trait de personnalité», «trouble affectif» » ;
>
> b) le terme est défini à l'aide de mots appartenant au lexique émotionnel «classique» (colère, gaîté, tristesse, dégoût, etc.) ou référant à des « états psychologiques agréables ou désagréables » (dépression, découragement, euphorie, soulagement, etc.) ;
>
> c) le terme est posé comme synonyme ou antonyme d'un mot remplissant au moins un des critères précédents (i.e. soit le terme est défini par un synonyme/antonyme, respectant lui-même au moins un des deux critères précédents ; soit le terme apparaît en tant que synonyme/antonyme dans la définition d'un mot respectant au moins un des deux critères précédents). Ce critère peut également être vérifié dans le dictionnaire de synonymes numérique CRISCO. » (Étienne & Battistelli, 2021, p. 13)

Publications personnelles

Article dans une revue internationale (1)

1. Battistelli, D., **Étienne, A.**, & Lecorvé, G. (2022). L'émotion à un niveau textuel : La fonction structurante des émotions observée à partir d'annotations. *Discours*. https://doi.org/10.4000/discours.12114

Articles dans des conférences internationales (4)

2. Blandin, A., Lecorvé, G., Battistelli, D., & **Étienne, A.** (2020, mai). Age Recommendation for Texts. *12th Language Resources and Evaluation Conference (LREC)*. https://hal.inria.fr/hal-02868118

3. Rahman, R., Lecorvé, G., Chevelu, J., Béchet, N., **Étienne, A.**, & Battistelli, D. (2020, décembre). Mama/Papa, Is this Text for Me? *International Conference on Computational Linguistics*. https://hal.archives-ouvertes.fr/hal-03112352

4. Dragos, V., Battistelli, D., **Étienne, A.**, & Constable, Y. (2022, juin). Angry or Sad ? Emotion Annotation for Extremist Content Characterization. *13th Language Resources and Evaluation Conference (LREC)*. https://hal.science/hal-03712950

5. **Étienne, A.**, Battistelli, D., & Lecorvé, G. (2022, juin). A (Psycho-)Linguistically Motivated Scheme for Annotating and Exploring Emotions in a Genre-Diverse Corpus. *13th Conference on Language Resources and Evaluation (LREC)*. https://hal.science/hal-03709860

Articles dans des conférences nationales (5)

6. Blandin, A., Lecorvé, G., Battistelli, D., & **Étienne, A.** (2020). Recommandation d'âge pour des textes. In C. Benzitoun, C. Braud, L. Huber, D. Langlois, S. Ouni, S. Pogodalla, & S. Schneider (Éds.), *6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs*

en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 2 : Traitement Automatique des Langues Naturelles (p. 164-171). ATALA. https://hal.science/hal-02784764

7. **Étienne, A.**, Battistelli, D., & Lecorvé, G. (2020). L'expression des émotions dans les textes pour enfants : Constitution d'un corpus annoté. *Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 2 : Traitement Automatique des Langues Naturelles*, 205-212. https://aclanthology.org/2020.jeptalnrecital-taln.19

8. **Étienne, A.**, Battistelli, D., & Lecorvé, G. (2020, novembre). Apports de la linguistique et du TAL à l'analyse des émotions dans les textes pour enfants. *Colloque Langage et éMOTions 3ème Edition*. https://hal.science/hal-03028735

9. Battistelli, D., **Étienne, A.**, Rahman, R., Teissèdre, C., & Lecorvé, G. (2022). Une chaîne de traitement pour prédire et appréhender la complexité des textes pour enfants d'un point de vue linguistique. *Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles. Volume 1 : conférence principale*, 236-246. https://aclanthology.org/2022.jeptalnrecital-taln.23

10. **Étienne, A.**, Battistelli, D., & Lecorvé, G. (2022, juin). Les émotions montrées et comportementales dans un corpus de fictions jeunesse. *Journées d'étude Stylistique outillée et analyse des émotions et sentiments en littérature*. https://hal.science/hal-04102954

Rapport de recherche (1)

11. **Étienne, A.**, & Battistelli, D. (2021). *Annotation manuelle des émotions dans des textes écrits avec la plateforme Glozz*. [Research Report]. MoDyCo ; Université Paris Nanterre. https://hal.archives-ouvertes.fr/hal-03263194

Bibliographie

Abdaoui, A., Azé, J., Bringay, S., & Poncelet, P. (2017). Feel : A french expanded emotion lexicon. *Language Resources and Evaluation*, *51*(3), 833-855.

Aboud, K. S., Bailey, S. K., Del Tufo, S. N., Barquero, L. A., & Cutting, L. E. (2019). Fairy tales versus facts : Genre matters to the developing brain. *Cerebral Cortex*, *29*, 4877-4888. https://doi.org/10.1093/cercor/bhz025

Alm, C. O. (2010). Characteristics of high agreement affect annotation in text. *Proceedings of the fourth linguistic annotation workshop*, 118-122.

Alm, C. O., Roth, D., & Sproat, R. (2005). Emotions from Text : Machine Learning for Text-based Emotion Prediction. *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, 579-586. https://aclanthology.org/H05-1073

Aman, S., & Szpakowicz, S. (2007). Identifying expressions of emotion in text. *International Conference on Text, Speech and Dialogue*, 196-205.

Balahur, A., Hermida, J. M., & Montoyo, A. (2012). Detecting implicit expressions of emotion in text : A comparative analysis. *Decision support systems*, *53*(4), 742-753.

Baretta, L., Tomitch, L. M. B., MacNair, N., Lim, V. K., & Waldie, K. E. (2009). Inference making while reading narrative and expository texts : An ERP study. *Psychology & Neuroscience*, *2*, 137-145. https://doi.org/10.3922/j.psns.2009.2.005

Baron-Cohen, S., Golan, O., Wheelwright, S., & Granader, Y. (2010). Emotion Word Comprehension from 4 to 16 Years Old : A Developmental Survey. *Frontiers in Evolutionary Neuroscience*, *0*. https://doi.org/10.3389/fnevo.2010.00109

Bassano, D. (1985). Procédures de traitement dans la compréhension d'énoncés modalisés chez l'enfant. *Revue française de pédagogie*, *70*(1), 35-40. https://doi.org/10.3406/rfp.1985.1550

Bassano, D., & Champaud, C. (1989). The argumentative connective même in French : An experimental study in eight- to ten-year-old children*. *Journal of Child Language*, *16*(3), 643-664. https://doi.org/10.1017/S0305000900010771

Battistelli, D. (2009). *La temporalité linguistique : Circonscrire un objet d'analyse ainsi que des finalités à cette analyse* [Habilitation à diriger des recherches, Université de Nanterre - Paris X]. https://theses.hal.science/tel-00452464

Battistelli, D., & Amardeilh, F. (2009). Knowledge claims in scientific literature, uncertainty and semantic annotation : A case study in the biological domain. *Semantic Authoring, Annotation and Knowledge Markup Workshop (SAAKM 2009)*.

Battistelli, D., Couto, J., Minel, J.-L., & R. Schwer, S. (2008). Représentation algébrique des expressions calendaires et vue calendaire d'un texte. *TALN 08*, 1-10. https://shs.hal.science/halshs-00275138

Battistelli, D., Étienne, A., & Lecorvé, G. (2022). L'émotion à un niveau textuel : La fonction structurante des émotions observée à partir d'annotations. *Discours*. https://doi.org/10.4000/discours.12114

Battistelli, D., Etienne, A., Rahman, R., Teissèdre, C., & Lecorvé, G. (2022). Une chaîne de traitement pour prédire et appréhender la complexité des textes pour enfants d'un point de vue linguistique (A Processing Chain to Explain the Complexity of Texts for Children From a Linguistic and Psycho-linguistic Point of View). *Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles. Volume 1 : conférence principale*, 236-246. https://aclanthology.org/2022.jeptalnrecital-taln.23

Bevilacqua, M., Blloshmi, R., & Navigli, R. (2021). One SPRING to Rule Them Both : Symmetric AMR Semantic Parsing and Generation without a Complex Pipeline. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(14), 12564-12573. https://doi.org/10.1609/aaai.v35i14.17489

Bianchi, F., Nozza, D., & Hovy, D. (2021). Feel-it : Emotion and sentiment classification for the italian language. *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 76-83.

Blanc, N. (2010). La compréhension des contes entre 5 et 7 ans : Quelle représentation des informations émotionnelles? [The comprehension of the tales between 5 and 7 year-olds: Which representation of emotional information?]. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *64*(4), 256-265. https://doi.org/10.1037/a0021283

Blanc, N., & Quenette, G. (2017). La production d'inférences émotionnelles entre 8 et 10 ans : Quelle méthodologie pour quels résultats ? *Enfance*, *4*(4), 503-511. https://doi.org/10.3917/enf1.174.0503

Blanchard, S., Korach, D., Pencreac'h, J., & Varone, M. (2001). *Le Robert et Nathan Vocabulaire*. Nathan.

Blandin, A., Lecorvé, G., Battistelli, D., & Étienne, A. (2020, mai). Age Recommendation for Texts. *Language Resources and Evaluation Conference (LREC)*. https://hal.inria.fr/hal-02868118

Blything, L. P., Davies, R., & Cain, K. (2015). Young children's comprehension of temporal relations in complex sentences : The influence of memory on performance. *Child Development*, *86*(6), 1922-1934. https://doi.org/10.1111/cdev.12412

Bostan, L.-A.-M., & Klinger, R. (2018). An Analysis of Annotated Corpora for Emotion Classification in Text. *Proceedings of the 27th International Conference on Computational Linguistics*, 2104-2119. https://aclanthology.org/C18-1179

Bronckart, J.-P. (1979). L'élaboration des opérations langagières. Un exemple à propos des structures casuelles. *Cahier de l'Institut de Linguistique de Louvain*, *5*, 139.

Buechel, S., & Hahn, U. (2017). Emobank : Studying the impact of annotation perspective and representation format on dimensional emotion analysis. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, 578-585.

Cardon, R., & Grabar, N. (2021). Simplification automatique de textes biomédicaux en français : Lorsque des données précises de petite taille aident. *Traitement Automatique des Langues Naturelles*, 275-277.

Chatar-Moumni, N. (2013). L'expression verbale des émotions : Présentation. *Langue française*, *180*(4), 3-11. Cairn.info. https://doi.org/10.3917/lf.180.0003

Chen, T., & Guestrin, C. (2016, août). XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/2939672.2939785

Cheng, X., Chen, Y., Cheng, B., Li, S., & Zhou, G. (2017). An emotion cause corpus for chinese microblogs with multiple-user structures. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, *17*(1), 1-19.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, *20*(1), 37-46.

Coltier, D., Dendale, P., & De Brabanter, P. (2009). La notion de prise en charge : Mise en perspective. *Langue française*, *162*(2), 3-27. Cairn.info. https://doi.org/10.3917/lf.162.0003

Creissen, S., & Blanc, N. (2017). Quelle représentation des différentes facettes de la dimension émotionnelle d'une histoire entre l'âge de 6 et 10ans ? Apports d'une étude multimédia. *Psychologie Française*, *62*(3), 263-277. https://doi.org/10.1016/j.psfr.2015.07.006

Dale, E., & Chall, J. S. (1948). A Formula for Predicting Readability. *Educational Research Bulletin*, *27*(1), 11-28.

Davidson, D. (2006). The Role of Basic, Self-Conscious and Self-Conscious Evaluative Emotions in Children's Memory and Understanding of Emotion. *Motivation and Emotion*, *30*(3), 232-242. https://doi.org/10.1007/s11031-006-9037-6

De Belder, J., & Moens, M.-F. (20100101). Text simplification for children. *Prroceedings of the SIGIR Workshop on Accessible Search Systems*, 19-26. https://lirias.kuleuven.be/retrieve/120012

Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A. S., Nemade, G., & Ravi, S. (2020). GoEmotions : A Dataset of Fine-Grained Emotions. *CoRR*, *abs/2005.00547*. https://arxiv.org/abs/2005.00547

Dijkstra, K., Zwaan, R. A., Graesser, A. C., & Magliano, J. P. (1995). Character and reader emotions in literary texts. *Poetics*, *23*(1-2), 139-157.

Diwersy, S., Goossens, V., Grutschus, A., Kern, B., Kraif, O., Melnikova, E., & Novakova, I. (2014). Traitement des lexies d'émotion dans les corpus et les applications d'EmoBase. *Corpus*, *13*, 269-293.

Dobrovolski, D. O., & Bresson, D. (1995). Petite syntaxe de la « peur ». Application au français et à l'allemand. *Langue française*, *105*(1), 107-119. https://doi.org/10.3406/lfr.1995.5297

Dragos, V., Battistelli, D., Etienne, A., & Constable, Y. (2022, juin). Angry or Sad ? Emotion Annotation for Extremist Content Characterization. *13th Language Resources and Evaluation Conference*. https://hal.science/hal-03712950

Dyer, M. G. (1983). The role of affect in narratives. *Cognitive science*, *7*(3), 211-242.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3-4), 169-200. https://doi.org/10.1080/02699939208411068

Elguendouze, S. (2020). Simplification de textes : Un état de l'art (Text simplification (State of the art) This paper presents the state of the art on text simplification, in particular the two main types of approaches, namely rule-based approaches and statistical approaches (or data-driven)). *Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 3 : Rencontre des Étudiants Chercheurs en Informatique pour le TAL*, 96-109. https://aclanthology.org/2020.jeptalnrecital-recital.8

Étienne, A., & Battistelli, D. (2021). *Annotation manuelle des émotions dans des textes écrits avec la plateforme Glozz.* [Research Report]. MoDyCo ; Université Paris Nanterre. https://hal.archives-ouvertes.fr/hal-03263194

Fleiss, J. L., Levin, B., Paik, M. C., & others. (1981). The measurement of interrater agreement. *Statistical methods for rates and proportions*, *2*(212-236), 22-23.

Flesch, R. (1948). A new readability yardstick. *Journal of applied psychology*, *32*(3), 221.

Flor, M., Klebanov, B. B., & Sheehan, K. (2013). Lexical Tightness and Text Complexity. *Proceedings of the Workshop on Natural Language Processing for Improving Textual Accessibility.*

Forti, L., Milani, A., Piersanti, L., Santarelli, F., Santucci, V., & Spina, S. (2019). Measuring Text Complexity for Italian as a Second Language Learning Purposes. *BEA@ACL.*

Fraisse, A., & Paroubek, P. (2015). Utiliser les interjections pour détecter les émotions. *Actes de la 22e conférence sur le Traitement Automatique des Langues Naturelles. Articles longs*, 279-290.

François, T., & Fairon, C. (2012). An "AI readability" Formula for French as a Foreign Language. *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 466-477. https://aclanthology.org/D12-1043

François, T., Müller, A., Degryse, B., & Fairon, C. (2018). AMesure, une plateforme web pour soutenir la rédaction simple de textes administratifs. *Repères-Dorif, 16.*

Gala, N., François, T., Javourey-Drevet, L., & Ziegler, J. C. (2018). La simplification de textes, une aide à l'apprentissage de la lecture. *Langue francaise*, *199*(3), 123-131.

Gala, N., Todirascu, A., Bernhard, D., Wilkens, R., & Meyer, J. P. (2020). Transformations syntaxiques pour une aide à l'apprentissage de la lecture : Typologie, adéquation et corpus adaptés. *7e Congrès Mondial de Linguistique Française.*

Ghazi, D., Inkpen, D., & Szpakowicz, S. (2015). Detecting emotion stimuli in emotion-bearing sentences. *International Conference on Intelligent Text Processing and Computational Linguistics*, 152-165.

Gosselin, L. (2005). *Temporalité et modalité.* De Boeck Supérieur; Cairn.info. https://www.cairn.info/temporalite-et-modalite--9782801113660.htm

Gosselin, L. (2010). *Les modalités en français : La validation des représentations.* Brill.

Gromer, B., & Weiss, M. (1990). *Lire*. A. Colin.

Gross, M. (1995). Une grammaire locale de l'expression des sentiments. *Langue française*, *105*(1), 70-87. https://doi.org/10.3406/lfr.1995.5294

Gui, L., Xu, R., Wu, D., Lu, Q., & Zhou, Y. (2018). Event-driven emotion cause extraction with corpus construction. In *Social Media Content Analysis : Natural Language Processing and Beyond* (p. 145-160). World Scientific.

Hamon, T., Fraisse, A., Paroubek, P., Zweigenbaum, P., & Grouin, C. (2015, juin). Analyse des émotions, sentiments et opinions exprimés dans les tweets : Présentation et résultats de l'édition 2015 du défi fouille de texte (DEFT). *Actes de la 22e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2015)*. https://hal.archives-ouvertes.fr/hal-01617180

Hickmann, M. (2012). Diversité des langues et acquisition du langage : Espace et temporalité chez l'enfant. *Langages*, *188*(4), 25-39.

Hickmann, M., Champaud, C., & Bassano, D. (1993). Pragmatics and metapragmatics in the development of epistemic modality : Evidence from French children's reports of think-statements. *First Language*, *13*(39), 359-388. https://doi.org/10.1177/014272379301303905

Imperial, J. M., & Ong, E. (2021). Diverse Linguistic Features for Assessing Reading Difficulty of Educational Filipino Texts. *ArXiv*, *abs/2108.00241*.

Kail, M., & Léveillé, M. (1977). Compréhension de la coréférence des pronoms personnels chez l'enfant et l'adulte. *L'Année psychologique*, *77*(1), 79-94. https://doi.org/10.3406/psy.1977.28180

Kerbrat-Orecchioni, C. (2000). Quelle place pour les émotions dans la linguistique du XXe siècle. *Plantin, C., Doury, M. & Traverso, V.(éds), Les émotions dans les interactions. Lyon: Presses Universitaires de Lyon*, 33-74.

Kim, E., & Klinger, R. (2018). Who feels what and why ? Annotation of a literature corpus with semantic roles of emotions. *Proceedings of the 27th International Conference on Computational Linguistics*, 1345-1359.

Kim, E., & Klinger, R. (2019a). An Analysis of Emotion Communication Channels in Fan-Fiction : Towards Emotional Storytelling. *Proceedings of the Second Workshop on Storytelling*, 56-64. https://doi.org/10.18653/v1/W19-3406

Kim, E., & Klinger, R. (2019b). Frowning Frodo, Wincing Leia, and a Seriously Great Friendship : Learning to Classify Emotional Relationships of Fictional Characters. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 647-653. https://doi.org/10.18653/v1/N19-1067

Krippendorff, K. (1995). On the reliability of unitizing continuous data. *Sociological Methodology*, 47-76.

Lecocq, P. (1998). *L' É.co.s.se une épreuve de compréhension syntaxico-sémantique (manuel et épreuve) : Deux volumes*. Presses Univ. Septentrion.

Leeman-Bouix, D. (1995). Pourquoi peut-on dire Max est en colère mais non *Max est en peur ? Hypothèses sur la construction être en N. *Langue française*, *105*(1), 55-69. https://doi.org/10.3406/lfr.1995.5293

Liu, C., Osama, M., & Andrade, A. de. (2019). DENS: A Dataset for Multi-class Emotion Analysis. *CoRR*, *abs/1910.11769*. http://arxiv.org/abs/1910.11769

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). Roberta : A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Mahamdi, M. (2022). *Analyse de l'interaction entre émotions et deux dimensions sémantiques : Les adverbiaux temporels et les modalités* [Rapport de stage]. Université Paris Nanterre.

Martin, L., Muller, B., Ortiz Suárez, P. J., Dupont, Y., Romary, L., de la Clergerie, É., Seddah, D., & Sagot, B. (2020). CamemBERT: a Tasty French Language Model. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7203-7219. https://www.aclweb.org/anthology/2020.acl-main.645

Mathet, Y., & Widlöcher, A. (2016). Évaluation des annotations : Ses principes et ses pièges. *Revue TAL*, *57*(2), 73-98.

Mathet, Y., Widlöcher, A., & Métivier, J.-P. (2015). The Unified and Holistic Method Gamma (γ) for Inter-Annotator Agreement Measure and Alignment. *Computational Linguistics*, *41*(3), 437-479. https://doi.org/10.1162/COLI_a_00227

Mathieu, Y. Y. (2005). Annotation of Emotions and Feelings in Texts. *Proceedings of the First International Conference on Affective Computing and Intelligent Interaction*, 350-357. https://doi.org/10.1007/11573548_45

Micheli, R. (2014). *Les émotions dans les discours. Modèle d'analyse, perspectives empiriques*. De Boeck Supérieur. https://www.cairn.info/les-emotions-dans-les-discours--9782801117385.htm

Micheli, R., Hekmat, I., & Rabatel, A. (2013). Les émotions : Des modes de sémiotisation aux fonctions argumentatives. *Semen. Revue de sémio-linguistique des textes et discours*, *35*.

Mohammad, S. (2012). #Emotional Tweets. **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, 246-255. https://aclanthology.org/S12-1033

Mohammad, S., Bravo-Marquez, F., Salameh, M., & Kiritchenko, S. (2018). SemEval-2018 Task 1 : Affect in Tweets. *Proceedings of The 12th International Workshop on Semantic Evaluation*, 1-17. https://doi.org/10.18653/v1/S18-1001

Mohammad, S., & Kiritchenko, S. (2018, mai). Understanding Emotions : A Dataset of Tweets to Study Interactions between Affect Categories. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. https://aclanthology.org/L18-1030

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational intelligence*, *29*(3), 436-465.

Neviarouskaya, A., & Aono, M. (2013). Extracting causes of emotions from text. *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, 932-936.

Öhman, E. (2020). Emotion annotation : Rethinking emotion categorization. *CEUR Workshop Proceedings*, *2865*, 134-144.

Öhman, E., Pàmies, M., Kajava, K., & Tiedemann, J. (2020). XED: A Multilingual Dataset for Sentiment Analysis and Emotion Detection. *Proceedings of the 28th International Conference on Computational Linguistics*, 6542-6552. https://doi.org/10.18653/v1/2020.coling-main.575

Oltéan, D. (2021). *Le rôle des adverbiaux temporels dans l'intensification des émotions* [Mémoire de recherche].

Paroubek, P., Pak, A., & Mostefa, D. (2010, mai). Annotations for Opinion Mining Evaluation in the Industrial Context of the DOXA project. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.
http://www.lrec-conf.org/proceedings/lrec2010/pdf/422_Paper.pdf

Piolat, A., & Bannour, R. (2009). EMOTAIX : Un scénario de Tropes pour l'identification automatisée du lexique émotionnel et affectif. [EMOTAIX: A scenario using Tropes for automatized identification of emotional and affective lexicon.]. *L'Année Psychologique*, *109*(4), 655-698. https://doi.org/10.4074/S0003503309004047

Piolat, A., Booth, R. J., Chung, C. K., Davids, M., & Pennebaker, J. W. (2011). La version française du dictionnaire pour le LIWC : modalités de construction et exemples d'utilisation. *Psychologie Française*, *56*(3), 145-159. https://doi.org/10.1016/j.psfr.2011.07.002

Plank, B. (2022). *The « Problem » of Human Label Variation : On Ground Truth in Data, Modeling and Evaluation.*

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In *Theories of emotion* (p. 3-33). Elsevier.

Pyykkönen, P., & Järvikivi, J. (2012). Children and situation models of multiple events. *Developmental Psychology*, *48*(2), 521-529. https://doi.org/10.1037/a0025526

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza : A Python Natural Language Processing Toolkit for Many Human Languages. *Proceedings*

of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. https://nlp.stanford.edu/pubs/qi2020stanza.pdf

Rabatel, A. (2013). Ecrire les émotions en mode emphatique. *Semen. Revue de sémio-linguistique des textes et discours*, *35*.

Rahman, R., Lecorvé, G., Chevelu, J., Béchet, N., Étienne, A., & Battistelli, D. (2020, décembre). Mama/Papa, Is this Text for Me? *International Conference on Computational Linguistics*. https://hal.archives-ouvertes.fr/hal-03112352

Romero, C. (2017). *L'intensité et son expression en français*. Ophrys.

Russo, I., Caselli, T., Rubino, F., Boldrini, E., & Martínez-Barco, P. (2011). EMOCause : An Easy-adaptable Approach to Extract Emotion Cause Contexts. *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*, 153-160. https://aclanthology.org/W11-1720

Sagot, B. (2010, mai). The Lefff, a Freely Available and Large-coverage Morphological and Syntactic Lexicon for French. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. http://www.lrec-conf.org/proceedings/lrec2010/pdf/701_Paper.pdf

Scherer, K. R. (2005). What are emotions ? And how can they be measured? *Social science information*, *44*(4), 695-729.

Schmidt, T., Dennerlein, K., & Wolff, C. (2021). Towards a Corpus of Historical German Plays with Emotion Annotations. *3rd Conference on Language, Data and Knowledge (LDK 2021)*.

Segui, J., & Léveillé, M. (1977). Etude sur la compréhension de phrases chez l'enfant. *Enfance*, *30*(1), 105-115. https://doi.org/10.3406/enfan.1977.2637

Sheehan, K., Kostin, I., & Futagi, Y. (2008). When Do Standard Approaches for Measuring Vocabulary Difficulty , Syntactic Complexity and Referential Cohesion Yield Biased Estimates of Text Difficulty ? *The Elementary School Journal*, *115*(2), 184-209.

Simoës-Perlant, A., & Lemercier, C. (2018). Evaluation du lexique émotionnel chez l'enfant de 8 à 11 ans. *ANAE - Approche Neuropsychologique des Apprentissages Chez L'enfant*, *155*, 417.

Strapparava, C., & Mihalcea, R. (2007). SemEval-2007 Task 14 : Affective Text. *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, 70-74. https://aclanthology.org/S07-1013

Strapparava, C., & Valitutti, A. (2004). WordNet-Affect : An Affective Extension of WordNet. *Vol 4.*, *4*.

Tamine, J., & Bonnet, C. (1982). La compréhension des métaphores chez les enfants. Une hypothèse et quelques implications pédagogiques. *L'information grammaticale*, *14*(1), 17-22. https://doi.org/10.3406/igram.1982.2352

Tartas, V. (2001). The development of systems of conventional time : A study of the appropriation of temporal locations by four-to-ten-year old children. *European Journal of Psychology of Education*, *16*(2), 197-208. https://doi.org/10.1007/BF03173025

Tartas, V. (2010). Le développement de notions temporelles par l'enfant. *Developpements*, *4*(1), 17-26.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words : LIWC and computerized text analysis methods. *Journal of language and social psychology*, *29*(1), 24-54.

Teissedre, C. (2012). *Analyse sémantique automatique des adverbiaux de localisation temporelle : Une application à la recherche d'information et à l'acquisition de connaissances.*

Ters, F., Reichenbach, D., & Mayer, G. (1988). *L'échelle Dubois-Buyse d'orthographe usuelle française.* éd. MDI.

Vajjala, S., & Meurers, W. D. (2014). Readability assessment for text simplification : From analysing documents to identifying sentential simplifications. *ITL – International Journal of Applied Linguistics*, *165*, 194-222.

van den Broek, P. (1997). Discovering the cement of the universe : The development of event comprehension from childhood to adulthood. In *Developmental spans*

in event comprehension and representation : Bridging fictional and actual events (p. 321-342). Lawrence Erlbaum Associates, Inc.

Vion, M., & Amy, G. (1984). Comprendre les relations agent-patient dans les énoncés simples en français, une étude génétique du traitement des structures clivées. *Archives de Psychologie*, *52*, 209-229.

Widlöcher, A., & Mathet, Y. (2012). The glozz platform : A corpus annotation and mining tool. *Proceedings of the 2012 ACM symposium on Document engineering*, 171-180.

Wilkens, R., Alfter, D., Wang, X., Pintard, A., Tack, A., Yancey, K. P., & François, T. (2022). FABRA: French Aggregator-Based Readability Assessment toolkit. *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 1217-1233. https://aclanthology.org/2022.lrec-1.130

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, *123*(2), 162.