Université Paris Cité

École doctorale Sciences du Langage (ED 622)

Centre de Linguistique Inter-langues, de Lexicologie, de Linguistique Anglaise et de Corpus-Atelier de Recherche sur la Parole (UR 3967)

L'impact des discours numériques sur l'accès aux droits pour les citoyens

Analyse outillée et contribution à la caractérisation linguistique des discours administratifs en ligne en France et au Royaume-Uni

Par Marie BOUCHET

Thèse de doctorat de SCIENCES DE LA TRADUCTION

Dirigée par Mojca PECMAN

Présentée et soutenue publiquement le 26 novembre 2024

Devant un jury composé de :

Anne CONDAMINES, Directrice de recherche, Université Toulouse - Jean Jaurès, Rapporteure

Pascaline DURY, Professeure, Université Lumière Lyon 2, Rapporteure

Patrick DROUIN, Professeur, Université de Montréal, Examinateur

Christopher GLEDHILL, Professeur, Université Paris Cité, Président du jury

Natalie KÜBLER, Professeure, Université Paris Cité, Examinatrice

Philippe MILLOT, Professeur, Université Lumière Lyon 2, Examinateur

Mojca PECMAN, Professeure, Université Paris Cité, Directrice de thèse

Résumé

Titre : L'impact des discours numériques sur l'accès aux droits pour les citoyens : analyse outillée et contribution à la caractérisation linguistique des discours administratifs en ligne en France et au Royaume-Uni

Amorcée dans les années 2000, la dématérialisation des démarches administratives est désormais achevée : l'accès aux droits pour les citoyens se fait désormais sur les plateformes numériques. Dès lors, le discours administratif devient numérique et les rituels d'interactions entre les administrations et leurs usagers sont profondément modifiés et recentrés sur un média : le site web. Cette thèse a pour objectif de participer à la caractérisation du discours numérique de l'accès aux droits. Pour ce faire, nous avons constitué un corpus bilingue à partir de sites d'accès aux droits officiels français et britanniques. Ce corpus a été exploré par démarche inductive, avec l'outil d'analyse de texte Sketch Engine, afin de réaliser une caractérisation lexicale, syntaxique et rhétorique. Ces analyses ont dévoilé une variation terminologique ainsi qu'une instabilité des dispositifs d'aide sociale et de droit au séjour. D'autres phénomènes observés, dus au caractère numérique du discours, tels que la présence de schémas lexico-syntaxiques répétés et de technosignes, nous ont conduits à élaborer de nouvelles méthodes d'analyse. Afin d'intégrer la dimension numérique et de recréer un fil discursif, nous avons constitué un corpus spécialisé composé de scénarios de navigation que nous avons annoté manuellement pour dégager les différentes fonctions discursives et les éléments interactifs. L'ensemble de ces analyses ont mené à la rédaction de recommandations méthodologiques pour les recherches en corpus sur les discours administratifs numériques. Dans un deuxième temps, nous avons étudié ces sites web en tant que genre discursif, en questionnant les phénomènes de spécialisation et de vulgarisation, afin de réenvisager la place de l'interactivité dans la compréhension et la diffusion du discours. Ces conclusions ont mené à la rédaction des recommandations à destination des experts produisant le discours de l'accès aux droits numérique.

Mots-clés : analyse du discours, discours numérique, linguistique de corpus, termi-

nologie, phraséologie, accès aux droits, genre discursif, langue de spécialité

Abstract

Title : The impact of digital discourse on citizens' access to rights : a tool-based analysis and contribution to the characterization of online administrative discourse in France and the United Kingdom

In the past two decades, administration services have shifted to online communication. Administrative information is now mainly available through websites and published by administration internal services. Websites are the main way to access information and play a crucial role in ensuring citizens access to their rights. This thesis aims to contribute to the characterisation of administrative online discourse. To conduct this study, we built a bilingual corpus of French and British online administrative websites. This corpus was explored using an inductive approach, with the Sketch Engine text analysis tool, in order to carry out a lexical, syntactic and rhetorical characterisation. Corpus analysis revealed terminological variation and instability in French welfare and immigration schemes, as well as in British welfare schemes. After identifying repeated lexical-syntactic patterns and technolinguistic elements, we designed a complementary method to study this discourse. In order to include the digital features and redefine threads of discourse, we built a specialised corpus based on user paths, which was manually annotated with discursive functions and interactive elements. These findings were integrated into methodological guidelines aimed at helping corpus research on digital administrative discourse. We then studied these websites as a discourse genre, questioning discourse specialisation and accessibility. We examined the role interactive elements and hyperlinks play in creating a coherent thread of discourse. These findings led to a set of guidelines aimed at helping experts in producing online administrative discourse.

Keywords : discourse analysis, digital discourse, corpus linguistics, terminology, phraseology, administrative discourse, language for specific purposes

Remerciements

Je tiens tout d'abord à remercier ma directrice, Mojca Pecman, pour son accompagnement bienveillant, depuis la conception du projet de thèse jusqu'à sa finalisation. Son expertise, ses intuitions et ses conseils avisés m'ont permis de réaliser ce projet de recherche, mais aussi de m'épanouir sur le plan académique. La réalisation de ce projet de thèse en trois ans n'aurait pas été possible sans sa réactivité et sa disponibilité.

Je souhaite également remercier Natalie Kübler, directrice du CLILLAC-ARP qui m'a accueillie au sein du laboratoire. Ses précieux conseils et le cadre de travail agréable et bienveillant mis à disposition des doctorants m'ont permis de mener à bien mes recherches dans de bonnes conditions.

J'adresse ensuite mes remerciements à Christopher Gledhill pour sa disponibilité, son expertise et ses retours qui m'ont été d'une grande aide pour améliorer la qualité de mes travaux.

Je tiens également à remercier Pascaline Dury pour son écoute attentive et ses retours lors des comités de suivi, ainsi que pour avoir accepté d'être rapporteure de cette thèse. J'adresse mes remerciements à Anne Condamines, qui a accepté d'être rapporteure de cette thèse, ainsi qu'à Patrick Drouin, Christopher Gledhill, Philippe Millot et Natalie Kübler pour l'honneur qu'ils m'ont fait en acceptant d'examiner ce travail.

De manière plus générale, je remercie toute l'équipe du laboratoire pour leur écoute et leur bienveillance. Je souhaite particulièrement remercier les doctorant.es du laboratoire, ceux qui m'ont accueillie à mon arrivée et ceux qui nous ont rejoints en cours de route. Je remercie notamment Anastasia Buturlakina et Maud Bénard qui ont été avec moi lors de ces trois années de thèse. Merci pour tous ces moments partagés, pour les discussions enrichissantes et pour votre soutien qui ont rendu ces trois années agréables et stimulantes.

Je remercie ensuite Brice et l'équipe du centre social 13 pour tous pour leur accueil lors des permanences d'écrivain public. C'est lors de ces permanences et des discussions avec les usagers que j'ai pu me rendre compte à quel point le discours de l'accès aux droits est important dans notre société.

Enfin, je souhaite remercier ma famille et mes amis pour leur soutien et leur encou-

ragement. Un grand merci à Cédric, Masha et Claire pour leurs relectures avisées.

Pour terminer, je remercie Amine Benamara, pour son expertise informatique, ses précieux conseils, sa bienveillance, ses relectures attentives et ses encouragements constants. Merci à Spyro d'avoir été là pendant ces longues heures de rédaction, de m'avoir fait rire avec tes bêtises et de m'avoir forcé à mettre le nez dehors tous les jours.

Table des matières

Introduction **1**

0.1 Présentation du sujet 1
0.2 Présentation du titre 3
0.3 Étudier un discours numérique interactif 4
0.4 Le discours administratif en tant que langue de spécialité vulgarisée . . 5
0.5 Problématique 6
0.6 Annonce du plan 8

I État de l'art **11**

1 Le discours de l'accès aux droits : contexte historique et situation actuelle **13**

1.1 Accès aux droits en France et au Royaume-Uni 13
1.1.1 Définition de la notion d'*accès aux droits* 14
1.1.2 L'accès aux droits en France 16
1.1.3 L'accès aux droits au Royaume-Uni 19
1.1.4 France et Royaume-Uni, une vision divergente 23
1.2 Une double transformation : numérique et idéologique 23
1.2.1 État plateforme : une idéologie aux origines américaines 24
1.2.2 Conditionnalité 27
1.2.3 Charge administrative - *Administrative burden* 30
1.3 Paysage documentaire du discours numérique de l'accès aux droits . . . 33
1.3.1 France 33
1.3.2 Royaume-Uni 37
1.4 Synthèse 39

2 Un discours spécialisé et vulgarisé **43**

2.1 Discours et genres discursifs 43
2.1.1 Définitions 43
2.1.1.1 Analyse du discours 43

2.1.1.2 Le discours 44
2.1.2 Genre de discours 46
2.1.3 Textes procéduraux 49
2.2 Un discours spécialisé 50
2.2.1 Définition de la notion de *discours spécialisé* 50
2.2.2 Définition de la notion de *communauté de discours* 52
2.2.3 Qui sont les experts de l'accès aux droits 53
2.2.4 Les niveaux de compétences de l'usager 55
2.2.5 La communauté de discours de l'accès aux droits 60
2.2.6 *Administrative literacy* 61
2.2.7 Langues de spécialité 63
2.2.7.1 Juridique 63
2.2.7.2 Politique 64
2.2.7.3 Administrative 64
2.2.7.4 Institutionnelle 65
2.3 Un discours vulgarisé 66
2.3.1 Qu'est-ce que la vulgarisation ? 66
2.3.2 Le discours de l'accès aux droits est-il vulgarisé ? 67
2.3.3 La simplification et le langage clair 68
2.3.4 Stratégies de simplification pour le discours de l'accès aux droits 69
2.3.4.1 Royaume-Uni 70
2.3.4.2 France 71
2.3.4.3 Autres outils 73
2.4 Synthèse 74

II Outils méthodologiques et théoriques 77

3 Redéfinir les normes de communication 79
3.1 Émergence d'un nouveau discours 79
3.1.1 De l'électronique au numérique 79
3.1.2 Nouveaux paradigmes de communication 82
3.1.2.1 Les sites web 83
3.2 Le discours numérique 84
3.2.1 Définition 84
3.2.2 Caractéristiques 85
3.2.3 Discours hypertextualisé 86
3.3 Spécificités linguistiques 88

3.3.1 Variation terminologique 89
3.3.2 Groupes nominaux complexes 90
3.3.3 Approche ontologique de la terminologie 91
3.3.4 Sigles et acronymes 92
3.3.5 Redondance/Répétitions 93
3.3.6 Unités phraséologiques 94

4 Méthode **97**
4.1 Choix du sujet 97
4.2 Linguistique de corpus 99
4.2.1 Essor de la linguistique de corpus : ouverture des possibilités . . 99
4.2.2 Approche adoptée 101
4.2.3 Définition d'un corpus 103
4.2.4 Une méthode adaptée à l'étude des langues de spécialité 104
4.2.5 Les corpus issus du web dynamique 105
4.3 Création d'un corpus spécialisé 109
4.3.1 Définition des critères pour la constitution du corpus 109
4.3.2 Choix des sites 110
4.3.2.1 Thématique aide sociale 111
4.3.2.2 Thématique droit au séjour 112
4.3.2.3 Thématique citoyenneté 113
4.3.3 Création d'un corpus exploitable à partir d'un site web 114
4.3.3.1 HTML, un outil d'aide pour le chercheur en linguistique 114
4.3.3.2 Protocole d'extraction 118
4.3.4 Description du corpus spécialisé dit classique 119
4.4 De nouvelles méthodes d'analyse 120
4.4.1 Parcours utilisateur 120
4.4.2 Les *moves* - analyse de l'organisation rhétorique 122
4.4.3 La rédactologie : s'intéresser à la rédaction professionnelle pour comprendre le discours de l'accès aux droits 123
4.4.4 Adaptation à notre corpus 124
4.4.4.1 Les parcours utilisateur pour l'analyse du discours . . 124
4.4.4.2 Les mouvements rhétoriques sur la page web 125
4.5 Un corpus réalisé à partir du parcours utilisateur 126
4.5.1 Mise en place d'une méthodologie complémentaire 127
4.5.2 Création des parcours utilisateur 128
4.5.3 Mise en place d'un schéma d'annotation 131
4.5.3.1 Identification des fonctions des unités discursives . . 131

4.5.4 Intégration des hyperliens . . . 139
4.5.5 Intégration des éléments interactifs . . . 139
4.5.6 Protocole de récolte du corpus parcours . . . 140
4.6 Outils et démarches . . . 141
4.6.1 Présentation des différents outils utilisés . . . 141
4.6.1.1 UAM Corpus Tool . . . 141
4.6.1.2 Sketch Engine . . . 141

III Résultats et analyses 143

5 Caractérisation linguistique 145
5.1 Caractérisation lexicale . . . 145
5.1.1 Groupes nominaux complexes . . . 145
5.1.1.1 Terminologie dans le corpus français sur le thème des aides sociales . . . 146
5.1.1.1.1 **allocation** - analyse . . . 147
5.1.1.1.2 **aide** - analyse . . . 152
5.1.1.1.3 **revenu** - analyse . . . 154
5.1.1.1.4 **complément** - analyse . . . 155
5.1.1.1.5 **prestation** - analyse . . . 157
5.1.1.1.6 **prime** - analyse . . . 159
5.1.1.1.7 **bourse** - analyse . . . 161
5.1.1.1.8 Synthèse - corpus AIDESOC-FR . . . 163
5.1.1.2 Terminologie dans le corpus anglais sur le thème des aides sociales . . . 164
5.1.1.2.1 *credit* - analyse . . . 165
5.1.1.2.2 *allowance* - analyse . . . 166
5.1.1.2.3 *benefit* - analyse . . . 169
5.1.1.2.4 *payment* - analyse . . . 171
5.1.1.2.5 Synthèse - corpus BEN-AN . . . 174
5.1.1.3 Analyse constrastive - corpus AIDESOC-FR et BEN-AN . . . 175
5.1.1.4 Terminologie dans le corpus français sur le thème du droit au séjour . . . 178
5.1.1.5 Terminologie dans le corpus anglais sur le thème du droit au séjour . . . 188
5.1.1.6 Analyse contrastive - ETR-FR et IMM-AN . . . 191

5.1.1.7 Terminologie dans le corpus français sur le thème de la citoyenneté . . . 193
5.1.1.8 Terminologie dans le corpus corpus anglais sur le thème de la citoyenneté . . . 199
5.1.1.9 Analyse constrastive - corpus CIT-FR et CIT-AN . . . 204
5.1.2 Synthèse de l'analyse terminologique anglais-français des trois thématiques . . . 205
5.1.3 Variation terminologique . . . 206
5.1.3.1 Analyse contrastive en corpus - allocation chômage/*Jobseeker's allowance* . . . 207
5.1.3.1.1 Analyse en corpus français . . . 207
5.1.3.1.2 Analyse en corpus britannique . . . 213
5.1.3.2 Analyse contrastive dans l'environnement textuel - allocation chômage/*Jobseeker's allowance* . . . 219
5.1.3.2.1 Analyse en texte français . . . 219
5.1.3.2.2 Analyse en texte britannique . . . 224
5.1.3.3 Variation des noyaux des termes complexes . . . 227
5.1.4 Acronymes et sigles . . . 231
5.1.5 Approche ontologique de la terminologie du discours numérique de l'accès aux droits . . . 235
5.2 Caractérisation syntaxique . . . 242
5.2.1 Pronoms . . . 242
5.2.1.1 Pronoms — corpus français . . . 242
5.2.1.2 Pronoms — corpus anglais . . . 247
5.2.1.3 Analyse contrastive . . . 252
5.2.2 Conditionnalité : si vous êtes [profil administratif]/if you're [profil administratif] . . . 252
5.3 Caractérisation lexico-syntaxique . . . 255
5.3.1 Schémas phraséologiques . . . 255
5.3.1.1 Schéma phraséologique autour des verbes modaux . . . 255
5.3.1.2 Phraséologie autour des organismes d'accès aux droits 258

5.3.1.2.1 Zoom sur la phraséologie autour de Caf . . . 259
5.3.1.2.2 Zoom sur la phraséologie autour de Pôle emploi . . . 260
5.3.1.3 Zoom sur la collocation vous + pouvoir/devoir . . . 262
5.4 Caractérisation stylistique et rhétorique . . . 263
5.4.1 Répétitions . . . 263
5.4.1.1 Citations ou reprises mot à mot . . . 263
5.4.1.2 Éléments définitoires . . . 264
5.4.1.3 Reformulation ou reprise partielle . . . 265
5.4.1.4 Répétitions liées au numérique . . . 266
5.4.1.5 Intertextualité au niveau des pages pour rediriger l'utilisateur . . . 267
5.4.1.6 Synthèse . . . 268

6 Caractérisation discursive **271**
6.1 Les sites web de l'accès aux droits en tant que genre discursif . . . 271
6.1.1 Structure de la page . . . 272
6.1.1.1 Structure de la page française . . . 272
6.1.1.2 Structure de la page britannique . . . 275
6.1.1.3 Analyse comparative . . . 278
6.2 Fonctionnalité des différentes parties du discours . . . 280
6.2.1 Navigation . . . 280
6.2.1.1 Analyse des structures des parties navigation dans le corpus français . . . 280
6.2.1.2 Analyse des structures des parties navigation dans le corpus anglais . . . 282
6.2.1.3 Analyse comparative . . . 284
6.2.2 Liens . . . 287
6.2.2.1 Analyse des liens en corpus français . . . 287
6.2.2.2 Analyse des liens en corpus britannique . . . 293
6.2.2.3 Analyse contrastive des liens . . . 296
6.2.2.4 Spécificités discursives relevées dans le sous-corpus liens français . . . 297
6.2.2.5 Spécificités discursives relevées dans le sous-corpus liens britannique . . . 302
6.2.2.6 Analyse comparative . . . 305

6.2.3 Interactivité 306
6.2.4 Cohérence et cohésion dans les textes numériques 307
6.3 Une arborescence ou un labyrinthe 310
6.4 Recommandations 315
6.4.1 Recommandations méthodologiques pour les recherches en corpus sur les discours administratifs numériques 316
6.4.1.1 Étudier les éléments ne faisant pas partie du contenu . 316
6.4.1.2 S'appuyer sur les éléments HTML pour appréhender le discours 317
6.4.2 Recommandations à destination des experts rédigeant ou produisant le discours étudié 318
6.4.2.1 Remettre la terminologie et la phraséologie au centre des préoccupations 318
6.4.2.2 Créer des outils d'aide à la navigation 319
6.4.2.3 Mettre en place des collaborations avec la recherche en sciences du langage 319
6.4.2.4 Multiplier les genres discursifs pour faciliter la compréhension 320

7 Conclusion **321**

Annexes **353**
.1 Exemples de l'utilisation des termes autour du concept de *unemployment benefits* 353
.2 Exemples de l'utilisation des termes autour du concept d'allocation chômage 359

Liste des éléments sous droits **361**

Table des figures

1.1 Frise chronologique de l'évolution historique de la dématérialisation des démarches d'accès aux droits en France 19
1.2 Frise chronologique de l'évolution historique de la dématérialisation des démarches d'accès aux droits au Royaume-Uni 22
1.3 Schéma récapitulatif des sites officiels d'accès aux droits en France . . . 35
1.4 Schéma récapitulatif des sites officiels d'accès aux droits au Royaume-Uni 38

2.1 Récapitulatif proposé par les modérateurs du groupe r/ukvisa diffusée sur Reddit 58
2.2 Illustration de la spécialisation et de la vulgarisation du discours de l'accès aux droits 76

3.1 Graphique représentant l'évolution des termes **discours numérique** et **discours électronique** dans Ngram Viewer 82

4.1 Tableau montrant les différences entre texte et corpus (Bonelli, 2010) . 103
4.2 Illustration du contenu d'une page du site `service-public.fr` . . 115
4.3 Illustration d'un parcours utilisateur sur le site `service-public.fr`, consulté le 10 septembre 2021 121
4.4 Reconstitution de l'arborescence correspondant à un parcours utilisateur 125
4.5 Résultats pour la collocation de [pouvoir solliciter] sur Sketch Engine . 127
4.6 Différentes parties d'une page du site `gov.uk` 133
4.7 Différentes parties d'une page du site `service-public.fr` 134
4.8 *Contenu complémentaire* sur le site `gov.uk` 136
4.9 *Contenu complémentaire* sur le site `gov.uk` 136
4.10 *Contenu complémentaire* sur le site `service-public.fr` 136
4.11 Rubrique *actualité* sur le site `gov.uk` 137
4.12 Rubrique *actualité* sur le site `service-public.fr` 137
4.13 Pop-up *technique* d'une page sur le site `gov.uk` 138
4.14 Pop-up *technique* d'une page sur le site `service-public.fr` . . . 138
4.15 Schéma d'annotation - partie fonctions 139

4.16 Schéma d'annotation - partie hyperliens 139
4.17 Schéma d'annotation - éléments interactifs 140
4.18 Schéma d'annotation complet 141

5.1 Patrons syntaxiques les plus fréquents dans le corpus CIT-AN 200
5.2 Capture d'écran du chapô d'une page du site service-public.fr concernant les droits au chômage des salariés 219
5.3 Capture d'écran du chapô d'une page du site service-public.fr concernant les droits au chômage des salariés 220
5.4 Capture d'écran du chapô d'une page du site service-public.fr concernant les droits au chômage des agents publics 221
5.5 Capture d'écran du chapô d'une page du site service-public.fr concernant les droits au chômage des agents publics 221
5.6 Capture d'écran du chapô de la page concernant le *New Style Jobseeker's Allowance* 224
5.7 Capture d'écran d'un paragraphe de la page concernant le *New Style Jobseeker's Allowance* 225
5.8 Capture d'écran d'un paragraphe de la page concernant le *Employment and Support Allowance*, consulté le 20 mai 2024 226
5.9 Capture d'écran d'un paragraphe de la page concernant le *Employment and Support Allowance* 226
5.10 Représentation ontologique des termes vedettes autour du concept de document de séjour en France 236
5.11 Terme vedette et variants dans la branche [visa] 238
5.12 Terme vedette et variants dans la branche [VLS-TS] 238
5.13 Terme vedette et variants dans la branche [titre de séjour] 239
5.14 Terme vedette et variants dans la branche [carte de séjour pluriannuelle] 240
5.15 Représentation ontologique des termes vedettes autour du concept de [document de séjour] au Royaume-Uni 241
5.16 Capture d'écran d'une rubrique du site France Travail (anciennement Pôle emploi) 261

6.1 Représentation graphique de la structure d'une page informative du corpus parcours — français 272
6.2 Représentation graphique de la structure d'une page de menu du corpus parcours — français 273
6.3 Rubrique info complémentaire de `service-public.fr` 274
6.4 Rubrique actualité de `service-public.fr` 274

6.5 Représentation graphique de la structure d'une page du corpus parcours — anglais 275
6.6 Rubrique *related content* d'un site web d'accès aux droits britannique . 276
6.7 Rubrique *explore* d'un site web d'accès aux droits britannique 276
6.8 Représentation graphique de la structure d'une page menu du corpus parcours — anglais 277
6.9 Représentation graphique de la structure d'une page menu du corpus parcours — anglais 278
6.10 Représentation de l'en-tête d'une page de site web d'accès aux droits français 280
6.11 Représentation du pied de page d'une page de site web d'accès aux droits français 281
6.12 Représentation de l'en-tête d'une page de site web d'accès aux droits anglais 282
6.13 Représentation de l'en-tête d'une page de site web d'accès aux droits anglais 283
6.14 Capture d'écran de l'en-tête d'une page du site `service-public.fr` 284
6.15 Capture d'écran de l'en-tête d'une page du site `gov.uk` 284
6.16 Capture d'écran du pied de page d'une page du site `service-public.fr` 285
6.17 Capture d'écran du pied de page d'une page du site `gov.uk` 286
6.18 Capture d'écran d'un menu déroulant 289
6.19 Capture d'écran d'un menu statique 290
6.20 Capture d'écran d'un menu statique en bas de page 290
6.21 Capture d'écran d'hyperliens dans le *contenu* 291
6.22 Capture d'écran d'hyperliens dans le *contenu* 291
6.23 Représentation graphique du pourcentage de liens dans le corpus français 292
6.24 Capture d'écran d'un menu 294
6.25 Capture d'écran d'hyperliens dans le contenu 295
6.26 Représentation graphique du pourcentage de liens dans le corpus britannique 295
6.27 Capture d'écran d'hyperliens dans le contenu 299
6.28 Capture d'écran du fil d'Ariane issu de `gov.uk` 308
6.29 Capture d'écran du fil d'Ariane issu de `service-public.fr` 308
6.30 Représentation d'une partie de l'arborescence du site `service-public.fr` 310
6.31 Représentation d'une partie de l'arborescence du site `gov.uk` 311

6.32 Capture d'écran des résultats de recherche de *JSA* 313
6.33 Capture d'écran du menu concernant le droit au séjour des étudiants étrangers 313
6.34 Capture d'écran des résultats de recherche de **carte de séjour étudiant** 314
6.35 Capture d'écran des résultats de recherche de **titre de séjour étudiant** . 315

Liste des tableaux

1.1 Les différents types de charges administratives 30
1.2 Récapitulatif des différents lieux de production du discours de l'accès aux droits en France 37
1.3 Récapitulatif des différents lieux de production du discours de l'accès aux droits au Royaume-Uni 39

4.1 Représentation simplifiée du texte balisé comme contenu par le code HTML 116
4.2 Balises HTML indiquant le contenu sur les pages des sites britanniques 117
4.3 Balises HTML indiquant le contenu sur les pages des sites français . . . 117
4.4 Récapitulatif des deux corpus dits classiques 119
4.5 Parcours utilisateur du site `gov.uk` concernant l'acte de naissance . . . 129
4.6 Parcours utilisateur du site `gov.uk` concernant la naturalisation 129
4.7 Parcours utilisateur du site `service-public.fr` concernant l'acte de naissance 130
4.8 Parcours utilisateur du site `service-public.fr` concernant la naturalisation par mariage 130
4.9 Parcours utilisateur du site *gov.uk* concernant les aides au logement . . . 130
4.10 Parcours utilisateur du site `service-public.fr` concernant les aides au logement 131
4.11 Tableau récapitulatif du corpus parcours dans les deux langues 131

5.1 Dispositifs d'aide financière pour le patron [nom commun][préposition][nom commun] ou [adjectif] 147
5.2 Résultats les plus fréquents pour le patron [**allocation**][préposition][nom commun] 148
5.3 Résultats les plus fréquents pour le patron [**aide**][préposition][nom commun] 152
5.4 Collocats relevés automatiquement du noyau **aide** 152

5.5 Résultats les plus fréquents pour le patron [**complément**][préposition][nom commun] 155
5.6 Collocats relevés automatiquement du nom noyau **complément** 156
5.7 Résultats les plus fréquents pour le patron [**prestation**][préposition][nom commun] 157
5.8 Collocats relevés automatiquement du nom noyau **prestation** 158
5.9 Résultats les plus fréquents pour le patron [**prime**][préposition][nom commun] 159
5.10 Collocats relevés automatiquement du nom noyau **prime** 160
5.11 Résultats les plus fréquents pour le patron [**bourse**][préposition][nom commun] 161
5.12 Collocats relevés automatiquement du nom noyau **bourse** 162
5.13 Résultats les plus fréquents pour le patron ([nom] ou [adjectif]) [*credit*] 165
5.14 Résultats les plus fréquents pour le patron ([nom] ou [adjectif]) [*allowance*] 166
5.15 Résultats les plus fréquents pour le patron ([nom] ou [adjectif]) [*benefit*] 169
5.16 Résultats les plus fréquents pour le patron ([nom] ou [adjectif]) [*payment*] 172
5.17 Résultats les plus fréquents du patron [mot][*Support Payment*] 173
5.18 Résultats les plus fréquents pour le patron [nom] [préposition] [nom] . . 178
5.19 Résultats les plus fréquents pour le patron [nom] [préposition] 179
5.20 Résultats les plus fréquents pour le patron [**titre de séjour**][nom] 180
5.21 Résultats les plus fréquents pour le patron [**titre de séjour**][adjectif] . . 180
5.22 Résultats les plus fréquents pour le patron [**carte de séjour**][nom] . . . 182
5.23 Résultats les plus fréquents pour le patron [**carte de séjour**][adjectif] . . 182
5.24 Résultats les plus fréquents pour le patron [**visa**] ([mot]) ([mot]) [nom] . 187
5.25 Relevés réalisés à partir du patron [nom] [nom] [visa] 189
5.26 Résultats les plus fréquents pour le patron [nom][préposition][nom] . . 193
5.27 Résultats les plus fréquents pour le patron [**acte de**][préposition][mot] . 195
5.28 Résultats les plus fréquents pour le patron [nom][adjectif] 196
5.29 Résultats les plus fréquents pour le patron [**justificatif**][préposition][mot] 198
5.30 Acronymes et sigles relevés en corpus CIT-FR 199
5.31 Résultats les plus fréquents pour le patron [adjectif][nom][nom] 200
5.32 Résultats les plus fréquents pour le patron [nom][préposition][nom] . . 201
5.33 Résultats les plus fréquents pour le patron [nom][nom][nom] 201
5.34 Noms formés à partir de *civil partnership* 202
5.35 Noms formés à partir de *marriage* 202
5.36 Résultats pour l'expression régulière [A-Z]3,4 sur le corpus CIT-AN . . 204
5.37 Récapitulatif des termes relevés autour du concept d'allocation chômage 207

5.38 Récapitulatif des modes d'utilisation des variants d'allocation chômage 213
5.39 Récapitulatif des termes et de leur utilisation en corpus autour du concept de l'allocation chômage 213
5.40 Récapitulatif des termes et de leur utilisation autour du concept de *unemployment benefits* 218
5.41 Textes et titres des pages étudiées 222
5.42 Résultats de l'expression régulière [A-Z]3,4 dans le corpus français . . 231
5.43 Résultats pour l'expression régulière [A-Z]3,4 dans le corpus anglais . . 233
5.44 Résultats pour l'expression régulière [A-Z]2 dans le corpus anglais . . . 234
5.45 Pronoms personnels les plus fréquents dans le corpus FrTenTen et le corpus d'étude français 242
5.46 Collocats les plus fréquents de **vous** détectés automatiquement par Sketch Engine 243
5.47 Verbes les plus fréquemment utilisés avec le pronom **vous** 244
5.48 Pronoms personnels les plus fréquents dans le corpus EnTenTen et le corpus d'étude anglais 248
5.49 Collocats les plus fréquents de *you* détectés automatiquement par Sketch Engine 248
5.50 Verbes les plus fréquemment utilisés avec le pronom *you* 249
5.51 Occurrences les plus fréquentes pour le schéma : if you're [profil] . . . 253
5.52 Occurrences les plus fréquentes pour le schéma : if you're applying for [dispositif] 253
5.53 Occurrences les plus fréquentes pour le schéma : si vous êtes [profil] . . 254
5.54 Auxiliaires modaux les plus fréquents dans le corpus d'étude 255
5.55 Auxiliaires modaux les plus fréquents dans le corpus web général ENTenTen 256
5.56 Collocats les plus fréquents avec les auxiliaires modaux - gauche détectés automatiquement par Sketch Engine 256
5.57 Collocats les plus fréquents de *will* détectés automatiquement par Sketch Engine 257
5.58 Collocats les plus fréquents de *You will need to* détectés automatiquement par Sketch Engine 257
5.59 Collocats les plus fréquents de *you will get* détectés automatiquement par Sketch Engine 257
5.60 Collocats les plus fréquents de *can* détectés automatiquement par Sketch Engine 258
5.61 Collocations automatiquement détéctées du schéma [caf] [verbe] 259

5.62 Collocats les plus fréquents de Pôle emploi détectés automatiquement par Sketch Engine 260
5.63 Collocats les plus fréquent de **vous pouvez** automatiquement détectés par Sketch Engine 262
5.64 Collocats les plus fréquent de **vous devez** automatiquement détectés par Sketch Engine 263

6.1 Statistiques des hyperliens présents dans le corpus parcours en français 288
6.2 Statistiques des sections *contenu* et *navigation* dans le corpus parcours en français 288
6.3 Statistiques des hyperliens présents dans le corpus parcours anglais . . . 293
6.4 Statistiques des sections *contenu* et *navigation* en anglais britannique . . 294
6.5 Noms les plus fréquemment utilisés dans le sous-corpus liens français . 297
6.6 N-grammes les plus fréquemment utilisés dans le sous-corpus de liens français 298
6.7 Verbes les plus fréquemment utilisés dans le sous-corpus liens français . 300
6.8 Acronymes et sigles présents dans le corpus parcours français 302
6.9 Noms les plus fréquemment utilisés dans le sous-corpus liens britannique 303
6.10 Verbes les plus fréquemment utilisés dans le sous-corpus lien britannique 304
6.11 Acronymes et sigles présents dans le corpus parcours britannique 305
6.12 Statistiques des éléments interactifs présents dans le corpus parcours français 306
6.13 Statistiques des éléments interactifs présents dans le corpus parcours britannique 306

1 Exemples de l'utilisation des termes autour du concept de *unemployment benefits* 358
2 Exemples de l'utilisation des termes autour du concept d'allocation chômage 360

Abbreviations

ANEF administration numérique des étrangers en France

CORPUS AIDESOC-FR corpus français sur le thème des aides sociales

CORPUS BEN-AN corpus anglais sur le thème des aides sociales

CORPUS CIT-AN corpus anglais sur le thème de la citoyenneté

CORPUS CIT-FR corpus français sur le thème de la citoyenneté

CORPUS ETR-FR corpus français sur le thème du droit au séjour

CORPUS IMM-AN corpus anglais sur le thème du droit au séjour

COSLA comité d'orientation pour la simplification du langage administratif

DILA direction de l'information légale et administrative

DITP direction interministrielle de la transformation publique

GDS Government Digital Service

SIG service d'information du gouvernement

Introduction

0.1 Présentation du sujet

En 2002, dans l'Actualité Terminologique, Langlois (2002) fait l'annonce suivante :

> Cette année, le gouvernement français a pris le taureau par les cornes en lançant une initiative audacieuse. Quand elle aura porté ses fruits, elle transformera de façon radicale le langage administratif, ou, si vous voulez, le français qu'on utilise dans la correspondance officielle. (Langlois, 2002, p. 5)

À la suite de la création du comité d'orientation pour la simplification du langage administratif (COSLA) en 2000, les linguistes et autres personnalités s'intéressant aux discours administratifs s'enthousiasment à la perspective d'un discours administratif simple et compréhensible par tous. Cet enthousiasme pour la simplification du langage administratif est à la hauteur du niveau de difficulté perçu par les usagers. En mai 2024, le Premier ministre français Gabriel Attal a déclaré vouloir « lancer un "travail colossal", pour faire de "la débureaucratisation, un objectif", et "la simplification, un cap" ». Il annonce alors le lancement du plan de simplification « Parlez-nous français », qui vise notamment à « lutter contre le jargon » (Graillot, 2024).

Vingt ans plus tard, il semble donc que la transformation radicale du langage administratif n'ait pas eu lieu. Malgré les nombreux plans de simplification qui se sont succédés, les mêmes préoccupations demeurent en 2024 comme en 2002. La simplification des formulaires, un des fers de lance du COSLA qui œuvre entre 2002 et 2006, fait l'objet de travaux similaires en 2024, avec le lancement du *Kit Formulaires*, dont le but est de créer « des formulaires accessibles, simples à remplir et compréhensibles pour les usagers » (Direction interministerielle de la transformation publique et Direction de l'information légale et administrative, 2024, p. 3). Le lexique, préoccupation phare du COSLA qui avait travaillé à la mise en place d'un lexique de termes administratifs, semble faire l'objet de nouvelles dynamiques de changement notamment dans une émission diffusée le 31 mai 2024 sur Action Publique TV : « Pour une relation entre citoyens et services publics sans jargon ». Cette formulation politique indique clairement un regain d'intérêt pour la terminologie administrative.

De l'autre côté de la manche, au Royaume-Uni, après la mise en place d'une centralisation des informations sur le site `gov.uk` au début des années 2010, le *Plain English* devient la star de la simplification du discours administratif et juridique, avec en 2014 l'introduction d'un guide de style intitulé *Content design : planning, writing and managing content* pour l'ensemble des contenus publiés pour les usagers des services publics. Dix ans plus tard, en janvier 2024, sur le blog du Government Digital Service (GDS), la simplification des formulaires est au cœur des préoccupations (Quilan, Quilan).

Mais que s'est-il passé en vingt ans pour que l'ensemble de ces préoccupations reviennent sur le devant de la scène ? C'est d'abord qu'il n'est pas aisé de simplifier un discours spécialisé, notamment quand cette spécialisation est héritée du discours administratif et juridique. C'est aussi que la transformation numérique est passée par là et a porté l'ensemble des communications administratives sur des supports numériques. Cette transformation a mené à une modification radicale des rituels d'interaction entre les administrations et leurs usagers.

À partir des années 2000, conscient des capacités transformatrices des nouvelles technologies et de leur inévitable intégration à nos sociétés, un grand nombre de pays amorce une transition numérique de leur administration et parmi eux, la France et le Royaume-Uni. Les transformations engagées ne sont pas linéaires et les évolutions des technologies poussent les administrations à se remettre constamment en question. Le Royaume-Uni, après avoir amorcé sa transition dans les années 1990, est contraint de revoir sa stratégie dans les années 2010 en bouleversant les structures mises en place pour créer le GDS. La France, qui entreprend également ce changement dans la fin des années 1990, réalise une transition incrémentale et numérise petit à petit ses services et ses démarches. Ces transitions ne font au départ pas l'objet de politiques linguistiques claires et les discours sont portés sur le choix du média qui semble le plus adapté aux besoins des administrations et des usagers. Le média choisi est le site web qui constitue désormais des pôles d'information importants où se diffuse le discours administratif vers les usagers.

Lors de permanences proposées aux usagers d'un centre social dans le 13e arrondissement de Paris, nous nous sommes rendu compte du poids écrasant de ces sites dans l'accès aux droits. Les sites web constituent le point d'entrée dans le discours, le point de contact avec les administrations et le lieu de réalisation des démarches dans une grande majorité des cas. Ils sont également un support de communication aux caractéristiques uniques qu'il convient de maîtriser et de comprendre. Le site web est installé dans nos pratiques discursives et informatives et, contrairement aux réseaux sociaux, il n'est pas vu comme un élément particulièrement disruptif alors qu'il peut constituer un obstacle à la réalisation d'actions citoyennes essentielles. Dans notre thèse, le site web et ses dif-

férentes caractéristiques occupent une place centrale dans notre conception du discours, car il s'agit d'un genre discursif à part entière.

Notre première hypothèse pour l'étude de ce discours est que les outils théoriques et méthodologiques apportés par l'étude des langues de spécialité et notamment la terminologie et la phraséologie peuvent permettre de caractériser le discours de l'accès aux droits. Cette caractérisation pourra ensuite servir de point de départ pour identifier les points de blocage dans la compréhension du discours. Les gouvernements étant intéressés par des questions de lexique et de jargons, la terminologie semble être la discipline idéale pour aborder le discours administratif. Les analyses réalisées pourront servir de base à la constitution d'outils d'aide à la compréhension et à la rédaction.

Notre seconde hypothèse est qu'il est indispensable d'intégrer la dimension numérique de ce discours à nos recherches. La transition étant achevée, il est évident que le discours est désormais numérique et le restera. Nous avons donc étudié le discours administratif de l'accès aux droits en tant que discours numérique natif, élaboré directement pour les plateformes en ligne et non porté sur support numérique après rédaction (Paveau, 2015). En effet, nos recherches ont montré que la composante numérique du discours modifiait le fil discursif.

> *Digital discourse is an interactive, dynamic, ever evolving multimodal discourse. At the same time a mosaic and a maze, with its breadcrumb trails of navigation histories in an attempt to prevent us from getting lost, it is nowadays the major medium of information.* [1] (Bouchet et Pecman, à paraître)

Les éléments interactifs et technosignes créent un labyrinthe discursif où l'utilisateur doit se repérer et recomposer son texte. La nature même du site web, en particulier sa délinéarisation, combinée à la complexité terminologique, forme un discours particulièrement difficile à comprendre et envisager dans son ensemble.

0.2 Présentation du titre

Pour présenter la recherche réalisée dans le cadre de cette thèse, nous avons opté pour le titre suivant :

> L'impact des discours numériques sur l'accès aux droits pour les citoyens : analyse outillée et contribution à la caractérisation linguistique des discours administratifs en ligne en France et au Royaume-Uni.

1. « Le discours numérique est un discours multimodal, interactif, dynamique et en constante évolution. A la fois mosaïque et labyrinthe avec ses historiques de navigation, tel le fil d'Ariane, tentant si bien que mal de nous aider à ne pas nous égarer, il est le principal moyen d'accès à l'information aujourd'hui. (Bouchet et Pecman, à paraître). »

Ce titre reflète l'ensemble des objectifs de notre recherche. Dans un premier temps, nous sommes partie de nos observations réalisées dans le cadre des permanences d'écrivain public. C'est pour cela qu'il nous a paru important de centrer le titre sur les usagers en faisant référence à l'impact que peut avoir ce discours sur l'accès aux droits. Identifier cet impact et les points d'amélioration possibles fait partie de nos objectifs principaux.

Nous avons choisi de réaliser une caractérisation linguistique de ce discours et nos aborderons plusieurs niveaux de l'analyse linguistique : terminologique, phraséologique, syntaxique et rhétorique. Nous nous sommes intéressée dans un premier temps à ce qui est accessible via le discours, par la terminologie et la phraséologie notamment, en mettant en avant les termes les plus saillants et les plus complexes du domaine et la phraséologie liée aux organismes d'accès aux droits. Dans un deuxième temps, nous avons distingué la caractérisation linguistique, propre au système de la langue, de la caractérisation discursive, propre aux réalisations de la langue dans un type de discours qui est dans notre cas l'accès aux droits pour les citoyens. Notre thèse vise donc cette double caractérisation qui a pour but d'apporter une contribution à l'évaluation de l'impact des discours et des modalités sur l'accès aux droits des citoyens.

Dans cette thèse, nous utilisons les termes discours de l'accès aux droits et discours administratifs numériques de façon interchangeable. Nous souhaitons préciser ici la nature exacte de ces deux termes et la raison de leur utilisation dans nos travaux. Nous définissons le discours de l'accès aux droits comme un type de discours administratif numérique. Il existe donc d'autres types de discours administratifs numériques qui ne sont par exemple pas consacrés aux usagers ou qui ne sont pas diffusés sur un site web. On peut penser par exemple aux communications et documents internes aux administrations qui se trouvent sur support numérique, dans des logiciels par exemple. Nous faisons donc référence au discours numérique de l'accès aux droits comme étant un discours administratif numérique.

0.3 Étudier un discours numérique interactif

Nous avons adopté une approche pluridisciplinaire qui a pour but de trouver des régularités dans la langue. L'approche de l'analyse du discours permet d'étudier le texte au-delà de la phrase (Schiffrin et al., 2015), en prenant en compte son contexte et son environnement de production. Pour cela, la notion de genre de discours a une place importante dans nos travaux puisque les genres permettent d'observer les « régularités sociohistoriques » du discours (Adam, 1997, p. 670) ou, comme le définit Bakhtine (1984), la stabilité des énoncés. Cette notion permet également d'identifier la place du discours dans le domaine de l'accès aux droits en général. Les textes étudiés se rapprochent éga-

lement des textes procéduraux puisqu'ils demandent à l'usager de réaliser des actions en produisant des explications pour les réaliser. Comme l'indique Adam (2001), ils disent « de et comment faire ».

Le discours de l'accès aux droits est également un discours numérique natif, créé directement sur la plateforme en ligne avec les outils d'internet (Paveau, 2017). Ces discours sont caractérisés par des technosignes et des technomots, dont font partie les hyperliens qui permettent de lier un discours source à un discours cible (Paveau, 2017). Ces hyperliens qui caractérisent le discours numérique et particulièrement le discours issu de sites web permettent aux utilisateurs de se déplacer dans l'arborescence du site. Ces éléments sont interactifs et rendent le fil discursif unique et personnalisable mais aussi délinéarisé.

0.4 Le discours administratif en tant que langue de spécialité vulgarisée

Le discours étudié s'apparente à une langue de spécialité car il est « vecteur de connaissances spécialisées » (Lerat, 1995, p. 31). Il s'agit d'une utilisation spécifique de la langue employée par un groupe de spécialistes possédant une expertise variée (Gledhill et Kübler, 2016). Le discours étudié est toutefois particulier puisqu'il rassemble une communauté de discours singulière composée d'experts et d'usagers. Les locuteurs du discours de l'accès aux droits possèdent une expertise issue de différentes expériences et connaissances. Cette expertise peut être professionnelle chez les fonctionnaires ou les rédacteurs par exemple, mais elle peut aussi être informelle, chez les militants ou les usagers. Dans tous les cas, elle répond à une demande ou à un besoin qui pousse un individu à endosser un rôle d'expert (Lelebina et Sardas, 2011). Dans ce domaine, les usagers ont diverses compétences, qu'ils utilisent dans les différents espaces en ligne afin de créer des communautés d'entraide et d'apprentissage du discours de l'accès aux droits. La compréhension du discours de l'accès aux droits est aussi liée à un concept représenté par l'*administrative literacy* qui comprend plusieurs compétences, notamment la connaissance du vocabulaire, la familiarité avec les formulaires et les documents, mais aussi la ténacité ou la répétition (Gordon, 1975).

Le discours administratif numérique est aussi un discours vulgarisé, puisqu'il est issu des langues de spécialité administrative, juridique, politique et institutionnelle, mais aussi car il est destiné aux usagers. La vulgarisation d'un discours consiste à diffuser vers l'extérieur, à des non-experts, des connaissances élaborées à l'intérieur d'une communauté (Authier, 1982, p. 34), ce qui est le cas du discours de l'accès aux droits. Il a tout de même un statut spécifique car la communauté des usagers possède un savoir

et une expertise hétérogènes. Pour simplifier ce discours spécialisé, les administrations peuvent utiliser des outils de simplification qui permettent d'adapter le discours à un certain public. Parmi ses outils, le langage clair représente une stratégie essentielle pour le gouvernement britannique, dont l'objectif est de permettre aux lecteurs de comprendre un texte dès la première lecture (Rossetti et al., 2020).

Le jargon administratif est également au cœur des préoccupations des usagers et des législateurs. Pour aborder la problématique du jargon qui semble essentiel dans ce domaine, nous avons choisi de passer par la terminologie et l'étude des termes.

> *Terminology is the study and the field of activity concerned with the collection, description, processing and presentation of terms.* (Sager, 2002, p. 2).

La langue de spécialité est utilisée pour communiquer de la façon « la moins ambiguë possible sur le plan lexical » (Dury et Lervad, 2008, p. 67) et nous avons donc étudié les phénomènes de variation et de synonymie qui sont présents dans ce discours et qui pourraient causer des problèmes de compréhension. La variation est un concept particulièrement intéressant dans cette étude, car elle peut témoigner d'une vulgarisation du discours avec la variation verticale, c'est-à-dire la variation dans le degré de spécialité des termes (Freixa, 2006). Les phénomènes de variation sont liés à la compréhension du domaine spécialisé et peuvent tout autant être des obstacles à la communication que des aides (Pecman, 2014).

Les difficultés de compréhension d'un discours spécialisé ne se limitent pas à la terminologie, mais peuvent s'étendre à la phraséologie qui comprend l'étude des structures, du sens et de la restriction combinatoire (Granger et Paquot, 2008). L'étude de la phraséologie est d'autant plus adaptée au discours administratif car il s'appuie fortement sur des préconstruits (Anderson, 2003). Le discours administratif s'appuie largement sur le discours juridique qui est caractérisé par l'utilisation de nombreux préfabriqués et une tendance à la synonymie (Rey, 2008).

L'étude de la terminologie et des unités discursives permet de faire un état des lieux du discours de l'accès aux droits et de mettre en avant les points de difficulté.

0.5 Problématique

Le discours de l'accès aux droits a été drastiquement modifié ces dernières années avec la transformation numérique. De plus, il n'a, à notre connaissance, jamais fait l'objet d'une caractérisation linguistique à part entière en tant que discours spécialisé. Les administrations communiquent depuis de nombreuses années avec leurs usagers. Ces communications sont unilatérales, c'est-à-dire que ces discours ne peuvent pas faire l'objet de conversation avec les administrations et les conventions de communication

sont fixées par ces mêmes administrations. La transformation numérique a accéléré cette dynamique de communication, en mettant le site web, un objet discursif dynamique et unilatéral, au centre des communications entre usagers et administration. Cette intensification des communications numériques ainsi que la diminution des communications bilatérales au guichet et par téléphone ont transformé les rituels d'interaction entre l'administration et ces usagers. Les discours numériques étant fragmentés et unilatéraux, les relations peuvent être inégales, la communication entre les deux peut facilement devenir inefficace. Cela est d'autant plus pertinent lorsque l'on considère que tous les usagers ne possèdent pas les mêmes compétences pour accéder à ce type de discours.

Pour limiter les inégalités causées par un discours numérique fragmenté, des initiatives ont vu le jour au sein de plusieurs domaines. On observe une volonté d'institutionnaliser les traducteurs-interprètes des services publics (Pointurier, 2016) qui pourraient aider à garantir l'accès aux droits des personnes ayant des difficultés en français, en particulier les réfugiés. On remarque également une volonté de simplifier les démarches administratives, par exemple avec l'instauration de l'aide *Universal Credit* au Royaume-Uni, un dispositif d'aide financière ayant pour objectif de diminuer le nombre de dispositifs d'aides financières en les centralisant dans une aide unique. Les institutions cherchent à rendre les informations plus claires grâce à diverses initiatives telles que la création d'un outil d'aide à la rédaction en style clair par les services administratifs britanniques ou d'une plateforme permettant de signaler les documents trop compliqués à l'administration française. Ces actions et ces recherches apportent des réponses et aident à limiter ces inégalités. Toutefois, la charge administrative qui pèse sur les usagers (Döring, 2018) reste considérable. Cette charge est d'autant plus lourde pour les personnes qui n'ont pas les compétences numériques requises. Il semble donc important d'aborder les divers aspects de cette problématique, afin de mieux comprendre ces transformations et leurs conséquences. Dans cette thèse, nous proposons d'aborder les questions suivantes :

— Comment élaborer une méthodologie d'analyse pour dégager les caractéristiques, à la fois terminologiques et phraséologiques, d'un discours nativement numérique et au cœur de la vie de tous locuteurs ? En combinant une approche outillée et descriptive, quelles problématiques terminologiques et phraséologiques peuvent être mises en lumière à travers l'analyse de ce discours ?
— Comment une approche discursive, terminologique et phraséologique peut-elle contribuer à caractériser efficacement le discours administratif numérique ? Comment peut-elle apporter des éléments concrets pour aider à la création de ressources à destination des différents acteurs et usagers ? Quels sont les types de locuteurs concernés par ce discours ?

Les objectifs scientifiques visés dans cette thèse sont donc doubles, ils se concentrent dans un premier temps à la définition du genre de discours étudié : un discours numérique natif dont l'objectif est de transmettre des informations aux usagers. La définition de ce genre de discours permet de définir une méthodologie adaptée à l'étude de ce discours. Le second objectif est de réaliser une caractérisation discursive et linguistique du corpus comparable, anglais-français, créé spécifiquement pour les besoins de cette recherche.

0.6 Annonce du plan

Afin d'aborder ces problématiques et de contribuer à la caractérisation du discours administratif numérique étudié, nous avons organisé cette thèse en trois parties. La première partie est consacrée à l'État de l'art afin de présenter le contexte de l'étude, d'identifier les éléments théoriques et de définir les concepts abordés. Dans un premier temps, nous retraçons l'histoire du discours de l'accès aux droits en France et au Royaume-Uni et nous définissons les notions principales liées à ce domaine. La double transformation numérique et idéologique est expliquée et nous dressons le paysage documentaire actuelle du discours numérique de l'accès aux droits. Dans un second temps, nous faisons un point sur les recherches en analyse de discours, mais aussi en langue de spécialité et en vulgarisation. Nous examinerons la communauté de discours de l'accès aux droits et des différents types d'expertise des locuteurs. La définition des notions de spécialisation et de vulgarisation nous permet de positionner le discours de l'accès aux droits, qui est un discours hybride.

Dans la deuxième partie de la thèse, nous présentons les outils méthodologiques et théoriques utilisés dans cette recherche. La première partie est consacrée aux concepts théoriques utilisés, notamment en définissant les différents types et les caractéristiques du discours numérique, mais aussi en définissant les spécificités linguistiques du corpus par rapport à la terminologie et la phraséologie. Nous présentons ensuite la méthodologie utilisée pour réaliser nos analyses, en présentant la linguistique de corpus, ainsi que les corpus créés pour notre recherche. Nous avons ainsi élaboré deux méthodologies différentes afin de créer deux corpus différents pour nos recherches.

Dans une troisième partie, nous présentons les résultats et analyses réalisées. Nous commençons par une caractérisation linguistique. Cette caractérisation est d'abord lexicale, en présentant la terminologie du domaine. Elle est ensuite syntaxique, avec l'étude des pronoms et de la conditionnalité dans le discours. Elle est aussi lexico-syntaxique, stylistique et rhétorique. Nous réalisons une étude de la phraséologie autour des organismes d'accès aux droits et des phrases répétées. Nous réalisons ensuite une caractérisation discursive. Nous mettons tout d'abord en avant la structure du genre et de ses

différentes unités discursives. Nous analysons ensuite les fonctionnalités des différentes parties du discours, en prêtant une attention particulière aux éléments interactifs (liens, menus déroulants, navigation, etc.), ce qui nous permet d'étudier la cohérence et la cohésion du discours. Avant de conclure, nous proposons des recommandations avec deux objectifs. La première partie de ces recommandations s'applique à la méthodologie des recherches pour les discours numériques. La deuxième partie est destinée aux experts rédigeant ou produisant le discours étudié.

Première partie

État de l’art

Chapitre 1

Le discours de l'accès aux droits : contexte historique et situation actuelle

"Ask ordinary citizens for an example of unreadable prose, and half of them will show you a government document"
– Don Byrne

« On n'en finit jamais avec l'administration, et même quand on y met le meilleur de soi, ou quand a minima on suit les règles si tant est qu'on les comprenne toutes, il y a quelque chose qui nous rattrapera toujours. »
– Géraldine Mosna-Savoye

1.1 Accès aux droits en France et au Royaume-Uni

Afin de proposer une caractérisation d'un discours, il convient de dresser un portrait du contexte dans lequel ce discours se manifeste et évolue. Dans les études en langue de spécialité, et en linguistique plus généralement, ce contexte est appelé domaine. Pour l'étude du discours de l'accès aux droits, il est ainsi particulièrement important de s'intéresser au domaine administratif des droits sociaux, à sa trajectoire historique et à sa situation actuelle, car il s'agit d'un champ moderne empreint d'enjeux sociopolitiques importants. Dans la vie quotidienne des citoyens, ce domaine renvoie à un ensemble d'actions et pratiques qui sont désignées par le terme « démarches administratives ». L'étude de son évolution et de l'ensemble de ses caractéristiques est une étape essentielle dans la constitution d'une caractérisation complète du discours. Nous nous intéressons spécifiquement au domaine en France et au Royaume-Uni. Dans ces pays, l'accès

aux droits est étroitement lié aux domaines politiques et juridiques et font partie intégrante de son évolution. Ces deux domaines influencent de façon marquée le discours de l'accès aux droits. De nouveaux termes ou éléments de discours sont introduits pour correspondre à des campagnes politiques, tandis que d'autres sont introduits par un décret ou une loi. L'accès aux droits est également un sujet controversé, qui est parfois au premier plan dans les médias, car il concerne un grand nombre d'acteurs : privés, publics, associatifs, particuliers, etc.

Dans ce premier chapitre, nous présentons des éléments afin de mieux comprendre l'accès aux droits, désormais étroitement lié aux outils numériques, ainsi que les évolutions récentes dans ce domaine. Nous commencerons par définir la notion d'accès aux droits, en l'accompagnant d'un bref historique de son évolution en France et au Royaume-Uni. Nous nous intéresserons ensuite à la position de l'État lors de cette transformation numérique des administrations, en montrant que ces changements ne sont pas seulement techniques, mais qu'ils s'accompagnent d'une transformation idéologique. Enfin, nous présenterons le paysage documentaire actuel au Royaume-Uni et en France.

L'accès aux droits a longtemps été réalisé sur des médias non numériques : flyers, conseillers administratifs, brochures, renseignements téléphoniques. Ces moyens ont été utilisés pour informer les usagers et les aider à réaliser leurs démarches administratives. Ces dernières années, ces démarches ont été profondément transformées par des programmes de dématérialisation des démarches administratives. Dans notre étude, nous nous intéressons à la dématérialisation grâce à une approche contrastive et outillée, en comparant ces évolutions au Royaume-Uni et en France. Dans ces deux pays, cette transition a bouleversé les rituels d'interaction entre administrations et usagers, en créant de nouvelles dynamiques discursives. Nous retraçons tout d'abord l'historique de l'accès aux droits et de sa transition numérique avant de mettre en lumière la transformation idéologique qui s'est opérée ces dernières années. Ces éclairages historiques et sociaux nous ont permis de positionner notre étude linguistique dans un contexte sociétal et mondial.

1.1.1 Définition de la notion d'*accès aux droits*

Nous allons tout d'abord définir la notion principale de notre étude : l'accès aux droits. D'après le Conseil Départemental de l'Accès au Droit (CDAD) :

> « Accès au droit » signifie que chaque personne peut bénéficier gratuitement d'une information sur ses droits et devoirs et être en mesure de les faire valoir. (Conseil Départemental de l'Accès au Droit, 2018)

Cette notion témoigne d'une volonté de mettre en place des ressources afin d'orien-

ter les citoyens et ressortissants vers les services publics adaptés à leurs besoins. L'accès aux droits consiste donc à réaménager les politiques sociales afin de toucher le plus de citoyens possible (Lafore, 2014). Des aménagements sont effectués pour permettre aux différentes personnes de bénéficier des biens, services et prestations publics disponibles pour eux. Une grande partie de l'accès aux droits repose naturellement sur la transmission d'informations adaptées aux publics concernés. Cette transmission d'informations se fait désormais en grande partie par écrits numériques, notamment sur les sites web (Lochak, 2022). Il s'agit de rendre accessibles les informations et les démarches. Il est donc essentiel de s'intéresser au discours utilisé par l'administration pour transmettre ces informations. L'aide à l'accès aux droits est définie en France dans la loi du 18 décembre 1998. D'après le service d'information du gouvernement (SIG) et la direction de l'information légale et administrative (2022), elle consiste d'abord à mettre à disposition, dans des lieux accessibles à tous, les services suivants :

— information sur les droits et devoirs des personnes ;
— orientation vers les organismes, services ou professionnels chargés d'assurer ou de faciliter l'exercice des droits ;
— aide pour accomplir les démarches nécessaires à l'exercice d'un droit ou l'exécution d'une obligation ;
— assistance au cours de procédures non juridictionnelles. (Service d'information du Gouvernement (SIG) et Direction de l'information légale et administrative (DILA), 2022)

Au Royaume-Uni, le concept d'accès aux droits ne fait pas l'objet d'une loi et n'existe pas sous la même forme. Il existe bien évidemment des services publics et des politiques d'accès qui y sont liées, mais ces programmes d'aide à l'inclusion ne font pas partie d'une politique globale et gouvernementale. Il existe cependant un concept similaire à l'accès aux droits, les *Welfare rights* (ou « droits sociaux »), qui sont décrits dans la Wiley Encyclopedia of Management Online comme suit :

> *Welfare rights are rights to or concerning well-being. Thus, "welfare" refers to [...] the organized provison for the basic well-being of the needy members of the community.* (Wellman, 2015)

Il est cependant important de noter que *welfare* fait référence à des droits sociaux, notamment à des aides financières : aide au logement, allocation chômage, familiale, etc. Là où la France rassemble sous un même concept les démarches pour obtenir des allocations, faire appel à une aide juridique, ou obtenir un titre de séjour, le Royaume-Uni utilise des termes bien distincts. *Rules for immigration* est un concept qui fait référence au droit au séjour et qui concerne donc les étrangers, tandis que le concept de *Welfare*

rights concerne les droits sociaux. Il existe donc deux concepts différents pour parler de deux domaines différents. En France, le droit au séjour et les droits sociaux font tous partie de l'accès aux droits. De plus, nous observons une différence fondamentale dans les définitions proposées. La France mentionne le droit de « chaque personne », tandis que la définition britannique indique que ces droits ne concernent que les « membres précaires de la communauté »[1] (Wellman, 2015). La définition britannique fait donc déjà état d'une conditionnalité des droits, que nous mettrons en lumière dans la section 1.2.2, dans la mesure où ces différences se reflètent dans le discours et la terminologie et peuvent en outre poser des problèmes de traduction.

Par ailleurs, dans la sphère francophone, la notion d'accès aux droits fait l'objet d'une variation orthographique suivant le domaine dans lequel elle apparaît. Dans le domaine juridique, et notamment dans les lois ou dans les communications du gouvernement, elle apparaît au singulier « au droit », comme le montre la définition présentée au début de cette section (Conseil Départemental de l'Accès au Droit, 2018). Toutefois, dans la recherche, notamment en sciences sociales en général, on retrouve le pluriel « aux droits », comme on peut l'observer dans les divers articles qui sont abordés dans les sections suivantes (Lafore, 2014; Lochak, 2022), qui fait référence à la pluralité des droits. Nos travaux de recherche se placent dans cette perspective et nous utiliserons donc le second variant : « accès aux droits ». Nous allons maintenant présenter les trajectoires historiques de la dématérialisation des démarches de ces deux pays.

1.1.2 L'accès aux droits en France

L'histoire de la transformation numérique des administrations françaises commence dans les années 1990 avec l'arrivée d'internet.

Le premier portail internet d'information grand public, AdmiFrance, est ouvert en 1998 (Figure 1.1). Il est ensuite remplacé en 2000 par le site `service-public.fr` (Figure 1.1). Ce site est toujours actif aujourd'hui[2], et il est désormais le site principal d'information administrative. En 1999, le site Légifrance est mis en ligne afin de diffuser les textes législatifs et réglementaires (Figure 1.1). En 2001, un portail de téléservice est ouvert afin de réaliser certaines démarches en ligne. Le premier plan de dématérialisation à grande échelle est mis en place dès 2004, il s'agit du Plan Administration Électronique (ADELE) (Figure 1.1). En parallèle, les collectivités territoriales se dotent d'Espaces Publics Numériques (EPN). Les EPN sont des lieux gérés par les collectivités territoriales, permettant d'assurer l'accès au numérique pour tous, en proposant notamment des ordinateurs connectés à internet, mais aussi des initiations et formations

1. "the needy members of the community"
2. consulté le 23/10/2023

aux outils numériques[3]. La mise en place de ces plateformes et de ces dispositifs représente le balbutiement de la dématérialisation. Lors de leurs créations, le numérique a pour seule fonction de compléter les services déjà présents. En 2008, le gouvernement met en place un plan de développement de l'économie numérique, intitulé France numérique 2012. Ce plan a pour but de permettre à tous les Français d'accéder aux services numériques, en modernisant les plateformes, les contenus et en rendant les sites internet publics plus accessibles (Figure 1.1). En 2010, un changement s'opère dans la relation usager-administration et la dématérialisation des démarches connaît une nette accélération. L'État engage une stratégie de transformation numérique des services publics. Cette transformation est tout d'abord technique, elle implique un changement de support et de procédure administrative. Elle entraîne également un changement idéologique que nous présenterons dans la partie suivante.

La volonté de dématérialiser s'accompagne d'une logique de simplification, notamment dans les mesures inscrites au « Choc des simplifications », qui visent la création de nouveaux services pour les usagers (Figure 1.1). Pour les étrangers, le gouvernement crée l'Administration Numérique pour les Étrangers en France (Anef) en 2014 (Figure 1.1). Cet organisme a pour but de dématérialiser le droit au séjour de « bout en bout », pour la préfecture et pour les usagers. Il vient s'ajouter à l'ensemble des démarches qui ont déjà été numérisées : allocations familiales, impôts, sécurité sociale, etc. En 2015, le système France Connect est déployé : il s'agit d'un agrégateur d'identité sécurisé pour l'accès en ligne des services publics. Ce service permet de centraliser l'accès à un grand nombre de services publics.

À partir de 2015, il y a une augmentation des initiatives d'entrepreneuriat numérique au sein de l'État : hackathons, start-up d'État, entrepreneurs, etc. En 2017, le programme Action Publique 2022 prévoit de dématérialiser l'ensemble des démarches les plus utilisées par les Français d'ici 2022 (Figure 1.1).

En 2020, avec la pandémie de COVID-19, la transformation des relations entre administration et usager prend un tournant décisif. La plupart des guichets administratifs ferment pour des raisons sanitaires, et le numérique joue désormais un rôle essentiel dans la communication avec les administrations. Cet événement a notamment eu un impact significatif sur l'accès aux droits au séjour. Les services numériques nécessaires pour la demande d'un visa ou titre de séjour reposent principalement sur la demande d'un rendez-vous en ligne. En 2020, la démarche était encore réalisée principalement en personne, à la préfecture, avec un agent. Progressivement et avec une accélération significative depuis la pandémie, la demande de ces titres se fait uniquement par des formulaires en ligne. C'est le cas des titres de séjour mention étudiant depuis octobre 2020.

3. Voici à titre d'exemple le site des EPN de Paris : http ://www.epn-paris.org, consulté le 23/10/2023

En 2021, cette accessibilité au téléservice s'adresse à la plupart des demandeurs de titre de séjour, qui peuvent être obligés de passer par la plateforme Anef (Figure 1.1). Cette dématérialisation s'accompagne de la fermeture de guichets. Le décret n° 2021-313, publié le 24 mars 2021 (L31, 2021b), met en place des démarches en ligne concernant le droit au séjour, notamment un téléservice pour les demandes de titre de séjour (Figure 1.1).

En novembre 2021, à la suite de plusieurs recours déposés en Guyane, la justice oblige les préfectures à offrir une alternative aux démarches numériques. Cette décision est suivie par deux décisions similaires aux tribunaux administratifs de Rennes et de Strasbourg. Ces jugements réaffirment le caractère facultatif de la dématérialisation des démarches. Les organisations sont tenues de proposer une solution non numérique pour la demande ou le renouvellement de titres de séjour (Cimade, 2021)[4]. Cependant, à Paris, il est quasiment impossible de contacter la préfecture sans passer par l'intermédiaire d'une plateforme numérique ou d'un médiateur. En 2021, 23 préfectures sont assignées en justice pour avoir forcé les usagers à faire leur demande par leur plateforme en ligne (Cimade, 2021). Ces accusations ne portent pas uniquement sur l'absence d'alternative non numérique. L'administration est accusée d'utiliser la dématérialisation pour restreindre le droit au séjour, mais aussi les droits sociaux (GISTI, 2023)[5]. Dans un article concernant la dématérialisation des démarches administratives, la chercheuse et juriste française Danièle Lochak parle d'une « politique délibérée, quoique non avouée, de dissuasion » (Lochak, 2022). Toutefois, le gouvernement semble vouloir continuer à se diriger vers un plan de dématérialisation car depuis le 6 février 2023, les demandes de naturalisation par décret doivent être obligatoirement faites sur la plateforme Anef (Figure 1.1).

Le rapport de la Défenseure des Droits intitulé « Dématérialisation des services publics : trois ans après, où en est-on ? » et publié en février 2022 corrobore ces préoccupations. Il recense un nombre de réclamations concernant les services publics en forte augmentation : 90 000 en 2021, contre 30 000 en 2014 (Defenseur des droits, 2022). Ce rapport demande à reconsidérer le « tout numérique » et préconise de « laisser à chaque usager le choix de son mode de relation avec l'administration ». C'est la fin de la pro-

4. La Cimade est une association de soutien aux étrangers en situation irrégulière, aux demandeurs d'asile et aux réfugiés fondée en 1939. Elle joue une rôle actif sur la scène associative, en proposant des permanences d'accès aux droits ainsi que différents types d'aide et d'accueil. Elle joue également un rôle politique (Drahy, 2004) et a pu récolter des données sur le terrain utilisées dans la recherche (Sadick, 2022)

5. Le Groupe d'information et de soutien des immigrés (GISTI) est une organisation militante formée en 1972. Elle contribue aux recherches dans le domaine notamment des politiques migratoires, avec la revue *Plein Droit*. L'organisation, très active dans la défense des droits des étrangers en France, est à l'origine de nombreux décrets que l'on peut retrouvé dans les articles suivants : Lochak (2009); Genevois (2009).

gression vers le tout numérique dont les diverses phases de transition s'étirent sur 30 ans. Nous présentons ci-dessous une frise chronologique de l'évolution historique de la dématérialisation des démarches d'accès aux droits en France entre 1996 et 2022 (Figure 1.1).

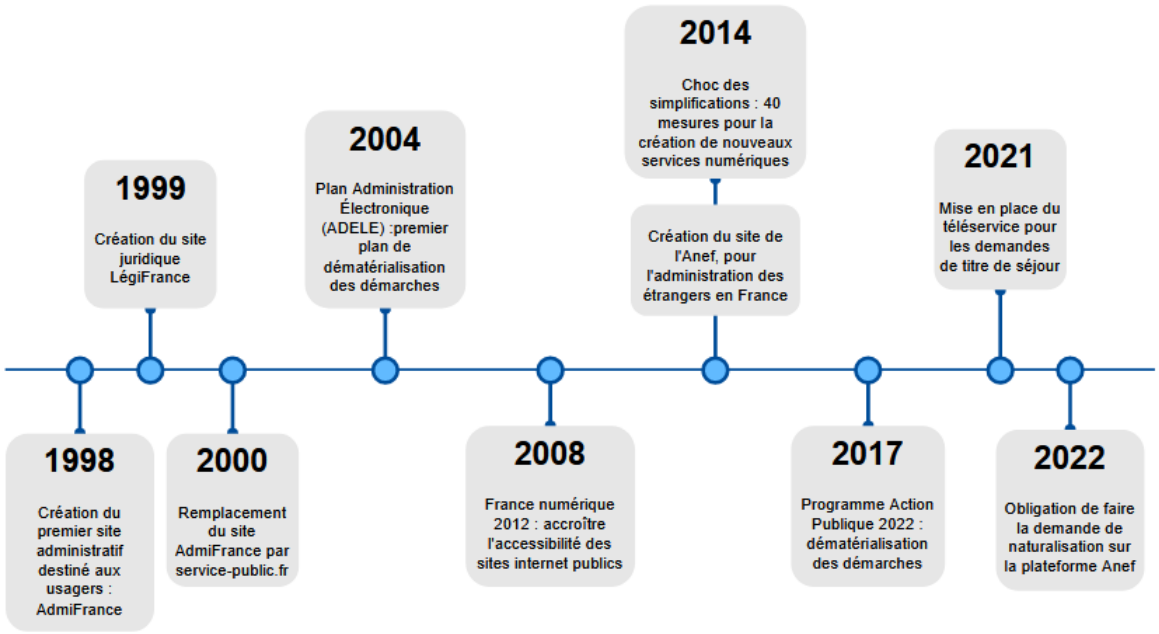


FIGURE 1.1 – Frise chronologique de l'évolution historique de la dématérialisation des démarches d'accès aux droits en France

La transformation numérique de l'administration est au cœur de problématiques sociales et politiques en France. Elle a lieu progressivement, au cours de deux décennies, et s'est intensifiée durant la dernière décennie. La transition peut désormais être considérée comme achevée, ou en phase d'achèvement, car l'ensemble des démarches sont disponibles en ligne, ou en déploiement. Les évolutions à venir vont forcément s'appuyer sur ce socle pour proposer des modifications nécessaires. Nous pensons donc qu'il est important de faire un état des lieux de cette transformation, et des transformations linguistiques et discursives qu'elle a engagées.

1.1.3 L'accès aux droits au Royaume-Uni

Au Royaume-Uni, l'histoire de la dématérialisation et de l'automatisation des démarches commence beaucoup plus tôt qu'en France. L'automatisation commence en 1944 avec le système de prélèvement de l'impôt sur le revenu à la source dénommé *PAYE* pour *Pay As You Earn*. Il s'agit d'un système de récupération des taxes par le prélèvement ou la retenue à la source instauré en temps de guerre et d'austérité économique. Cette retenue permet de gagner en efficacité et d'économiser de l'argent public

(UK Parliament, 2022). Cette mesure ne sera instaurée en France que 75 ans plus tard. Ce système, même s'il n'implique pas directement une transformation numérique, pose les jalons d'une approche d'automatisation et de personnalisation des démarches administratives et fiscales pour faciliter la collecte des taxes. Il montre que l'efficacité des démarches et leur rapidité sont un sujet important pour l'administration britannique.

Cette volonté d'automatisation et d'intégration des nouvelles technologies se poursuit dans les années 1960/1970, lorsque le gouvernement investit afin de s'équiper en ordinateurs. En 1972, on compte 4 640 agents formés en technologie de l'information (Weiss, 2019a). Toutefois, les évolutions technologiques rapides, ainsi que des réformes des services gouvernementaux mènent à plusieurs échecs de transition informatique (Agar, 2016). De plus, les innovations technologiques étant tellement rapides, le gouvernement britannique doit faire appel à plusieurs prestataires externes afin de mener à bien des projets, ce qui implique des coûts parfois élevés (Agar, 2016). Plusieurs évolutions ont toutefois lieu entre 1970 et 2000, comme la numérisation du *General Index of National Insurance Contributors* en 1982 (Figure 1.2).

En 1996, un rapport propose la création du site *Directgov* qui devient le principal site dédié à la communication entre gouvernement et citoyens (Figure 1.2). Le pays semble donc s'engager dans une transition numérique dans le domaine de la communication entre usagers et administration. Entre 2000 et 2010, le Royaume-Uni tente de mettre en place plusieurs projets afin de continuer cette transformation numérique, qui connaissent de nombreux échecs et mènent à d'importantes pertes financières (Weiss, 2019b). Ces échecs sont un moment charnière dans la transition numérique du Royaume-Uni. En effet, les fortes pertes financières forcent le gouvernement à revoir sa stratégie de transition, et à chercher de l'aide auprès d'experts externes au gouvernement, qui ont plus d'expérience avec les nouvelles technologies. En 2009, pour tenter de se redresser financièrement, le gouvernement engage un Défenseur du numérique, Martha Lane Fox, qui est une entrepreneuse issue du secteur privé (Figure 1.2). Cette dernière est chargée de faire un rapport sur le site du gouvernement *Directgov*, afin d'en identifier les lacunes. Son rôle est de donner une nouvelle direction à la transformation numérique du gouvernement en rédigeant des recommandations afin que celui-ci puisse réaménager sa transition. Dans ses recommandations publiées en 2010, Martha Lane Fox prône de nombreux changements stratégiques, qui seront le fondement de la politique gouvernementale telle qu'elle existe aujourd'hui (Figure 1.2).

Les deux premières recommandations concernent les services proposés par le gouvernement. La première incite à créer un site unique à tous les départements, capable de mettre en place des solutions sur l'ensemble du gouvernement, et d'obliger les départements à améliorer leurs rapports avec les citoyens. La deuxième recommande à

nouveau de faire un site unique, en développant les interfaces de programmation d'applications et en les rendant disponibles à des tiers. Sa troisième recommandation nous intéresse particulièrement, puisqu'elle adresse le sujet de la communication entre l'administration et ses usagers. La Défenseure du numérique indique qu'il est nécessaire de « réinventer les méthodes de publication numériques du gouvernement » (Fox, 2010). Cette recommandation veut encourager la centralisation de la publication des informations, en créant un site et une entité unique responsable de ces publications. Enfin, ces efforts de centralisation du numérique sont complétés par la quatrième recommandation, qui indique qu'il est nécessaire de nommer un « PDG du numérique », responsable de l'ensemble de la transition (Fox, 2010). Ces recommandations prônent un changement de direction radical, mais surtout un remaniement des méthodes employées, comme en témoigne le titre de son rapport « une révolution plutôt qu'une évolution »[6]. En introduisant un PDG, la Défenseure du numérique indique clairement la ligne idéologique de cette transformation : se tourner vers un fonctionnement similaire au secteur privé. Malgré des efforts très tôt vers l'automatisation des démarches administratives, et l'introduction du numérique, c'est ces recommandations qui donnent l'impulsion de créer un organisme gouvernemental dédié au numérique. En 2010, le plan nommé *Race Online 2012* est mis en place afin d'encourager les citoyens à utiliser internet et les services qui y sont proposés (Figure 1.2). En 2011, le rapport de la Défenseure du numérique conduit à la création du GDS, qui dépend du *Cabinet office* (Figure 1.2). Il est chargé de la dématérialisation des différents services du gouvernement. En 2012, le GDS met en place une nouvelle stratégie de modernisation nommée *Digital by Default*. Ce plan, comme son intitulé l'indique, a pour ambition de faire du numérique le moyen de communication par défaut entre l'usager et l'administration.

La mise en place de cet organisme unique entraîne également à la création d'un site web unique servant aux communications entre citoyens et gouvernement : le site `gov.uk`. La création de ce site unique, sur lequel il est possible d'obtenir l'ensemble des informations concernant les services publics, mais aussi de faire un grand nombre de démarches, représente l'accomplissement des changements initiés en 2009 avec l'arrivée de la Défenseure du Numérique (Figure 1.2).

Le GDS est toujours actif à ce jour et propose de nombreux services de communication, y compris des podcasts et articles de blog. La politique est de simplifier les rapports entre administrations et administrés. Le langage clair est utilisé systématiquement dans toutes les communications et le guide de style de l'organisme est accessible en ligne[7].

Nous proposons une frise chronologique des différentes étapes de la transition nu-

6. Titre complet original : "Directgov 2010 and beyond : revolution not evolution" (Fox, 2010)
7. https ://www.gov.uk/guidance/style-guide

mérique de l'administration britannique avec la Figure (1.2).

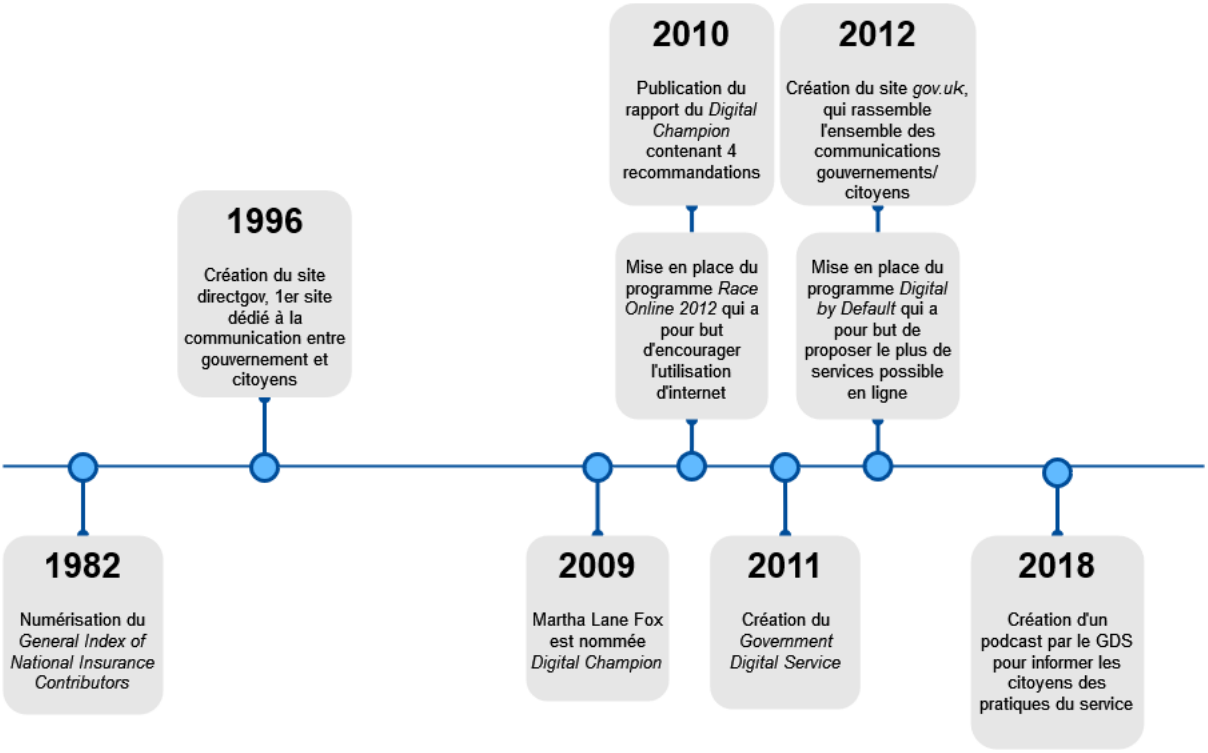


FIGURE 1.2 – Frise chronologique de l'évolution historique de la dématérialisation des démarches d'accès aux droits au Royaume-Uni

La frise chronologique retraçant les principales évolutions historiques de la dématérialisation des démarches d'accès aux droits au Royaume-Uni, présentée dans la Figure 1.2, montre que ce changement a été amorcé très tôt. Toutefois, les pertes financières écopées au début de la transition numérique ont poussé le Royaume-Uni à adopter une nouvelle stratégie d'abord issue du secteur privé, que nous explorerons plus en détail dans la section 1.2, mais qui aujourd'hui a un fonctionnement unique. La transition numérique est articulée autour des années 2009 à 2012, et particulièrement autour de la publication des recommandations par la Défenseure du numérique en 2010, qui constitue un tournant. En effet, les fondements du système tel qu'il existe aujourd'hui sont créés durant ces années charnières. La révolution numérique de l'administration britannique est depuis quelques années achevée et stable, même si elle évolue au fil des innovations actuelles. Le système mis en place est donc une communication principalement numérique, centralisée et normalisée.

Nous aborderons dans la section 1.3.2 l'émergence de publications et d'organismes parallèles, comme le Citizen Advice Bureau, qui est un organisme caritatif, financé en partie par des fonds publics.

1.1.4 France et Royaume-Uni, une vision divergente

Dans ce travail de recherche, nous aborderons le discours de l'accès aux droits sous l'angle d'une approche contrastive. Nous considérons que la similarité des dispositifs mis en place, mais aussi des politiques, permet de comparer les deux discours. La trajectoire des transformations numériques prise par les deux pays permettra d'étudier la complexité de ce discours sous plusieurs angles, tout en faisant ressortir les traits communs à ce discours dans les deux langues. Nous prendrons en compte une partie de l'histoire et des différentes politiques des deux pays ce qui permettra de réaliser une caractérisation informée. Les deux pays étudiés présentent deux histoires distinctes. Tandis que la France multiplie les dispositifs et sites internet, le Royaume-Uni décide de faire de `gov.uk` le lieu d'interaction principal avec l'administration. Ces différences de politique créent une disparité dans la masse discursive britannique et française, dont nous présenterons les éléments significatifs d'analyse dans les chapitres 4 et 5. Cependant, nous pouvons également observer de nombreux points communs entre les deux stratégies gouvernementales. En effet, le Royaume-Uni décide de s'inspirer du secteur privé et de faire appel à Martha Lane Fox, entrepreneuse issue du privé en 2009 à la suite d'échecs de mise en place de certains dispositifs. Le gouvernement français décide de faire de même quelques années plus tard en cherchant à imiter l'esprit « start-up » du privé. Ces deux transformations sont toutes deux de nature idéologique, et concomitantes à la transformation numérique. La transformation (ou révolution) numérique de l'administration, mais aussi plus généralement du gouvernement et de la société, pousse à reconsidérer nos modes de fonctionnements et nos relations. Dans ce contexte, l'État a également engagé une reconfiguration de la relation entre administration et usagers.

1.2 Une double transformation : numérique et idéologique

Au Royaume-Uni comme en France, la dématérialisation des services publics s'accompagne d'une transformation idéologique issue du *New management*. Elle est décrite ainsi par Chappoz et Pupion (2012, p. 1) :

> une nouvelle forme de gestion publique basée entre autres sur une culture du résultat et l'emprunt de pratiques et d'outils issus du privé.

Les objectifs de cette transformation sont de gagner en efficacité et de replacer la responsabilité sur l'usager. Lochak (2022) décrit ces objectifs :

> Le « tout numérique » s'impose ainsi comme la nouvelle norme du service public, au cœur des réformes de modernisation de l'État. L'entreprise s'ins-

> crit dans une logique managériale qui promeut l'efficacité et la réduction des coûts, conformément aux principes de légitimation du nouveau management public, calé sur les modèles du privé. (Lochak, 2022, p. 3)

La juriste met en avant ici les mécanismes économiques et administratifs mis en place par le gouvernement. De plus, elle rappelle que la transformation numérique fait partie d'un ensemble de réformes qui visent à moderniser l'État. Plus récemment en France, la révolution numérique a mené à la création d'un nouvel imaginaire politique de l'État : celui de l'État start-up. Les services de l'État adoptent le modèle managérial des start-up (Levasseur, 2021). C'est ce qu'explique Henri Verdier dans un entretien avec Julie le Bolzer, journaliste des Échos :

> Il est impératif d'injecter au sein de l'État les méthodes (agiles, mode projet...), les valeurs, mais aussi les profils de la révolution numérique. (Le Bolzer, 2019)

Cette transformation idéologique a des conséquences sur différents aspects de la révolution numérique des administrations, et de l'État en général. Nous allons tout d'abord présenter la direction prise par les gouvernements lors de la transition numérique, celle de l'État plateforme, puis nous ferons un point sur la notion de conditionnalité dans le domaine de l'accès aux droits, qui prend de plus en plus d'importance. Nous présenterons ensuite le concept de charge administrative qui est un facteur important dans l'étude du discours. Ces analyses nous permettront d'envisager la transformation numérique des administrations dans leur contexte politique et sociologique.

1.2.1 État plateforme : une idéologie aux origines américaines

L'État plateforme fait partie intégrante de la transformation idéologique qui fut à l'œuvre en France, au Royaume-Uni, mais aussi dans de nombreux pays d'Europe et d'Amérique du Nord à partir des années 2010. Il s'agit d'une idéologie émanant des États-Unis, puis reprise et adaptée dans les deux pays qui concernent notre étude. La notion d'État plateforme, ou de gouvernement 2.0 est abordée pour la première fois par Tim O'Reilly, un entrepreneur spécialisé dans l'informatique résidant aux États-Unis. Après des études de Lettres classiques, Tim O'Reilly évolue dans le milieu de l'édition et se consacre dès 1983 à la publication de manuels d'informatique en lançant sa maison d'édition (O'Reilly & Associates devenu O'Reilly Media) spécialisée dans l'informatique, et dont le siège se trouve aujourd'hui à Sébastopol en Californie[8]. Dans son article

8. Il a par exemple popularisé les concepts d'open source, et de web 2.0, qu'il a consacré sa carrière a étudier. Il a publié de nombreux livres sur les technologies numériques et leur place dans la société.

consacré au gouvernement 2.0 (O'Reilly, 2011), il reproche à l'État d'être un distributeur automatique : l'usager utiliserait les services dont il a besoin, sans jamais être actif. Ce mode de fonctionnement créerait donc des usagers dépendant de l'État. À l'inverse, l'État plateforme est imaginé comme une solution alternative où l'ensemble des acteurs privés, publics et des particuliers participent à l'élaboration de services. O'Reilly a pour ambition d'amener le succès des grandes entreprises comme Google ou Apple dans le gouvernement. Ces grandes entreprises utilisent des plateformes (comme le *Play Store* ou *App Store*) afin d'encourager les entrepreneurs à innover en créant des services, des applications, etc., puis en les rachetant pour les intégrer à leur offre. Pour réaliser cela, le numérique est bien sûr une composante essentielle. Dans son ouvrage, il détaille les composantes essentielles de l'État plateforme dans sa description du gouvernement 2.0 :

> *Government 2.0, then, is the use of technology — especially the collaborative technologies at the heart of Web 2.0—to better solve collective problems at a city, state, national, and international level.* O'Reilly (2011, p. 14)

L'entrepreneur fait référence aux deux piliers de cette idéologie : la collaboration et les nouvelles technologies. Il s'agit d'allier les deux afin de créer un nouveau type de gouvernement. Les gouvernements se tournent donc vers la création de plateformes de données ouvertes, comme le site `data.gouv.fr` en France ou `data.gov.uk` au Royaume-Uni. Ils souhaitent ainsi encourager le développement d'initiatives de la part des citoyens afin de développer des services publics. En France, cette théorie est reprise par Nicolas Colin et Henri Verdier en 2015 dans l'ouvrage *L'Âge de la multitude, entreprendre et gouverner après la révolution numérique*. Les auteurs se réapproprient le concept de l'État plateforme en y ajoutant le concept de « multitude », une force créatrice contenue dans le peuple, et qui peut être exploitée par le gouvernement (Jeannot, 2020). Cette multitude est décrite ainsi par les auteurs :

> L'ensemble des individus pouvant créer de la valeur dans une organisation sans pour autant être employés ou mandatés par cette organisation, c'est ce que nous avons déjà appelé la multitude. La multitude du point de vue des entreprises est en quelque sorte du capital humain demeuré en dehors de l'organisation et non accessible dans le cadre d'une relation entre un client et son fournisseur. (Colin et Verdier, 2015, p. 92)

Dans un article retraçant l'histoire de l'État plateforme en France publié en 2020, le chercheur Gilles Jeannot fait l'état d'une utopie décrite dans l'ouvrage de Colin et Verdier, qui prendrait la forme suivante :

> Imaginons que 100 applications proposent aux contribuables des interfaces différentes, des expériences innovantes pour déclarer leur revenu et payer leurs impôts en ligne. [...] Que 100 applications mélangent les ressources

> d'administrations distinctes pour proposer un service unique intégrant plusieurs guichets : déclarer une naissance, inscrire son enfant à l'école, prendre rendez-vous pour une consultation à l'hôpital, déposer une plainte au commissariat. Que ces mêmes services se mélangent à d'autres n'ayant rien à voir avec l'administration : pouvoir faire ses courses, échanger des opinions sur des films ou des livres, et, dans la même application, demander une aide au logement à la caisse d'allocations familiales. Imaginons encore que tous les services rendus par l'administration puissent l'être dans notre compte Gmail, sur notre page Facebook ou en marge de notre fil Twitter. (Colin et Verdier, 2015, p. 250-251)

Les auteurs décrivent ainsi la forme que pourrait prendre cette idéologie en France. Cet ouvrage a d'autant plus d'impact dans le paysage politique français que Henri Verdier est nommé directeur interministériel du numérique et du système d'information et de communication de l'État (DINSIC) en septembre 2015. Il est ensuite nommé Ambassadeur du numérique en 2018. Il joue donc un rôle majeur dans la transformation numérique de l'État et met en place de nombreux dispositifs afin d'encourager les start-up à créer des services.

Le Royaume-Uni s'est également saisi de ce concept qui a pris la forme d'un programme nommé *Government as a platform (GaaP)*, qui a vu le jour en 2013 suite à l'ouvrage de Tim O'Reilly. Un des principes de ce programme est résumé dans cette phrase :

> *Build once, use many.* (Heaton, 2021)

Son rôle est donc de gagner en efficacité afin d'éviter des pertes financières, mais aussi d'augmenter l'interopérabilité des services numériques. Dans ce cas, la simplicité d'utilisation et de mise en place est vue comme un atout majeur pour les systèmes. La centralisation de ces systèmes est également un élément important de ce programme. Les retours des usagers sont essentiels dans la mise en place de ce projet. L'objectif annoncé de ce programme est de construire une infrastructure commune pour les systèmes et les process numériques. Cette infrastructure faciliterait alors la mise en place de services publics centrés sur l'utilisateur (Bracken, 2015). Afin de construire une infrastructure commune, mais aussi de mettre en place des bonnes pratiques communes, l'organisme GDS a mis en place une approche ascendante permettant aux employés de s'impliquer dans les projets afin de créer une cohésion entre les différents départements.

L'État plateforme a mené le Royaume-Uni à une forme centralisation de l'ensemble des services publics proposés par l'État, avec au centre, la plateforme gov.uk. Cette centralisation et normalisation de la communication numérique interne et externe a été portée par l'organisme GDS qui s'est installé dans le rôle de coordinateur des différents

départements. En revanche, en France, le concept d'État plateforme a rapidement été repris par des entrepreneurs français comme Henri Verdier, afin de développer des start-up d'État. Ces start-up d'État ont mené à une décentralisation des plateformes d'accès aux services publics et à une multiplication des sources d'informations, comme nous le présentons dans la section 1.3.1. Les deux pays ont donc pris deux directions différentes en ce qui concerne l'État plateforme. Il est toutefois important de noter que récemment, le gouvernement français semble prendre un tournant similaire au Royaume-Uni, en débutant une centralisation de l'administration numérique qui s'est illustrée avec la refonte de `service-public.fr` et l'expansion du service d'identification France Connect.

La notion de multitude introduite dans les années 2015 n'a pas complètement disparu des évolutions amenées par le numérique et se retrouve notamment dans le concept devenu connu de *crowdsourcing*, traduit en français par *collaboration de masse*. Il s'agit d'une pratique encore présente, qui vise à encourager l'engagement citoyen, comme le montre par exemple la résurgence du concept de *science participative* ou encore le dispositif DansMaRue mis en place par la ville de Paris, une application qui a pour but d'améliorer l'espace public et de lutter contre les incivilités (Mericskay, 2021). Toutefois, dans le domaine des politiques de diffusion des informations, le crowdsourcing a une place moins centrale.

1.2.2 Conditionnalité

La conditionnalité est une notion qui est de plus en plus importante dans le domaine de notre étude. Elle est le résultat d'un changement d'idéologie politique qui accompagne les années de transition numérique. La conditionnalité fait référence au fait que certains droits sont conditionnés par des obligations. Willmann définit la conditionnalité ainsi :

> La notion de « conditionnalité », telle qu'elle est entendue en droit social au sens large (droit de l'emploi, droit de la protection sociale), s'entend d'un certain nombre de conditions (d'où l'expression « conditionnalité ») imposées aux chômeurs et bénéficiaires de minima sociaux, autres que celles généralement admises au bénéfice d'un revenu de remplacement (condition d'âge, de santé, de non-activité, de période minimale d'activité antérieure, de résidence, d'inscription sur la liste des demandeurs d'emploi, d'actes positifs de recherche d'emploi...). Willmann (2012, p. 1)

Par exemple, en France comme au Royaume-Uni, l'obtention des droits au chômage est conditionnée par un entretien professionnel dans les locaux des agences concernées, dont la date et l'horaire sont dictés par l'agence. L'usager doit donc être en mesure de s'y

rendre. Il ne s'agit ici que d'exemples, mais une grande partie des aides sociales, et des droits au logement, séjour, etc., sont accessibles uniquement sous conditions (explicites et implicites). En France, on peut par exemple penser à l'obligation de faire son actualisation sur le site Pôle Emploi chaque mois afin de rester inscrit comme demandeur d'emploi (une condition indispensable pour recevoir des allocations). Cette condition, explicite, en implique une autre qui est implicite : avoir accès à un ordinateur ou à un téléphone chaque mois pour se connecter à son compte. Un oubli d'actualisation entraîne une radiation immédiate, et oblige l'usager à recommencer le processus d'inscription.

L'ensemble de ces conditionnalités qui viennent s'ajouter en plus de « celles généralement admises » témoignent d'un changement d'idéologie. Ce changement est amorcé au début des années 2000 au Royaume-Uni. Le sociologue et conseiller politique Guiddens (1998) propose dans un ouvrage d'encourager « l'esprit d'entrepreneuriat ». Pour Dwyer (2004), il s'agit d'une renégociation des rapports entre l'État et ses citoyens. On passe d'une *Welfare society* à une *active society*. L'État met en place un certain nombre d'obligations à remplir afin d'accéder aux droits. Ces obligations peuvent être implicites ou explicites.

Au Royaume-Uni, ces obligations n'ont fait que s'étendre à un plus grand nombre de droits ces dernières années (Dwyer et al., 2020). Dans la dernière décennie, le grand changement concernant les droits sociaux s'est traduit par l'introduction de l'*Universal Credit*, une aide financière qui remplace six dispositifs. Afin de recevoir cette aide financière, les usagers doivent signer un engagement (nommé *Claimant Commitment*). Il s'agit d'un contrat qui définit les conditions d'obtention de l'aide. Contrairement aux dispositifs précédents, les conditions d'obtention sont très personnalisées et sont assorties de plusieurs types de sanctions : amendes, radiation (jusqu'à 3 ans), et poursuites judiciaires si une fraude est avérée (Dwyer et Wright, 2014). D'après Bennett (2012), ce dispositif élargit la conditionnalité qui s'applique pour les bénéficiaires, qu'ils travaillent ou non. Le dispositif *Universal Credit* a donc introduit un nouveau système de conditionnalité, mais il a aussi déplacé sur les citoyens les responsabilités de l'État et des employeurs en matière d'aide sociale (Dwyer et Wright, 2014, p. 33).

En France, la mise en place du Revenu de solidarité active (RSA) en 2008 correspond également à cette logique de conditionnalisation et de contractualisation des aides. Le RSA, qui remplace le revenu minimum d'insertion (RMI), introduit des conditions à remplir pour le bénéficiaire. La chercheuse Élisa Chelle s'intéresse à l'introduction de ce dispositif et indique que « le pendant institutionnel de la conditionnalité est clairement réaffirmé » (Chelle, 2019, p. 163). On retrouve également la notion d'engagement avec l'introduction d'un contrat d'accueil et d'intégration pour les étrangers souhaitant obtenir une carte de séjour (1 an) ou de résidence. L'obtention d'un titre de séjour a toujours

été soumise à un certain nombre de conditions, néanmoins la Loi du 24 juillet 2006 les renforce en ajoutant notamment une condition linguistique qui implique une connaissance « suffisante » de la langue française. L'article de loi suivant montre une volonté d'étendre la conditionnalité :

> La délivrance d'une première carte de résident est subordonnée à l'intégration républicaine de l'étranger dans la société française, appréciée en particulier au regard de son engagement personnel à respecter les principes qui régissent la République française, du respect effectif de ces principes et de sa connaissance suffisante de la langue française. (L31, 2021a)

L'utilisation du terme « subordonnée à » indique que de nouvelles conditions sont introduites. La loi du 7 mars 2016 renforce les conditions d'évaluation du niveau linguistique et du respect des « principes qui régissent la République française ». En 2020, de nouvelles exigences linguistiques sont introduites pour la naturalisation par décret. Une nouvelle version du Code de l'entrée et du séjour des étrangers et du droit d'asile (Ceseda) est également entrée en vigueur au 1er mai 2021, avec des mises à jour régulières [9] (Ces, 2023). Elle renforce par exemple les conditions liées à l'accueil des ressortissants étrangers avec un visa court séjour en France. Pour accueillir un ressortissant étranger (hors Europe), il est désormais obligatoire de faire établir une attestation d'accueil à la mairie de son domicile. Cette attestation est soumise à des conditions financières et matérielles pour la personne accueillant, et pour la personne accueillie (L31, 2021b).

Une seconde partie de conditionnalité vient naturellement de la dématérialisation des démarches. En effet, la disparition des guichets et les difficultés de plus en plus grandissantes à contacter les administrations par un canal non numérique rendent obligatoire l'utilisation du numérique. Pierre Mazet (2019), chercheur en sciences sociales, introduit la notion de « conditionnalités implicites » du numérique. D'après lui, la transformation numérique crée une « montée en conditionnalité » qui n'est pas reconnue, mais pourtant bien présente. En effet, aux premiers abords, la dématérialisation semble rendre les démarches plus accessibles, car il n'y a plus besoin de se déplacer, mais d'après Mazet la situation est plus complexe car la dématérialisation rendrait les démarches mois accessibles et donc l'accès aux droits plus restreint :

> il est à présent exigé et considéré comme relevant de leur responsabilité que les individus possèdent (au double sens de la matérialité et des compétences) le médium numérique. (Mazet, 2019, p. 2)

Cette conditionnalité est matérielle : l'usager doit posséder un matériel adéquat pour réaliser ses démarches (ordinateur, téléphone portable, connexion internet, etc.). Cependant, elle concerne aussi les compétences numériques, car l'usager doit savoir utiliser les

9. Dernière mise à jour le 26 juin 2023

sites internet et les outils numériques à sa disposition afin de réaliser les démarches. Plus l'usager est capable de naviguer efficacement sur internet, plus il lui sera facile d'accéder à ses droits.

La transformation numérique est donc accompagnée d'une transformation idéologique de l'administration qui apporte une conditionnalité croissante pour l'accès aux droits. Cette conditionnalité est explicite, avec un plus grand nombre de conditions à remplir pour l'usager, mais aussi implicite, avec l'utilisation du numérique quasi obligatoire. Elle place une grande responsabilité sur l'usager, ce qui est sans doute une charge lourde.

1.2.3 Charge administrative - *Administrative burden*

Nous avons montré que la conditionnalité a une place grandissante dans les relations entre administrés et administration. Nous allons maintenant introduire la notion de charge administrative qui a une place importante dans la transformation numérique de l'administration. La charge administrative existe lorsqu'un citoyen perçoit une démarche administrative comme pénible ou contraignante (Burden et al., 2012, p. 741). Un grand nombre de démarches administratives, à partir du moment où elles sont restreintes à une certaine partie de la population, présentent une charge car l'administré doit prouver qu'il est éligible. Les différents types de charges sont détaillés dans la Table 1.1 ci-dessous, telles que décrites par Moynihan et al. (2015).

Type de charge	**Application**
Pédagogique	Les citoyens doivent s'informer sur le programme, savoir s'ils sont éligibles, connaître la nature des prestations et savoir comment accéder aux services.
Psychologique	Les citoyens peuvent être stigmatisés lorsqu'ils reçoivent certaines aides. Ils subissent une perte d'autonomie et du stress lié aux démarches.
Conformité	Les citoyens doivent remplir des formulaires de demande et de renouvellement, fournir des documents attestant de leur statut.

TABLE 1.1 – Les différents types de charges administratives

D'après la Table 1.1, les coûts peuvent être pédagogiques, psychologiques et de conformité. Ils relèvent de plusieurs types de compétences et sont susceptibles d'affecter les administrés de différentes façons. Pour réduire la charge pédagogique, il semble important de donner des informations claires et faciles d'accès. D'après certaines études, cette charge peut être considérablement réduite en envoyant des informations et des rappels aux personnes éligibles. Ces actions mènent à une augmentation du nombre de bé-

néficiaires des différentes aides (Bhargava et Manoli, 2012). Le coût psychologique joue également un rôle important dans le recours, ou non, des citoyens à leurs droits. En effet, les citoyens peuvent être fortement stigmatisés lorsqu'ils sont bénéficiaires de certaines aides. Cette stigmatisation peut être aggravée par les rapports entre le gouvernement et les administrés (Moynihan et al., 2015). D'après Holler et Tarshish (2022), l'attente, le manque de communication et les erreurs administratives pourraient renforcer cette stigmatisation et détériorer les rapports entre administration et administrés. Les chercheurs remarquent que l'attente est un marqueur de relation hiérarchique. Cela signifie que l'attente vécue par les citoyens établit un rapport de pouvoir entre le gouvernement et le citoyen. Ce rapport de pouvoir peut être renforcé par le manque de communication : les informations ne sont pas personnalisées ou l'usager n'arrive pas à obtenir des informations sur son dossier. On remarque que la notion de temps n'est pas soulignée dans ces rapports or chacune des trois charges implique un coût en temps. Pourtant toute forme d'intégration du citoyen dans le système socio-économico-administratif l'est : écolier, étudiant, apprenti, salarié… ont tous des volumes horaires bien définis pour avoir droit à ces statuts, et par conséquent aux droits qu'ils renferment. Tout citoyen, s'il peut jouir de ses droits, doit utiliser une partie de son temps pour en faire les demandes (à l'exception de quelques allocations automatiques comme l'allocation jeune enfant). Ce coût en temps serait à calculer par les administrations afin de les en informer.

Enfin, des recherches ont montré que la charge administrative pourrait être un outil utilisé par les gouvernements afin de limiter l'accès aux droits (Baekgaard et al., 2021; Piven et Cloward, 2012). Madsen et al. (2022) montrent que la dématérialisation des démarches peut renforcer cette charge administrative. En effet, les démarches sont maintenant faites en autonomie, sur un ordinateur ou un téléphone portable. Il est donc nécessaire d'avoir des connaissances préalables sur les démarches à faire, contrairement aux démarches en présentiel où un agent est disponible afin de fournir les informations nécessaires.

En étudiant la transformation numérique des administrations et les changements idéologiques qui l'ont accompagnée, des experts notent qu'un des buts principaux des gouvernements est de gagner en efficacité (Girault, 2017). Les démarches en ligne permettent de traiter les dossiers directement. De plus, le numérique permet également d'utiliser des algorithmes, comme Parcours Sup, ou le logiciel de contrôle de la Caf, qui réduisent considérablement le temps de travail des agents administratifs (Dubois et al., 2018; Frouillou et al., 2019)[10]. Avec le numérique, la responsabilité est replacée sur le

10. A noter que Parcours Sup peut induire une charge de travail supérieure pour les professeurs de l'enseignements supérieur et même parfois pour les étudiants. Toutefois, ce surplus de temps est induit par un déplacement des responsabilité sur le logiciel, qui réalise une tâche précedemment faite par des agents administratifs, mais qui crée de nouvelles tâches, a réaliser par les étudiants et les professeurs, et non plus

citoyen, face à la machine, qui est chargé de chercher les informations adéquates, de se procurer les papiers, de remplir les formulaires et de les transmettre. Henri Verdier, directeur interministériel du numérique et du système d'information de 2015 à 2018, et depuis 2018 ambassadeur du numérique fut responsable de la transition numérique de l'État et des administrations. Dans son ouvrage, co-écrit avec Colin et Verdier (2015), il expose son idéologie de l'État plateforme, inspiré par l'article de Tim O'Reilly et que nous avons présenté dans la section 1.2.1. L'État plateforme suppose une participation active des citoyens et s'oppose à un gouvernement « distributeur » (Kettl, 2008) où l'on paierait des impôts pour recevoir des services. Le citoyen devrait devenir « co-responsable et coproducteur des biens publics » (Algan et Cazenave, 2017, p. 13). Il aurait donc la responsabilité de la bonne marche de la société. C'est ce que Colin et Verdier décrivent dans leur ouvrage en indiquant qu'il est « préférable de laisser les citoyens régler eux-mêmes les problèmes qu'ils peuvent régler, de les y aider et de les en valoriser » Colin et Verdier (2015, p. 249).

Ce replacement de la responsabilité sur l'usager est bien présent dans la transformation numérique où il est de la responsabilité du citoyen d'aller en ligne afin d'effectuer ses démarches, de remplir ses formulaires, de téléverser ses documents administratifs, etc. La juriste Michèle Lockart décrit un usager qui aurait une position d'entrepreneur :

> Un public dont on attend que, davantage responsabilisé, il devienne l'entrepreneur de sa propre vie administrative. (Lochak, 2022, p. 3).

Comme nous l'avons montré dans la section 1.2.2, la transition numérique de l'administration demande une montée en compétence chez les utilisateurs. La charge administrative qui lui incombe est donc alourdie, et il se trouve responsable de certaines tâches qui, auparavant, étaient réalisées par des guichetières et guichetiers. C'est un principe que l'on retrouve dans les *Governement Design Principles* du gouvernement britannique, qui sont les grands principes de rédaction du site web officiel du gouvernement. Le second principe de rédaction est « *Do less* » (Central Digital and Data Office, 2019), et indique que le gouvernement ne doit réaliser que le minimum nécessaire, et s'appuyer sur les autres acteurs du domaine et les usagers afin de se concentrer sur l'essentiel.

En conclusion, la transformation numérique de l'État n'est pas un simple changement de support : du papier au numérique. La dématérialisation des démarches administratives implique une transformation profonde des rapports entre administrations et usagers. Elle est guidée par des volontés politiques et sociologiques, et répond à des objectifs idéologiques. Notre étude du discours de l'accès aux droits doit donc prendre en compte ces éclairages afin de proposer une analyse complète.

---

par le personnel administratif, qui s'est trouvé supposément remplacé par la machine.

1.3 Paysage documentaire du discours numérique de l'accès aux droits

Dans les deux parties précédentes, nous avons présenté l'histoire de l'accès aux droits au Royaume-Uni et en France. Nous avons également introduit des concepts indispensables pour comprendre l'enjeu et la place de ces discours dans la société de chacun de ces pays. Dans cette troisième partie, nous allons présenter le paysage documentaire de ce discours dans les deux langues. L'étude de ce paysage documentaire permettra d'identifier les différentes plateformes numériques où il se développe, mais aussi la place du discours émanant de sources officielles qui fait l'objet de notre étude. En recensant l'ensemble des plateformes numériques du discours de l'accès aux droits, nous pourrons étudier la place du discours officiel dans l'ensemble des discours.

1.3.1 France

En France, il existe de nombreux sites officiels concernant différents aspects de l'accès aux droits [11]. Le site `legifrance.gouv.fr` est le site officiel du gouvernement pour la diffusion de l'ensemble des textes législatifs. Ce site forme un socle de connaissances et de textes juridiques accessibles à partir desquels une grande partie des textes d'accès aux droits sont rédigés. Concernant le discours de l'accès aux droits, le site le plus connu est le site officiel de l'administration française, `service-public.fr`, qui est édité et mis en ligne par la direction de l'information légale et administrative (DILA). Il rassemble l'ensemble des informations concernant tous les domaines de l'accès aux droits. La DILA est ue administration publique française placée sous l'autorité du Premier ministre, dont le rôle est de diffuser le droit en France [12]. Il est intéressant de noter qu'une partie de ce site a été reproduit sur un autre site. Ce second site, issu du domaine public également, propose une ergonomie et un design complètement différent, mais réutilise les mêmes textes. Il s'agit du site `demarches.interieur.gouv.fr`, publié par le ministère de l'Intérieur. [13]

En parallèle, chaque organisme responsable d'une catégorie de droit possède un site qui lui est propre. C'est le cas pour la Caisse d'allocations familiales (Caf), avec le site `caf.fr`, de l'organisme Pôle emploi, avec le site `pole-emploi.fr` [14], de la Maison Départementale pour les Personnes Handicapées (MDPH), avec le site `mdphenligne.cnsa.fr` et de chacune des préfectures de France. Les différents services du Crous

11. l'ensemble des URL présentées dans cette section ont été consulté le 20/10/2023
12. https ://www.dila.premier-ministre.gouv.fr/, consulté le 20/10/2023
13. En date du 16 novembre 2023, ce site n'est plus disponible.
14. Pôle Emploi a été renommé France Travail en janvier 2024. Toutefois, notre corpus a été récolté avant cette date, et nous continuons donc d'appeler l'organisme Pôle emploi tout au long de cette thèse.

sont également disponibles sur plusieurs sites, correspondant généralement à la ville ou à l'académie, ou bien à ses services comme le site permettant de trouver un logement : `trouverunlogement.lescrous.fr`.

Avec la publication du décret n° 2021-313, le 24 mars 2021, la direction générale des étrangers en France met en place des démarches en ligne concernant le droit au séjour, notamment un téléservice pour les demandes de titre de séjour. Ces téléservices s'étendent à partir du 6 février 2023 à la demande de naturalisation par décret, qui doit obligatoirement se faire en ligne par la plateforme. Ces démarches sont disponibles sur le site suivant : `administration-etrangers-en-france.interieur.gouv.fr`, souvent désigné par son sigle : Anef. Les sites des mairies proposent également des sites avec des informations concernant l'accès aux droits, notamment les grandes villes. Voici quelques exemples de sites publiés par les mairies recensant des informations concernant l'accès aux droits : `paris.fr`, `marseille.fr`, `lyon.fr` et bien d'autres encore.

En résumé, l'ensemble des services publics de l'État possède en même temps une page sur `service-public.fr`, mais aussi un site dédié à ce service ou ce type de service. Ces sites permettent d'obtenir des informations sur les étapes des démarches et les conditions d'obtention des droits et services, mais aussi de réaliser ces services. Les mêmes informations concernant un certain type de démarche sont donc souvent présentes sur plusieurs sites, mais rédigées par différents organismes gouvernementaux. C'est le cas par exemple des informations relatives aux droits au séjour. Ces informations sont bien sûr disponibles sur `service-public.fr`, mais également sur le site du ministère de l'Intérieur, ou encore sur le site de l'Anef, enfin les informations sont également disponibles sur un site édité par l'agence Business France, qui se charge notamment d'attirer les travailleurs étrangers en France [15].

La Figure 1.3 propose une illustration par catégories des différents sites d'accès aux droits disponibles en France. Il ne s'agit pas d'un schéma exhaustif mais d'une représentation visuelle des différentes catégories existantes et de la multiplicité des sites. Le discours de l'accès aux droits issus d'organismes officiels est caractérisé par la pluralité de ses sources et de ses plateformes.

15. https ://www.businessfrance.fr/, consulté le 20/10/2023



FIGURE 1.3 – Schéma récapitulatif des sites officiels d'accès aux droits en France

Il existe des entreprises privées qui proposent des services d'aide à l'accès aux droits, et donc également des sites. Le plus connu et utilisé en France est le site `mes-allocs.fr`, qui propose un accompagnement administratif sous forme d'abonnement, à 29,90 € par trimestre. Cette entreprise a également un site entièrement gratuit, qui propose des fiches explicatives détaillées sur chacune des aides. Il est important de noter que la légalité de ce site demeure en question car l'association nationale des assistants de service social a déposé plainte en 2019 en indiquant que les prestations sociales « ne peuvent faire l'objet de commerce » (Conseil d'Administration de l'ANAS, 2022). Pour l'instant, le site est toujours disponible en date du 3 septembre 2024. Il existe d'autres entreprises privées, réalisant un profit dans le sillon de celle-ci. Les prestataires de services aidant dans la réalisation d'un visa sont notamment très nombreuses. Mais ces services sont quelque peu différents, car ils se fondent sur les lois de pays étrangers dans lesquels l'usager souhaite se rendre. Néanmoins, leurs sites présentent un discours d'accès aux droits, car il indique les démarches administratives à réaliser afin d'obtenir un visa. Il ne s'agit pas du droit français, et c'est pourquoi nous ne les incluons pas dans cette étude.

En plus des sites web, les réseaux sociaux sont des lieux importants pour le partage des informations concernant l'accès aux droits. Certains organismes officiels sont présents sur les réseaux, comme la Caf, ou Pôle emploi qui possèdent tous deux un compte Instagram. Leur présence reste toutefois marginale, et très peu d'informations pertinentes sont communiquées sur ces plateformes par les instances officielles. À titre d'exemple, le dernier poste publié par le compte Instagram de la Caf date de janvier

2023 et consiste à souhaiter de meilleurs vœux aux allocataires. Au 3 septembre 2024, le compte Instagram de la Caf n'est plus accessible et semble avoir été effacé.

Les usagers, eux, sont très actifs et créent de nombreux groupes d'entraide. Pour cela, deux sites regroupent l'ensemble des initiatives d'entraide des usagers : Reddit et Facebook. Il existe sur Facebook un grand nombre de groupes privés, destinés à donner des informations sur les différentes démarches. Chaque groupe est spécialisé dans un domaine (naturalisation, allocations familiales, statut d'artiste auteur, etc.). Dans certains de ces groupes, des professionnels des organismes officiels participent aux discussions et répondent aux questions sur leur temps libre.

Enfin, les réseaux sociaux vidéos tels que YouTube ou TikTok sont également concernés par ce discours. En effet, on retrouve de nombreuses vidéos d'aide à l'accès aux droits. Certaines vidéos sont postées par les organismes eux-mêmes [16]. Différentes entreprises, des organismes de formation en français [17] aux agences immobilières, [18] proposent des contenus concernant l'accès aux droits. Certaines chaînes, comme la chaîne Quid Droit, sont consacrées uniquement à l'accès aux droits et sont présentes sur YouTube, Instagram et TikTok. Cette chaîne est animée par un étudiant en droit, qui semble avoir constitué un savoir professionnel de par sa formation, et non-professionnel au fil de ses interactions avec l'administration et de ses recherches. Nous abordons les différents types d'expertises dans la section 2.2.3 du chapitre 2.

L'étude du paysage documentaire français du discours de l'accès aux droits permet de constater que le discours est présent à différents niveaux de la société et concerne différents acteurs. Du côté des instances officielles, nous observons une fragmentation du discours et une multiplication des plateformes de communication unidirectionnelle (Figure 1.3). Du côté des usagers, nous remarquons également une diversité des plateformes et des méthodes de communication, mais aussi des types de communications : forums, vidéos, discussions, etc.

Un résumé du paysage documentaire comprenant les types de plateformes, les acteurs et leurs objectifs est proposé dans la Table 1.2.

16. vidéo concernant l'ASF postée par les Allocations Familiales : https ://www.youtube.com/watch ?v=YG4uavR4i8U, consultée le 10 octobre 2023
17. https ://www.youtube.com/watch ?v=BqDHFfh2PP0, consultée le 10 octobre 2023
18. https ://www.youtube.com/watch ?v=m8qWQvbZN0o, consultée le 10 octobre 2023

Acteurs	Organismes gouvernementaux	Organismes privés	Usagers
Plateforme principale	Site internet public	Site internet privé	Réseaux Sociaux
Plateformes secondaires	Réseaux sociaux (YouTube)	Réseaux sociaux	Forum
Objectifs	Diffusion des informations	Aide aux usagers contre rémunération	Compréhension de l'information et entraide

TABLE 1.2 – Récapitulatif des différents lieux de production du discours de l'accès aux droits en France

1.3.2 Royaume-Uni

Nous allons maintenant présenter le paysage documentaire du Royaume-Uni. Tout d'abord, comme en France, le gouvernement britannique propose un site unique afin de publier les textes législatifs : `legislation.gov.uk`. C'est à partir des textes juridiques publiés sur ce site que l'ensemble des informations sur le site `gov.uk` sont publiées.

Comme nous l'avons dit précédemment, le site principal de l'accès aux droits au Royaume-Uni est le site : `gov.uk`, résultant de la fusion de sites précédemment édités par chaque département, et supprimés depuis. Son objectif est de rassembler des informations concernant l'accès aux droits. Dans le paysage documentaire du discours de l'accès aux droits britannique officiel, il joue le rôle de figure centrale et quasi unique. Lors de la mise en place de l'aide *Universal Credit*, pour accompagner la transition, le gouvernement a mis en place un site dédié à ce dispositif d'aide qui est le suivant : `understanduniversalcredit.gov.uk`. Il bénéficie de son propre design graphique, mais est cependant rattaché et édité par l'organisme GDS, responsable de la diffusion des informations sur `gov.uk`. Il est également important de noter que l'Irlande possède un site dédié indépendant de `gov.uk` : `nidirect.gov.uk`.

La Figure 1.4 propose une illustration des sites d'accès aux droits officiels au Royaume-Uni. Ce schéma montre une certaine diversité des sites au niveau municipal, mais pour le reste l'ensemble des informations sont principalement présentes sur le site `gov.uk`. Le paysage documentaire est donc plutôt centralisé.

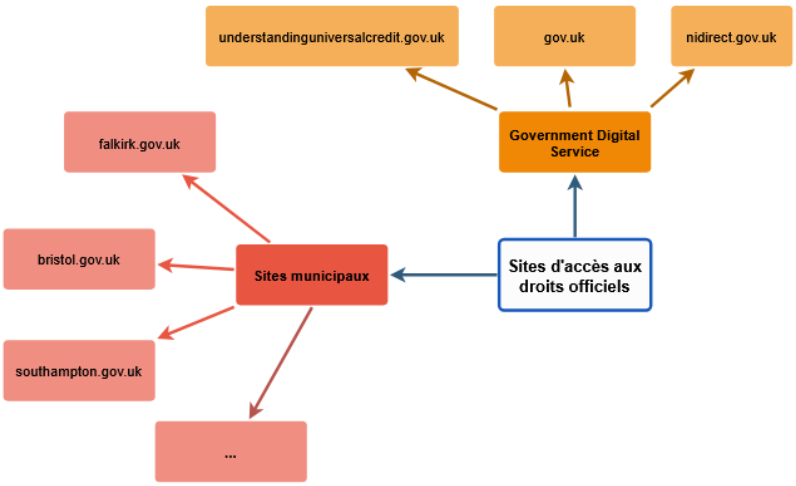


FIGURE 1.4 – Schéma récapitulatif des sites officiels d'accès aux droits au Royaume-Uni

Sur le domaine `gov.uk`, il existe un grand nombre de sites consacrés à chaque département, mais qui ne sont pas consacrés à l'accès aux droits. Ils peuvent cependant être spécialisés dans un domaine, le site `ofgem.gov.uk` concerne par exemple l'énergie pour les entreprises et pour les particuliers. L'usager peut donc y trouver certaines informations spécialisées, mais qui ne relèvent pas directement du discours de l'accès aux droits.

Au Royaume-Uni, un grand nombre de services publics sont fournis par des associations qui jouent donc un rôle essentiel dans l'accès aux droits. C'est le cas notamment de l'association *Citizen Advice Bureau* dont nous avons parlé dans la section 1.1.3. Cette association, qui bénéficie de financements publics, propose des services de conseils. Elle édite également le site web suivant : `citizensadvice.org.uk`. Cette association s'apparente à un service public. Il s'agit d'un site très complet qui propose des informations sur tous les thèmes de l'accès aux droits. D'autres associations plus spécialisées proposent également des informations sur leur site comme Age UK, Moneyhelper, Ginger Bread, etc. Enfin, il est possible de faire appel à une entreprise afin d'obtenir un visa. Les entreprises en question disposent d'un site web présentant parfois des informations, mais comme dans le contexte français, ces informations concernent plutôt les droits des pays étrangers dans lesquels l'usager souhaite se rendre.

Tout comme en France, il existe une communauté d'usager sur les réseaux sociaux. Ces communautés utilisent principalement Reddit et Facebook. Les groupes Facebook

sont privés et spécialisés, concernant un thème unique, comme le dispositif *Universal Credit* par exemple. Sur Reddit, il existe des fils de discussions spécialisés et organisés, comme le fil *r/ukvisa* qui concerne l'obtention de visas, ou encore *r/DWPhelp* qui concerne l'ensemble des aides proposé par le département *Work and Pension*. Ces fils de discussions sont autogérés et disposent de règles et de modérateurs. Enfin, certains organismes officiels proposent des vidéos explicatives, comme le site `nidirect.gov.uk`. Ces vidéos sont également proposés par des particuliers sur des réseaux sociaux comme YouTube.

Un résumé du paysage documentaire comprenant les types de plateformes, les acteurs et leurs objectifs est proposé dans la Table 1.3.

Acteurs	**Organismes gouvernementaux**	**Organismes associatifs**	**Usagers**
Plateforme principale	Site internet public	Site internet associatif	Réseaux Sociaux
Plateformes secondaires	Blog, podcast	Réseaux sociaux	Forum
Objectifs	Diffusion des informations	Aide aux usagers contre rémunération	Compréhension de l'information et entraide

TABLE 1.3 – Récapitulatif des différents lieux de production du discours de l'accès aux droits au Royaume-Uni

1.4 Synthèse

Le contexte du discours numérique de l'accès aux droits est essentiel à sa compréhension et à sa caractérisation. L'exploration des trajectoires prises par chacun des pays met en lumière les particularités historiques, culturelles et sociales qui ont pu influencer l'évolution de ce discours.

La première partie de cette thèse propose une vision de l'ensemble des transformations numériques de l'administration, ainsi qu'une description de la situation actuelle. En effet, les deux pays ont connu à partir du début de 21e siècle des évolutions centrales. En France, un des jalons de la transformation est la mise en place du site `service-public.fr` en 2000. La création de ce site est suivie de la dématérialisation des démarches, en 2004 pour les premières, puis en 2014 pour l'administration des étrangers en France. Enfin, en 2023, il devient obligatoire de faire certaines démarches en ligne comme la naturalisation. Au Royaume-Uni, c'est en 2011 que le gouvernement prend un tournant clé avec la création du *Governement Digital Service* (GDS), puis en

2012 du site gov.uk. La création du GDS permet de mettre en place une réelle stratégie de transformation numérique qui permet à l'accès aux droits numériques d'évoluer au Royaume-Uni.

Ces connaissances sur l'histoire et l'évolution de ces services nous aident à comprendre les enjeux de ce discours. Ils s'avèrent être multiples. Tout d'abord, l'intégration des technologies numériques dans les administrations s'est faite progressivement, et a été sous l'influence des différents facteurs économiques et sociologiques. Ces facteurs sont notamment présentés dans la section 1.2. La transformation numérique s'est faite au fil des évolutions politiques et nationales, en suivant par exemple les logiques du public, sous l'influence du *New Management*, mais aussi dans une logique d'efficacité afin de réduire les coûts pour les administrations. Le numérique est perçu comme plus économique étant donné qu'une partie des tâches (notamment la numérisation des données) sont réalisées par l'usager, que l'administration n'a plus qu'à analyser. Ces dernières années, les gouvernements ont vécu plusieurs périodes d'austérité, qui ont pu voir dans le numérique une réponse aux besoins économiques.

Par conséquent, il n'est pas possible d'étudier le discours de l'accès aux droits sans connaître ces différents facteurs d'influence, comme nous l'avons montré dans la section 1.1. Par ailleurs, les deux pays proposent des visions divergentes de l'intégration du numérique, ce qui permet aussi d'expliquer la présence d'une disparité dans les deux discours, qui présentent tout autant de facteurs communs (1.1.4). Le Royaume-Uni propose une vision centralisée, dotée d'un service central (le GDS) ayant pour mission la transformation numérique et la diffusion des informations. Ces informations sont également concentrées sur un site principal (gov.uk), qui permet de réaliser l'ensemble des démarches. La France propose une vision plurielle, composée d'une multitude de services publics et d'agences. Les sources d'informations sont donc multiples et les politiques de communication peuvent être différentes suivant les sites.

Les changements idéologiques qui ont accompagné cette transition numérique sont un élément d'analyse et de compréhension essentiel, qui est abordé dans la seconde partie de ce chapitre. L'État plateforme, venu des États-Unis, est une idéologie qui consiste à concevoir un État en tant que plateforme numérique qui serait enrichie par les technologies et la collaboration des citoyens. En France, cela prend la forme d'un État dit "start-up", qui encourage les initiatives des usagers et petites entreprises pour améliorer les services de l'état. Nous avons également pu questionner des enjeux culturels à l'œuvre dans cette transition, notamment autour du numérique et de son accessibilité. Les services utilisent des technologies et du matériel qui nécessitent un apprentissage. Il est donc nécessaire que la culture numérique soit intégrée dans notre société. En outre, certains facteurs ont pu avoir une influence indirecte, tandis que d'autres, à l'instar de

la conditionnalité, qui consiste à augmenter les conditions d'accès matérielles (ordinateur, téléphone, etc.) et légales (revenu, imposition, etc.), ont joué un rôle important dans la conception des outils administratifs numériques. L'étude du discours de l'accès aux droits ne peut se faire sans la prise en compte de la charge administrative, qui est le coût de la réalisation des démarches. L'usager doit prouver qu'il est éligible, et la montée en conditionnalité, matérielle avec le numérique, mais aussi légale avec les conditions d'éligibilité peuvent être perçus comme une charge et être contraignant pour les usagers. La compréhension du discours et sa diffusion font partie intégrante de cette potentielle charge. Les changements idéologiques, qu'ils soient liés au numérique ou aux politiques nationales, sont donc des éléments de compréhension du discours.

Enfin, la dernière partie décrit le paysage documentaire de l'accès aux droits actuel en France et au Royaume-Uni, dont le récapitulatif est représenté dans les Figures 1.2 et 1.3. Chaque pays possède un site principal, `gov.uk` pour le Royaume-Uni, et `service-public.fr` pour la France. Ces deux sites sont placés sous la direction du gouvernement, avec le GDS pour le Royaume-Uni et la DILA pour la France. Toutefois, là où le Royaume-Uni choisit de centraliser la diffusion d'information et le développement du numérique dans une seule agence (le GDS), la France développe une multitude de sites et de plateformes, qui proposent chacune leurs démarches et leurs informations : la Caf, Pôle Emploi, le Crous, les préfectures ou encore l'agence Business France et son site `welcometofrance.fr`. Les deux pays proposent également des informations concernant les démarches au niveau local, dans les *councils* pour le Royaume-Uni et dans les villes pour la France. Les deux pays utilisent également les réseaux sociaux ou d'autres canaux (podcast, blogs, etc.) pour la diffusion de certaines informations, mais le site reste le mode de diffusion principal dans les deux pays. C'est pourquoi notre étude du discours de l'accès aux droits se concentre sur le discours des sites web.

Le contexte historique et culturel, ainsi que les enjeux présentés dans ce premier chapitre, permettront de mettre en lumière les caractéristiques du discours de l'accès aux droits numérique. Dans le chapitre suivant, nous nous intéressons aux différentes théories linguistiques connues à ce jour et permettant d'aborder les discours spécialisés et numériques, ainsi que de procéder à leur analyse de manière adaptée. Cette exploration nous permettra de définir notre propre démarche scientifique et d'expliquer les choix de modèles d'analyse.

Chapitre 2

Un discours spécialisé et vulgarisé

"Do the hard work to make it simple"

– Government Digital Service of the United Kingdom

2.1 Discours et genres discursifs

Nous avons présenté dans le chapitre 1 le contexte historique et la situation actuelle du discours de l'accès aux droits. Nous allons maintenant nous intéresser aux concepts de discours, de spécialisation et de vulgarisation. Dans une première partie, nous définirons le cadre théorique utilisé dans cette étude, notamment en nous intéressant à l'analyse du discours. Dans une deuxième partie, nous étudierons le concept de discours spécialisé et de langue de spécialité. Nous présenterons notamment les discours spécialisés en lien avec l'accès aux droits tels que les discours juridique, politique et administratif. Enfin, nous conclurons ce chapitre en abordant les modalités et le rôle de la vulgarisation de ces discours.

2.1.1 Définitions

La première partie de ce chapitre a pour but de définir et situer la discipline de l'analyse du discours, ainsi que ses concepts clés, notamment le discours et le genre de discours.

2.1.1.1 Analyse du discours

L'analyse du discours est une discipline qui rassemble de multiples courants. Il s'agit d'une approche pluridisciplinaire, qui se manifeste différemment selon le domaine d'étude concerné. Paltridge (2012, p. 2) en propose la définition suivante :

> Discourse analysis is an approach to the analysis of language that looks at patterns of language across texts as well as the social and cultural contexts in which the texts occur.

Nous pouvons retrouver dans cette définition deux caractéristiques de la discipline. La première caractéristique est l'étude du langage afin d'y trouver des schémas et des régularités. La seconde caractéristique est la mise en contexte de ces régularités en analysant l'environnement discursif, social et culturel d'un texte. Il existe de multiples courants d'analyse du discours qui chacun se rattachent à des domaines ou à des courants historiques. Certains étudient de grands ensembles de textes écrits, tandis que d'autres étudient les situations d'énonciation (Angermüller, 2007). Pour les analystes du discours, les textes ne sauraient exister sans leur contexte. Il est donc indispensable de le prendre en compte dans l'analyse d'un discours. Paveau (2009) considère que les unités non verbales sont des parties intégrantes du discours et qu'il est nécessaire de les envisager dans un continuum entre verbal et non verbal. C'est la position que nous adopterons dans notre étude. La linguistique de corpus et les diverses approches en terminologie, que nous aborderons dans le chapitre 3, notamment en terminologie textuelle, sont souvent utilisées avec l'analyse du discours. D'après Condamines :

> la terminologie textuelle se rapproche de l'analyse du discours qui considère que toute analyse de textes est un construit et est soumis à une interprétation. (Condamines, 2005, p. 30)

La chercheuse précise que les recherches en corpus ont un objectif interprétatif, en analysant les objectifs des locuteurs. En outre, l'approche par corpus est très répandue, et reste un pivot des recherches en analyse du discours, comme le montre le chapitre « Constituer un corpus en analyse du discours, un moment crucial » extrait de l'ouvrage *Méthodes et outils informatiques pour l'analyse des discours* (Sitri et Barats, 2017). En d'autres termes, l'analyse du discours permet au chercheur d'étudier la place du discours au sein de la société. L'ensemble des pratiques en analyse du discours se croisent et parfois se complètent. Elles ont un point commun central : l'étude du discours.

2.1.1.2 Le discours

Le discours se distingue de la langue. En effet, pour les linguistes, la langue est un système tandis que le discours est une réalisation concrète de la langue, un usage (Maingueneau, 2021b). Notre étude a ainsi pour objectif de caractériser la langue d'accès aux droits à travers son usage concret, donc le discours d'accès aux droits, en prenant appui sur des textes utilisés par la communauté d'experts. Le discours est un concept utilisé dans de nombreuses disciplines, et qui par conséquent possède de multiples définitions.

Dans un ouvrage, Schiffrin et al. (2015) intitulé *The Handbook of Discourse Analysis*, les auteurs retiennent trois notions de la définition du discours. Le discours est :

— le texte au-delà de la phrase,
— l'usage de la langue,
— un ensemble de pratiques sociales plus large.

Pour Paltridge (2012), repris par Maingueneau, un discours est également un énoncé au-delà du mot et de la phrase, « agencé pour que la communication réussisse » (Maingueneau, 2021a, p. 4). Nous retenons pour l'instant deux composantes à notre notion de discours :

— un discours est étudié dans son ensemble,
— un discours est la langue dans son usage.

Charaudeau et Maingueneau, dans l'ouvrage intitulé *Dictionnaire d'Analyse du Discours* publié en 2002, ajoute une caractéristique à cette définition : « l'usage de la langue dans un contexte particulier » (Charaudeau et Maingueneau, 2002, p. 185). Le discours est donc la langue dans son usage, mais aussi dans son contexte. Il est donc nécessaire d'étudier l'ensemble de la communication et non seulement le texte. Bathia (2004) enrichit cette définition du discours en proposant la suivante :

> *the use of written language to communicate meaning in a particular context* (Bathia, 2004, p. 22)

Le discours a donc un but, celui de communiquer un message. Ce message est transmis dans un contexte particulier. En plus de prendre en compte l'usage de la langue, un discours inclut donc le contexte et l'objectif.

Toutefois, d'autres chercheurs ont une vision un peu différente du discours. Burger, par exemple, reprend la notion de contexte en y ajoutant une composante :

> Par 'discours' nous entendons les messages contextualisés, c'est-à-dire dotés d'un sens interprétable par et dans une communauté donnée (Burger, 2018, p. 2)

Cette définition introduit la notion de communauté de discours, dont nous discuterons de façon plus approfondie dans la partie 2.2.2 en lien avec les travaux de Swales (1990). Le discours, même s'il n'a pas de destinataire, possède une dimension interactive. Maingueneau (2007) parle d'interactivité constitutive. Il y a un émetteur et un récepteur, même si l'interactivité peut ne pas être intentionnelle. Dans le cas de notre discours, les émetteurs sont principalement les gouvernements et les organisations chargées de service public et les récepteurs attendus sont les usagers.

Pour certains chercheurs (Benveniste, 1966; Bathia, 2004; Maingueneau, 2021b), le discours est également une forme d'action. Les linguistes parlent alors d'acte de langage,

car le discours vise à avoir un impact sur le monde. Adam (1999) propose une définition plus formelle en écrivant la formule suivante :

> Discours = Texte + contexte/conditions de production et de réception-interprétation (Adam, 1999, p. 39)

Cependant, il est important de noter qu'il réfutera plus tard cette définition. En effet, pour l'auteur, nous n'avons pas accès au contexte comme « donnée extralinguistique objective » (Adam, 2006). L'auteur préfère ainsi la notion de cotexte, qui fait référence à l'ensemble des textes qui entoure la partie du discours étudié. Pour les textes écrits, le cotexte peut être implicite, explicite, voire directement accessible. Dans notre travail de recherche, nous étudierons tant le contexte de production, qui a été présenté dans le chapitre 1 de cette thèse, que le cotexte disponible pour nos textes.

Certains chercheurs proposent une nouvelle formulation de la définition du discours :

> Discours = usage de la langue + type d'unité linguistique Schiffrin (1994, p. 20)

On retrouve dans cette définition les deux composantes principales : la langue en tant qu'en ensemble d'unités linguistiques et son usage situationnel, dans une définition rédigée à la manière d'une formule.

Les différentes définitions proposées dans cette partie mettent en lumière une notion cyclique. Le discours est contextualisé, il s'agit d'un acte de langage qui en même temps définit son contexte, mais qui est aussi défini par son contexte. Il ne s'agit donc pas d'un phénomène unidirectionnel, où un contexte agirait sur le discours ou, à l'inverse, un discours agirait sur son contexte, mais d'un va-et-vient entre ces deux éléments, si bien que le texte est indissociable de son contexte, et de son cotexte. Les deux, le texte d'un côté et le contexte et le cotexte de l'autre, sont coconstruits. Nous verrons que dans le discours numérique en particulier, l'environnement discursif et le discours sont co-conçus (Paveau, 2017). En d'autres termes il n'existe pas de langue sans contexte, ni vraisemblablement de contexte sans langue. C'est précisément à ces aller-retours, qui constituent le discours, que nous souhaitons nous intéresser.

2.1.2 Genre de discours

Pour Beacco et Moirand (1995, p. 1), l'analyse du discours a pour objectif d'étudier les emplois de la langue, et donc de s'intéresser aux différentes catégories du discours, afin de le caractériser. Le genre peut être défini comme un « type de production discursive stabilisée, reliée à une pratique sociale dans un espace historique » (Delavigne, 2022).

L'objectif de notre étude est la caractérisation du discours numérique de l'accès aux droits. Pour cela, nous allons nous intéresser à la notion de genre de discours. L'émer-

gence des genres de discours part d'une observation faite dans le cadre d'une étude des genres littéraires et rhétoriques (Adam, 1997). Adam interroge la notion de genre dans le champ de l'analyse du discours et ses diverses définitions. Il met en avant les deux caractéristiques suivantes :

— les genres sont des régularités sociohistoriques observables dans le champ pratique qui est le leur. (Adam, 1997, p. 670)

— les genres règlent globalement, et de façon plus ou moins contraignante, les différents plans de la structure des textes. (Adam, 1997, p. 671)

Ces deux caractéristiques sont indispensables à l'analyse des genres de discours, ou genres discursifs, et seront intégrées à notre analyse. La notion de régularité est particulièrement documentée par un grand nombre de chercheurs qui s'intéressent à cet objet d'étude. Bakhtine, dans sa définition, parle de stabilité des énoncés :

> Chaque sphère d'utilisation de la langue élabore ses types relativement stables d'énoncés, et c'est ce que nous appelons les genres de discours. (Bakhtine, 1984, p. 265)

Cette stabilité se rapporte à la régularité évoquée par Adam. Il est important que les chercheurs puissent décrire cette stabilité ou régularité, bien qu'elle soit relative, en lien avec un domaine ou une sphère d'utilisation. Pour Bronckart (1996, p. 138) :

> S'ils sont intuitivement différenciés, les genres ne peuvent jamais faire l'objet d'un classement rationnel stable et défini.

Leur caractère instable rend leur caractérisation difficile. Les discours numériques en particulier sont en constante évolution. En effet, ces énoncés correspondent à des pratiques sociodiscursives qui passent par le texte et présentent des régularités micro-linguistiques (Adam, 2001). Toutefois, l'observation de ces régularités ne suffit pas. Comme l'explique Swales (1990), il ne suffit pas de plusieurs événements communicationnels pour définir un genre. L'ensemble de ces événements doit répondre à des objectifs communs. Un genre est donc défini par :

> various patterns of similarity in terms of structure, style, content and intended audience Swales (1990, p. 58)

L'observation de ces pratiques permet de catégoriser les discours, et donc d'identifier des genres de discours. Dans le cas des discours numériques, les pratiques sont guidées par l'évolution des technologies et de leur utilisation. Les genres numériques sont donc particulièrement instables, et difficiles à définir. Dans son étude des discours, Bakhtine affirme que :

> La richesse et la variété des genres du discours sont infinies, car la variété virtuelle de l'activité humaine est inépuisable et chaque sphère de cette

> activité comporte un répertoire des genres du discours qui va se différenciant et s'amplifiant à mesure que se développe et se complexifie la sphère donnée. (Bakhtine, 1984, p. 265)

La sphère évoquée par Bakhtine se rapporte au domaine étudié qui dans le cas de notre étude est l'accès aux droits. Nous avons présenté l'historique et le paysage documentaire de l'accès aux droits dans le chapitre 1. Les genres font donc référence à des pratiques discursives présentant des régularités dans un domaine donné. Pour Rastier (1989), c'est ce qui rattache le texte au discours. À l'instar de Rastier, de nombreux linguistes considèrent que la langue n'est pas le seul système à l'œuvre dans une production linguistique. D'autres « codifications sociales » sont présentes, ce qui forme les genres de discours. Les genres de discours influencent la structure du texte, mais aussi la phraséologie, et parfois la terminologie. L'étude et la caractérisation des genres de discours sont étroitement liées aux discours spécialisés, dont nous parlerons plus amplement dans la section 2.2.

Dans notre étude des sites web de l'accès aux droits, nous souhaitons nous intéresser à ces régularités dans le discours étudié. Notre étude porte sur un discours numérique, nous parlerons donc de rituels d'interaction (Beacco et Moirand, 1995, p. 4). Le terme rituel d'interaction nous semble adapté à la description des genres discursifs numérique, car il met en valeur les deux types d'interactions à l'œuvre dans le discours étudié : les interactions entre l'administration et l'usager, et celles entre les interlocuteurs et le texte numérique. Bathia (2004) indique que les genres sont des communications conventionnées. En effet, l'application des normes linguistiques et sociales liées à un genre de discours dénote d'une certaine intégration des rituels d'interaction normaux ou socialement acceptés. Dans notre travail de recherche, nous définirons donc un genre de discours comme des régularités linguistiques mises en œuvre pour réaliser (ou satisfaire) un rituel d'interaction.

Nous ajouterons qu'il s'exerce un rapport de pouvoir dans la diffusion des genres de discours, et notamment dans notre domaine. En effet, comme indiqué dans le chapitre 1, il existe un rapport hiérarchique entre l'administration, qui est décisionnaire, et les usagers. Si les genres de discours sont des rituels d'interaction, alors c'est l'administration qui définit ces rituels auxquels les usagers doivent se conformer. Certains de ces rituels sont imposés par la société en général et non spécifiques à l'accès aux droits. Cependant, dans des genres très codés comme le discours numérique de l'accès aux droits, l'usager doit se conformer aux attentes de l'administration qui lui propose, voire impose, pour accéder à l'information dont il a besoin, de se plier à un rituel d'interaction précis (naviguer à travers les pages web à la recherche d'informations, définir à travers les énoncés à structure conditionnelle sa situation avant de pouvoir affiner sa recherche, ne pas pouvoir

s'adresser à un agent humain pour poser des questions, s'armer de patience, etc.) Nous pensons donc que l'étude des genres est essentielle dans la caractérisation du discours. L'étude réalisée dans le cadre de cette thèse permettra de décrire les nouveaux rituels d'interaction introduits par le numérique, de définir les genres établis par ces rituels, et d'en décrire le fonctionnement.

Ce panorama nous permet d'appuyer notre étude sur la notion de genre de discours, ainsi que la notion de genre textuel. Les deux appellations semblent à la fois synonymes et proposer un point de vue différent de cette même notion. Le genre textuel désigne un type, une unité ou une réalisation particulière du discours (par exemple, article de loi, formulaire, rapport, etc.) relativement stable. Le genre de discours semble désigner l'ensemble des genres (écrits et oraux). La différence entre les deux, si elle existe, est donc subtile : la notion de genre de discours offre une vision plus large de l'ensemble des genres (écrits et oraux) et des dynamiques à l'œuvre dans l'évolution de ces genres et leur stabilisation, tandis que la notion de genre textuel semble mettre en avant le résultat stabilisé des choix opérés à un moment donné qui caractérisent un type ou genre de texte particulier.

2.1.3 Textes procéduraux

Une grande partie des textes collectés dans notre corpus ont pour objectif d'informer les usagers, mais aussi de les guider dans leur démarche. Ces textes présentent des caractéristiques et objectifs qui s'apparentent à des textes procéduraux. Il s'agit de textes qui :

> communiquent verbalement des procédures, c'est-à-dire des ensembles organisés d'opérations et/ou d'actions à exécuter en vue de transformer un état initial (i.e., la situation de départ) en un état final souhaité (i.e., le but recherché). (Heurley, 2001, p. 40)

Ces textes sont également définis comme des textes qui disent « de et comment faire » (Adam, 2001), c'est-à-dire qu'ils demandent au lecteur de faire une action et produisent des explications pour la réaliser. Ils ont suscité de nombreuses recherches, comme en témoignent la thèse et les multiples travaux de Coutherut (2014, 2016, 2021), qui s'intéresse à divers aspects de ses textes et notamment aux recettes de cuisine. L'aspect procédural des textes est également abordé dans le domaine de l'administration et de la simplification, par exemple dans les travaux de Vecchiato (2019) portant sur l'italien. Ce sont des textes qui mettent en avant la fonction illocutoire du langage puisqu'ils visent à faire agir le lecteur. L'objectif est d'obtenir une aide ou d'acquérir un statut reconnu par l'administration, pour cela, les sites décrivent une série de démarches à réaliser, afin

que l'administré puisse déposer sa demande et obtenir les aides auxquelles il a le droit. Le caractère procédural ne concerne toutefois pas l'ensemble de corpus, qui contient également une partie informative. Cependant, certains résultats obtenus lors des explorations de corpus peuvent être expliqués par le statut procédural de ces textes, montrant l'importance de la prise en compte des types de textes ou genres de discours dans notre étude.

2.2 Un discours spécialisé

Le domaine de l'accès aux droits numérique est l'ensemble des législations et conditions d'accès aux droits des citoyens mis à disposition sur une plateforme numérique, afin que chacun soit en mesure de faire les démarches nécessaires à l'obtention de ses droits. Le discours de l'accès aux droits est donc un discours qui cherche à transmettre des connaissances spécialisées dans un ou plusieurs domaines. Cette section de notre étude se concentre sur la définition et les spécificités du discours spécialisé. Nous allons également nous intéresser au concept de communauté de discours et d'expertise. Enfin, nous étudierons les connaissances spécialisées engagées dans l'accès aux droits.

2.2.1 Définition de la notion de *discours spécialisé*

Dans cette première section consacrée au discours spécialisé, nous allons définir les concepts centraux qui gravitent autour de la notion de spécialisation. Pour les besoins de notre étude, nous allons tout d'abord définir la langue de spécialité avant de préciser les quelques différences théoriques qu'il existe entre langue de spécialité et discours spécialisé. Ces disciplines ne bénéficient pas de concepts ou de définitions universelles, et les frontières entre les termes et leurs utilisations sont instables. Bien que sur le plan opérationnel de la conduite de notre analyse en corpus, la différenciation de ces deux concepts n'a pas d'incident, elle peut se manifester sur le plan de l'interprétation de nos résultats d'analyse. Nous ne jugeons donc pas nécessaire de différencier les deux concepts à ce stade, mais il nous est utile de le faire en prévision de l'interprétation des données collectées depuis notre corpus. En effet, celles-ci vont nous permettre de faire des constatations sur le fonctionnement de la langue de spécialité à partir des observations faites en discours. Comparativement à la dichotomie langue-discours abordée dans la section 2.1.1.2, un discours spécialisé est donc l'usage de la langue de spécialité. Cependant, Petit (2010) s'interroge sur le rapport entre langue de spécialité et discours spécialisé et formule l'hypothèse que pour les discours spécialisés, les caractéristiques spécialisées viendraient du discours et non de la langue. On s'intéresserait donc à ces différentes caractéristiques, selon différents points de vue en langue de spécialité.

La langue de spécialité est un objet mouvant et vivant, dont la définition a longtemps été mal établie. (L'Homme, 2011). Selon L'Homme (2011), elle ne peut être définie par des critères purement linguistiques, tels que ses propriétés lexicales, morphologiques, syntaxiques et stylistiques. Il convient donc de trouver d'autres critères définitoires, davantage extralinguistiques, tels que son usage par un groupe de spécialistes, à l'intérieur d'un domaine du savoir particulier. L'Homme propose ainsi la définition suivante :

> Sous-ensemble linguistique comprenant l'ensemble des moyens d'expression (lexicaux, morphologiques, syntaxiques et stylistiques), utilisés la plupart du temps par un groupe de spécialistes, à l'intérieur d'un domaine du savoir humain. (L'Homme, 2011, p. 31)

Cependant, la notion de sous-ensemble ou de « sous-langue » ne fait pas l'unanimité chez les chercheurs et chercheuses. D'autres considèrent qu'il ne s'agit pas d'une sous-langue, mais bien d'une utilisation particulière de la langue. Pour Lerat (1995, p. 31), une langue de spécialité est « une langue naturelle considérée en tant que vecteur de connaissances spécialisées ». Un discours spécialisé est donc avant tout une utilisation spécifique de la langue afin de transmettre des connaissances spécialisées. Il répond à un besoin : celui de communiquer de façon concise et sans ambiguïté dans un domaine donné (Rouleau, 1995). Le domaine spécialisé est donc une caractéristique essentielle de la langue de spécialité. À ce titre, il est important de le définir. Petit (2010) en donne la définition suivante :

> Tout secteur de la société constitué autour et en vue de l'exercice d'une activité principale qui, par sa nature, sa finalité et ses modalités particulières ainsi que par les compétences particulières qu'elle met en jeu chez ses acteurs, définit la place reconnaissable de ce secteur au sein de la société et d'un ensemble de ses autres secteurs et détermine sa composition et son organisation spécifiques. (Petit, 2010, p. 20)

Une langue de spécialité peut faire référence à la langue parlée par les professionnels dans leurs métiers, ou par les experts d'un domaine. Toutefois, l'idée qu'une langue de spécialité est parlée uniquement par un groupe d'experts est depuis quelques années discutée. Gledhill et Kübler proposent en 2016 la définition suivante pour l'anglais de spécialité :

> *a specific configuration of the English language that is used by a group of specialists for that group's particular purposes (whatever their language origins, whatever their degree of expertise)* (Gledhill et Kübler, 2016, p. 66).

Dans cette définition, les locuteurs non natifs, mais aussi les personnes à niveaux d'expertise variés sont pris en compte. En effet, la définition de Gledhill et Kübler suggère

que la langue de spécialité peut être utilisée par des experts aussi pour communiquer en dehors de leur communauté d'experts. C'est régulièrement le cas dans le processus de vulgarisation, quand ils s'adressent à des spécialistes d'autres disciplines ou au grand public. C'est la situation que nous retrouvons lorsque les experts de l'administration des services publics s'adressent à des usagers de leurs services et rédigent des textes pour eux. Par ailleurs, il arrive régulièrement qu'un locuteur utilise une langue de spécialité pour communiquer avec un expert, sans qu'il ne soit expert lui-même. Ces locuteurs s'approprient les connaissances expertes, dont ils peuvent finir par acquérir la maîtrise à certains degrés. Cela peut être le cas pour les personnels en formation, ou pour des professionnels de domaines adjacents. C'est particulièrement le cas de la langue de l'administration pour laquelle les usagers présentent un niveau d'expertise variable.

2.2.2 Définition de la notion de *communauté de discours*

Une composante importante de la définition d'un discours spécialisé est l'utilisation de ce discours par une communauté. La notion de communauté de discours, introduite par Swales (1990), vient de l'observation que l'utilisation de la langue dans un groupe est dictée par des normes sociales. Cependant, dans une communauté de discours, d'autres facteurs sont à prendre en compte. Afin de rendre la notion de communauté de discours plus facile à appréhender, Swales (1990, p. 24) décrit six critères définitoires. D'après sa définition, une communauté de discours :

— possède des objectifs publics communs ;
— dispose de mécanismes de communication entre ses membres ;
— utilise ces mécanismes afin de transmettre des informations ;
— fait usage et possède plusieurs genres de discours ;
— utilise un lexique spécifique (la terminologie) ;
— présente un équilibre entre novice et experts.

Ces critères permettent d'envisager de façon plus précise les contours de cette notion. Ils permettent de se rendre compte qu'il s'agit d'une notion complexe et protéiforme. Dans un article rédigé près de 25 ans plus tard, Swales (2016) nuancera cette conception de communauté de discours. Les modalités de communication ayant changé radicalement depuis 1990, il semble naturel que cette notion se transforme également. Les communautés sont devenues moins fermées, car elles peuvent être accessibles par internet, et requièrent moins d'interactions tout en partageant leurs objectifs et en possédant des perspectives communes. Swales ajoute alors deux critères supplémentaires. Une communauté de discours possède ainsi en outre :

— un lien silencieux entre ses membres ;

— des perspectives communes.

(Swales, 2016)

Un individu fait partie d'une communauté de discours à partir du moment où il partage les objectifs de la communauté et où il utilise les moyens et normes linguistiques adaptés à la réalisation de ces objectifs (c'est-à-dire une langue de spécialité). Cependant, une communauté de discours peut prendre diverses formes et il s'agit d'un objet difficile à définir. Dans le cas du discours de l'accès aux droits, nous identifions plusieurs communautés de discours qui communiquent entre elles.

2.2.3 Qui sont les experts de l'accès aux droits

Plusieurs critères de la définition de communauté de discours sous-entendent l'existence d'experts du domaine. Nous interrogeons dans cette partie ce qui constitue un expert dans le domaine de l'accès aux droits.

Tout d'abord, l'expertise repose sur la figure de l'expert, qui possède certaines compétences et savoir-faire. Un expert « a acquis des compétences ou des connaissances spécialisées qui permettent de maîtriser un sujet particulier grâce à son expérience et à sa formation » (Ericsson, 2014, p. 1). L'expertise est donc évaluée au niveau individuel, à partir des compétences d'un individu. Dans un article Lelebina et Sardas (2011) proposent une réflexion et une synthèse des recherches sur l'expertise. À partir des travaux d'autres chercheurs, ils apportent des précisions sur le statut de l'expertise :

> L'expertise est considérée comme une construction sociale où le statut d'expert est défini par rapport aux non-experts et aux clients — demandeurs d'expertise. (Huber, 1999 ; Mieg, 2001)

En effet, l'expertise se manifeste de manière évidente dans l'interaction, dans la mesure où la définition de l'expertise se fait par rapport à une non-expertise ou un manque de connaissances. L'article de Lelebina et Sardas (2011) nous apprend également qu'une partie intégrante du statut d'expert est la reconnaissance par les autres de son expertise. Il s'agit d'un rôle qu'endosse un individu.

Dans le cas de l'accès aux droits, il semble légitime de s'interroger sur les experts de ce domaine. En effet, les acteurs sont multiples et issus de divers milieux et secteurs. Il existe évidemment des experts juridiques : les législateurs sont sans aucun doute experts des lois et politiques mises en place dans les services publics. Cependant, ils sont experts du volet juridique et ne sont pas experts des conditions d'accès à ces droits pour les usagers. Nous considérons donc que leur expertise n'est pas toujours adaptée pour l'accès aux droits. Les avocats et juristes dont le rôle est de défendre les usagers peuvent être considérés comme des experts puisqu'ils possèdent des compétences sur la

législation, mais aussi sur les conditions d'accès et les usagers eux-mêmes. La direction de l'information légale et administrative en France et le Government Digital Service au Royaume-Uni sont également experts étant donné que ce sont eux qui rédigent les textes à destination des usagers. Naturellement, l'ensemble des fonctionnaires ou employés des organisations de service public ont également ce statut. C'est le cas également des écrivains publics et membres d'associations experts, qui font partie du *Citizen Advice Bureau* au Royaume-Uni ou du *Pimms méditation* en France et qui sont des médiateurs experts, connaissant les conditions d'accès aux droits (en théorie et en pratique), mais aussi la législation. La relation hiérarchique entre experts, détenteurs du savoir, et usagers semble se confirmer. Ces acteurs sont reconnus comme ayant les compétences nécessaires par leurs clients ou usagers. Toutefois, il n'existe pas, comme on pourrait le penser initialement, d'expertise unidirectionnelle qui placerait l'administration en tant que seule détentrice de l'expertise. Certains acteurs, comme les militants, possèdent également une expertise. Ils proposent ce que Trombert (2013)) appelle une « contre-expertise technico-légale », qui permet parfois à l'usager de contester les décisions de l'administration. Cette contre-expertise est basée sur une connaissance des droits, mais aussi des pratiques de l'administration et de son fonctionnement. Elle peut provenir de groupes militants composés d'usagers, comme nous l'explique Trombert (2013) dans l'étude portant sur un groupe de chômeurs précaires. Ces groupes peuvent également être composés de professionnels du droit, de la rédaction ou de l'action sociale comme les collectifs GISTI ou CIMADE, qui soutiennent l'accès aux droits des immigrés et engagent parfois des actions judiciaires contre les préfectures (GISTI, 2023).

Nous considérons donc qu'il existe dans ce domaine différents types d'expertises qui se complètent. Nous avons réalisé une classification de ces différents experts sous forme de liste numérotée allant des « plus » experts vers les « moins fg experts.

1. **Les législateurs.**
   Ce sont des professionnels du droit qui sont experts du volet juridique.
2. **Les professionnels rédacteurs des sites d'accès aux droits.**
   Ces experts possèdent généralement une spécialisation ou sont en charge d'une rubrique en particulier. Ils maîtrisent le discours et ses particularités linguistiques et numériques.
3. **L'ensemble des fonctionnaires et des employés des organisations de services publics.**
   Bien qu'ils ne soient pas forcément concernés par le discours de l'accès aux droits dans leur quotidien, ils disposent souvent de connaissance et de savoir-faire adjacent qui leur permettent de le maîtriser.
4. **Les médiateurs : assistant(e) social(e), écrivains publics et personnels de**

lieux d'accès aux droits.
Ces personnes sont expertes des conditions d'accès aux droits et connaissent souvent très bien les discours, les pratiques de l'administration ainsi que les démarches numériques.

5. **Les militants.**
Des personnes aux profils divers, qui possèdent des connaissances sociales et techniques, et une contre-expertise.

6. **Les usagers.**
Leurs interactions avec ce discours leur permettent de développer des compétences dans le domaine, au point où certains deviennent usagers-experts et conseillent des usagers moins avancés.

2.2.4 Les niveaux de compétences de l'usager

L'usager est une figure qui a émergé dans les années 1980 pour remplacer la notion d'administré. Il s'agit d'une reconfiguration des relations avec l'État et les services publics (Weller, 2018). Ces changements n'ont cessé d'être remodelés et actualisés avec la dématérialisation des rapports entre usager et administration. Plus récemment, l'idéologie de l'État plateforme et le système start up utilisés dans les administrations françaises et britanniques, ont renforcé la figure de l'usager. Désormais, l'usager doit être force d'innovations et d'avancées (Colin et Verdier, 2015).

Parmi les experts de l'accès aux droits présentés dans la section 2.2.3, nous avons identifié l'usager comme étant un des multiples experts du domaine. Contrairement aux professionnels du domaine, il posséderait un savoir d'usage (Sintomer, 2008), une expertise qui est donc différente de celle des professionnels, et qui ne peut qu'être considérée comme acquise à divers degrés selon les usagers. Trombert décrit un type d'usager possédant une expertise en rédigeant la définition suivante :

> (...) usagers capables de maîtriser les outils et les règles des institutions pour produire une contre-expertise qui parfois l'emporte sur l'expertise institutionnelle initiale. (Trombert, 2013, p. 4)

L'usager peut donc être considéré comme un expert dans certaines situations. Dans le domaine de la santé, le statut du patient expert est depuis plusieurs années reconnu dans la recherche (Boudier et al., 2012; Elberse et al., 2011). Il s'agit généralement de personnes atteintes de maladies chroniques, ou de certaines maladies mentales [1], qui vont

1. Cela semble être notamment le cas avec certaines maladies mentales qui sont effectivement des maladies chroniques et obligent le patient à apprendre sur sa maladie tout au long de la vie pour savoir vivre avec elle.

mettre à profit leur expérience concernant la maladie pour participer à la recherche ou aider d'autres patients. Delavigne (2022) souligne ainsi, dans le cadre d'une étude sur les forums de discussions portant sur le domaine médical, que ce type de personnes font émerger « de nouvelles formes d'expertise ». Ce savoir est transmis dans un cadre parfois informel, et parfois formel. En effet, ces patients experts font parfois partie des cursus de certaines universités (Jouet et al., 2010). On retrouve également un projet du Ministère de la Santé intitulé « L'usager expert : Acteur de l'amélioration des pratiques professionnelles » de 2015 (Ministère de la Santé et des solidarité, 2015). Ce projet avait pour but d'intégrer les patients-experts au parcours de soin afin d'améliorer la prise en charge des patients, et la relation soignant/soigné. L'expertise de ces usagers est donc reconnue par les professionnels du domaine et les institutions, la caractéristique principale de cette expérience étant son mode d'acquisition qui se fait par l'expérience.

Le même phénomène peut être observé pour les usagers des services publics non médicaux. Il s'agit d'usagers utilisant fréquemment les divers services publics, souvent dépendants des aides sociales. Tout comme les malades contraints de se rendre régulièrement à l'hôpital et/ou divers lieux de soins, leurs interactions fréquentes avec les services de l'État les forcent à fréquenter les diverses organisations d'accès aux droits. Il s'agit par exemple de collectifs de travailleurs précaires dépendant du RSA et de Pôle Emploi et devant rendre compte de leur recherche d'emploi aux agents référents, comme le souligne Trombert (2013).

L'expertise acquise par ces usagers concerne généralement les différentes étapes de démarches à suivre, ainsi que les pratiques de l'administration. En effet, pour certains l'accès aux droits est sans embûche, mais pour beaucoup il s'agit d'un parcours complexe, notamment quand ils se retrouvent pour la première fois dans cette situation, qui leur permet d'acquérir de nombreuses connaissances. Par ailleurs, certaines aides demandent un renouvellement régulier. C'est le cas de la Complémentaire Santé Solidaire, la demande de logement social, les allocations logement de la Caf, la prime d'activité et bien d'autres. Pour d'autres droits sociaux, il est obligatoire de remplir un formulaire tous les mois, ou tous les trois mois, de déclarer ses revenus, ses changements d'activités, ou encore de téléverser son bulletin de salaire. Nous pouvons constater que l'accès aux droits induit des obligations, tant au niveau des critères d'éligibilité qu'au niveau des démarches à accomplir. Lorsqu'une erreur s'immisce dans les dossiers, les démarches peuvent devenir encore plus chronophages. Pour certaines organisations, l'obtention d'un rendez-vous en vue de régler ces problèmes est très compliquée : l'usager doit se connecter à une certaine heure sur un site pour obtenir un rendez-vous en présentiel, ou patienter de longues minutes au téléphone.

Prenons l'exemple d'un usager souhaitant demander un logement social à Paris n'ayant

pas accès à un ordinateur : la personne devra prendre un rendez-vous afin d'obtenir les informations, puis rassembler les papiers nécessaires à la demande (parfois une vingtaine de pièces différentes), avant de reprendre rendez-vous pour déposer le dossier.

Au cours des prochaines années, elle devra actualiser son dossier tous les ans, mais aussi chaque fois que sa situation changera (changement d'adresse, d'emploi, naissance d'un enfant, mariage, divorce, etc.). De plus, si sa situation est urgente, elle ira probablement voir un agent de médiation, un écrivain public par exemple, afin d'écrire une lettre au maire de son arrondissement, ou à tout autre représentant jugé compétent. Elle pourra, si sa situation le lui permet, engager un dossier auprès du comité de médiation du Droit au logement opposable (DALO). Suivant sa situation, elle devra attendre entre 1 et 10 ans avant d'avoir le droit de déposer ce dossier (parfois composé d'une vingtaine de pièces également), qui permettra d'étiqueter sa situation comme "urgente" auprès de la ville de Paris. Enfin, il est important de noter que si la personne ne parvient pas à faire le renouvellement de sa demande chaque année, celle-ci est annulée et il est nécessaire de reprendre les démarches au début. Les personnes demandeuses de logement social sont souvent bénéficiaires de minima sociaux (RSA, CMU,...), ce qui multiplie le temps passé à faire des démarches.

Après plusieurs années d'errance administrative, ces personnes deviennent souvent expertes de leur situation, mais aussi de l'accès aux droits en général. Certaines redirigent leurs paires, prennent en charge leurs dossiers, créent des associations d'aide et participent activement sur des forums ou des groupes d'entraide sur les réseaux sociaux. Certains groupes mettent en place des systèmes collaboratifs afin de formaliser les pratiques de l'administration. Ils créent par exemple des tableaux Excel afin d'identifier les délais de traitement des dossiers, comme le subbredit[2] r/ukvisa disponible sur Reddit. Ce groupe rassemble des personnes autour du thème des demandes de visas au Royaume-Uni en leur permettant de discuter des diverses étapes d'une demande et de s'entraider[3].

Cette communauté qui s'est formée autour des demandes de visa a mis en place un document partagé afin de calculer le temps d'attente moyen pour obtenir un *Spouse/Fiancé Visa*, qui permet aux conjoints à l'étranger de rejoindre leur conjoint au Royaume-Uni. Ce document participatif est disponible sous forme de tableau Excel partagé via un Google drive. Grâce aux données participatives récoltées dans ce tableau, les organisateurs ont pu déterminer le délai d'attente moyen pour l'obtention de ce type de visa. Ces statistiques rassemblent un nombre important d'informations concernant le dépôt du visa telles que le statut du visa (standard ou prioritaire), le pays d'origine de la demande,

2. Un subbredit est une rubrique du site consacré à un thème en particulier, où les utilisateurs peuvent ouvrir des discussions.

3. La description indiquée par les modérateurs est la suivante : *« A subreddit to provide advice to people applying for Visas to come and stay or live in the UK. »*.

la date de dépôt initial et des différentes étapes (entretiens, dépôt des empreintes, etc.). Toutefois, ce document rassemble également des informations annexes, plus avancées concernant les sujets suivants :

— Si les demandeurs ont contacté un parlementaire et le délai de réponse de ce dernier : il est d'usage de contacter un parlementaire afin d'appuyer son dossier et d'accélérer la procédure.
— Si les demandeurs avaient contacté le service d'aide UKVI. Le cas échéant, combien cela leur avait coûté . Chaque demande faite au service UKVI coûte 2,74 £ en ligne, et chaque appel coûte 0,69 £ par minute.
— Quel délai leur avait été donné par l'autorité en charge de délivrer le visa.
— Depuis quand il n'avait pas vu leur partenaire.
— S'ils avaient reçu un email « ambigü »[4].

Ces sujets ayant été choisis, on peut deviner qu'ils doivent jouer un rôle important dans la demande de visa. La Figure 2.1 est une synthèse de ces informations récoltées sur l'année 2022.

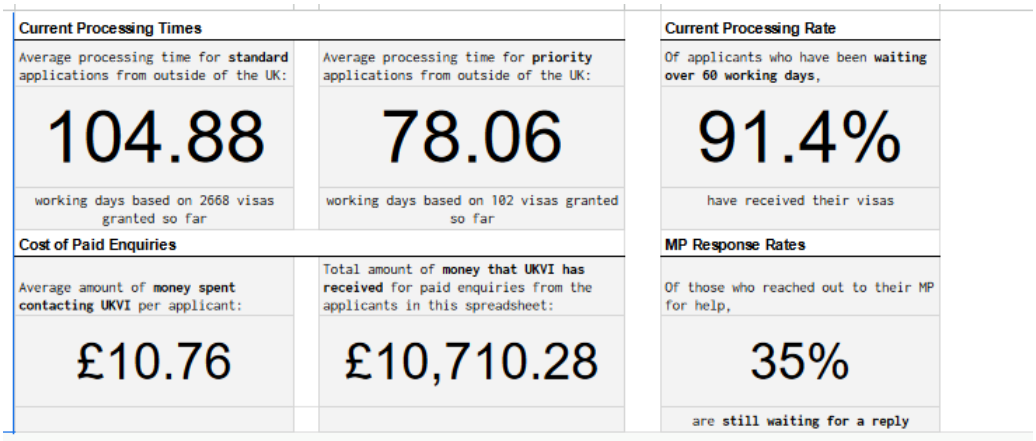


FIGURE 2.1 – Récapitulatif proposé par les modérateurs du groupe r/ukvisa diffusée sur Reddit

Cette synthèse, ainsi que le document partagé sur lequel elle est basée, montre une connaissance avancée du processus de demande de visa, mais également des éléments problématiques qui peuvent apparaître lors du processus. Elle illustre l'organisation et le type d'expertise que les usagers peuvent proposer. La synthèse propose plusieurs informations qui ne sont pas disponibles sur d'autres canaux d'informations, officiels ou non. Ces statistiques ne représentent qu'une partie de la population qui a accepté de remplir les informations demandées, elles ne constituent pas de données complètes, mais plutôt de données informelles, utilisées par les utilisateurs pour comprendre leur situation.

4. Texte original : « *Did you receive a 'not straight-forward' email ?* »

Toutefois, il est intéressant de noter que ces statistiques sont réalisées sous un format diffusable, ou du moins lisible facilement, comme en témoigne le tableau de la Figure 2.1. On peut donc supposer une volonté de communication et de visibilité de la part des organisateurs du groupe Reddit.

Le même phénomène s'observe pour les ressortissants étrangers en France souhaitant obtenir une carte de séjour ou la naturalisation. La naturalisation étant un processus complexe et opaque, de nombreux groupes voient le jour. On peut y observer une expertise des usagers qui au sein de ces groupes proposent des conseils personnalisés, suivant le parcours et selon les questions des participants. Certains réécrivent même le résumé de certaines procédures de façon simplifiée ou plus adaptée. Un membre d'un groupe de retour d'expérience concernant la naturalisation sur le réseau social Facebook reprend par exemple l'ensemble des étapes de la procédure de naturalisation par décret dans un post :

> Le processus de demande de naturalisation en France peut varier en fonction de la situation individuelle de chaque demandeur, mais voici généralement les étapes impliquées dans le traitement d'un dossier de naturalisation : (...) (anonyme, 2023)

Dans l'introduction de ce post, on remarque que l'usager reformule les informations inscrites sur les sites d'accès aux droits. Cette reformulation, qui ne présente pas d'erreur, témoigne d'une compréhension et d'une intégration de cette démarche.

Les experts professionnels et formés sont certes de précieuses aides puisqu'ils possèdent des codes et des connaissances indispensables. Cependant, les expériences des usagers sont tout aussi importantes, car elles témoignent des pratiques dans ce domaine. De plus, d'après la définition d'expertise et d'experts introduite dans la section précédente, ce type d'usager correspond bien aux critères énoncés. Tout d'abord, ces usagers possèdent « des compétences ou des connaissances spécialisées » (Ericsson, 2014, p. 1). Ils répondent aux demandes d'autres individus « demandeurs d'expertises » (Huber, 1999 ; Mieg, 2001), qui posent des questions sur les forums et groupes de paroles. Enfin, ils font l'objet d'une reconnaissance à ce titre, ce qui, comme nous l'avons montré dans la section précédente 2.2.3, est essentiel dans la définition d'un expert (Lelebina et Sardas, 2011).

Dans la description de ces phénomènes de groupes d'entraide en lien avec les usagers experts, nous observons qu'il existe différents niveaux de compétences des usagers. Certains usagers sont novices, et sollicitent l'aide des services professionnels, ou encore de leurs paires. D'autres usagers, au contraire, possèdent une certaine expertise qu'ils mettent à profit afin d'aider leurs paires. Enfin, les phénomènes de groupes d'entraide nous permettent d'identifier également une expertise de groupe. L'ensemble des

expériences des individus, lorsqu'elles sont mises en commun, permettent d'accéder à certaines connaissances et d'acquérir une expertise. Ces dernières peuvent s'appuyer sur des statistiques, qu'on pourrait qualifier d'« officieuses ». Ces informations et connaissances ne sont pas communiquées par les administrations, et ne sont parfois pas systématiquement quantifiées officiellement, mais elles sont mises en lumière par les usagers eux-mêmes.

Notre étude a permis de questionner le statut de l'usager de l'accès aux droits. Il ne s'agit pas toujours d'un individu novice dépendant de l'administration. Au contraire, l'usager peut devenir, avec le temps, une figure complexe, au profil et au statut divers et qui possède différents niveaux de compétences. Ces différents niveaux de compétences, lorsqu'ils sont rapportés à la spécialisation du discours, posent la question du contour ou périmètre de sa communauté de discours.

2.2.5 La communauté de discours de l'accès aux droits

Notre étude de ce discours spécialisé a mis en avant les multiples acteurs qui constituent le domaine de l'accès aux droits. Nous avons établi qu'il s'agit d'un domaine spécialisé composé de plusieurs genres de discours, dont les sites web d'informations et d'accès aux droits, qui sont notre objet d'étude. Dans le cas du discours de l'accès aux droits, la communauté de discours englobe les personnes qui côtoient ces sites et ces discours régulièrement. Nous estimons qu'à partir du moment où un usager consulte les informations et communique à ce sujet — notamment à travers les sites — en remplissant les formulaires, en envoyant des mails ou dans toutes autres circonstances, il fait partie de la communauté de discours. Les sites internet gouvernementaux de diffusion de l'information administrative sont le point de départ de la diffusion du discours spécialisé au grand public. Les articles sont rédigés par des experts formés et destinés à une diffusion large à la population. L'objectif public de ce discours est d'informer les citoyens de leurs droits afin qu'ils puissent y accéder. Cet objectif est des plus complexes, puisqu'il ne s'agit pas toujours d'une diffusion d'experts, détenteurs de l'information, vers un public dit profane. Comme nous l'avons montré, les professionnels ne détiennent pas toujours toutes les connaissances, et les usagers peuvent apporter une expertise complémentaire. De plus, notre étude ne porte que sur les canaux officiels de diffusion des informations (sites web). Il nous a ainsi paru important de noter qu'il existe un nombre important de canaux non officiels où professionnels et usagers se confondent. Nous avons présenté l'ensemble de ces canaux dans la section 1.3 du chapitre 1, mais également dans la section précédente (section 2.2.4) consacrée aux compétences des usagers. Le domaine étudié est particulier, car il regroupe un grand nombre de sous-domaines et de domaines spécialisés. Il existe donc plusieurs sous-communautés de discours suivant les secteurs

et le type d'aide. Il existe également une communauté de discours autour des outils numériques de l'administration.

La communauté de discours de l'accès aux droits est donc très hétérogène. Elles rassemblent différents acteurs possédant différents niveaux de connaissances. Il est possible d'identifier deux objectifs complémentaires qui déterminent l'appartenance à ces communautés de discours. La première catégorie d'individus faisant partie de cette communauté a pour objectif de faciliter l'accès aux droits : par leur reformulation, la publication des informations officielles, les explications, etc. La seconde catégorie a pour objectif d'accéder à ces droits et maîtrise ce discours afin d'y arriver. Il existe bien sûr un mouvement entre ces deux catégories, et un usager voulant initialement accéder à ses droits pourra ensuite se trouver dans la première catégorie en aidant à la diffusion du discours. Ces deux catégories sont complémentaires car l'accès aux informations et la compréhension des textes sous-entendent des compétences spécifiques que nous allons présenter dans la partie suivante.

2.2.6 *Administrative literacy*

La communication avec les administrations et la réalisation de tâches administratives relèvent de compétences spécifiques. En 1975, Gordon introduit la notion de *Bureaucratic Competence*, qui fait référence aux compétences nécessaires pour réaliser des « interactions administratives ». Certaines compétences non acquises par les usagers pourraient rendre la communication avec les administrations plus difficile. Cela pourrait, à terme, entraîner des inégalités d'accès aux informations et aux démarches. Les compétences concernées sont « le vocabulaire, la familiarité avec les formulaires et les documents, la prise de conscience de l'importance de la ténacité et la compréhension que la répétition et le traitement impersonnel d'informations extrêmement personnelles sont souvent caractéristiques. » (Gordon, 1975, p. 198). Une partie des compétences décrites par l'auteur relève de la connaissance du discours utilisé et du domaine en général. L'usager doit donc être familier avec les documents administratifs. Cette particularité des compétences bureaucratiques confirme qu'il s'agit d'un discours spécialisé. En effet, pour interagir avec succès avec les professionnels, les usagers doivent intégrer des compétences linguistiques de cette communauté de discours.

Dans un second temps, une nouvelle notion émerge d'abord en anglais, *Administrative literacy* Administrative literacy puis, sous influence de l'anglais, elle fait son apparition en français, sous le nom de littératie administrative. Elle fait référence aux compétences nécessaires pour réaliser des tâches administratives, notamment trouver des formulaires et effectuer des démarches (Cestnik et Kern, 2014). Döring (2018, p. 1) en donne la définition suivante :

> *Administrative literacy is the capacity to obtain, process, and understand basic information and services from public organizations needed to make appropriate decisions.*

Le concept d'*administrative literacy* peut être vu comme une formalisation des compétences requises afin d'intégrer la communauté de discours administratif. En effet, dans les compétences demandées, on retrouve des compétences linguistiques telles que la connaissance de la terminologie, mais aussi des genres textuels utilisés, notamment les formulaires et les documents officiels demandés dans le cadre de la réalisation d'une démarche administrative. Les usagers doivent pouvoir communiquer avec les autres membres de cette communauté discursive afin de formuler leur demande. Döring (2018, p. 12) identifie sept types de compétences qui forment la littéracie administrative :

1. une compétence fonctionnelle, qui fait référence à la compréhension de la terminologie et de la phraséologie du domaine ;
2. une compétence communicative, qui permet, grâce à la connaissance de la terminologie et de la phraséologie, d'interagir avec d'autres membres de la communauté de discours afin d'obtenir ou de comprendre des informations ;
3. une compétence structurelle, qui implique une connaissance des structures administratives ;
4. une compétence procédurale, qui implique de connaître les rituels et normes d'interaction afin de s'y soumettre ;
5. une compétence civique, qui implique de replacer les aides et réglementations administratives dans un contexte sociétal. L'objectif de certaines réglementations et obligations n'est pas toujours compréhensible pour l'individu, car elles visent un objectif plus large qui bénéficierait à la société tout entière ;
6. la littéracie médiatique, qui fait référence à la capacité de comprendre, d'interpréter et d'analyser les contenus médiatiques ;
7. la littéracie numérique, qui permet de naviguer sur l'ensemble des outils numériques.

Ces compétences et savoir-faire sont essentiels, d'autant plus que dans le discours étudié, les institutions ont le monopole de la diffusion des informations officielles (Muracciole et Massé, 2018). Ce sont eux qui dictent les conventions de communication, auxquelles l'usager doit se plier. Toute transformation des rituels de communication est instaurée par l'administration : les usagers sont contraints de s'adapter à ces changements. Il y a ainsi une relation de dépendance entre les usagers et les services publics (Döring, 2018). Ces derniers opèrent des changements et prennent des décisions auxquels les usagers doivent se plier afin d'obtenir leurs droits. La transformation des rituels

de communication qui nous intéresse dans cette thèse, et qui semble particulièrement importante a été amenée par la transformation numérique.

Les chercheurs Cestnik et Kern (2014) montrent dans leurs recherches que depuis la transformation numérique, littéracie administrative et littéracie numérique sont fortement liées. Malgré ce lien, une bonne connaissance des outils numériques n'est pas suffisante dans l'utilisation des services numériques du gouvernement. Il est nécessaire d'avoir les deux cocompétences afin de pouvoir réaliser les opérations administratives nécessaires à l'accès aux droits.

L'introduction de la notion d'*Adminitrative literacy* montre qu'il est nécessaire d'avoir des connaissances diverses et notamment de connaître la terminologie et la phraséologie du discours afin de le comprendre, d'interagir avec les membres de la communauté de discours et d'accéder à ses droits. Nous revenons maintenant vers la notion de langue de spécialité, introduite dans la section 2.2 afin de préciser les contours et les caractéristiques générales des langues de spécialités administrative, juridique et politique.

2.2.7 Langues de spécialité

Le discours de l'accès aux droits concerne différents domaines spécialisés. Pour rappel, une langue de spécialité est « un sous-système linguistique correspondant à l'emploi de la langue dans un domaine de connaissances unique » (Pecman, 2007, p. 5). Le discours étudié est issu de plusieurs langues de spécialité. De ce fait, les particularités linguistiques du discours proviennent de différentes langues de spécialité qui s'entrecroisent afin de former un nouveau type de discours. Dans cette section, nous allons présenter les différentes langues de spécialité qui composent et influencent ce discours.

2.2.7.1 Juridique

Le discours de l'accès aux droits est inévitablement lié au discours juridique. En effet, une grande partie des contenus informatifs publiés sur les sites sont issus des lois, décrets, arrêts et autres législations. Nous allons donc présenter les caractéristiques du discours juridique français et britannique. Les chercheurs s'intéressent à ce discours très tôt en se questionnant sur ses particularités par rapport à la langue générale, mais aussi par rapport aux autres langues de spécialité. Les termes et la syntaxe de la langue juridique sont très spécifiques et le rôle de la terminologie fait l'objet de discussion (Gémar, 1990). En France, elle est décrite comme possédant les caractéristiques suivantes (Roy, 2013) :

— abondance de mots,
— phrases longues (parfois excessivement longues),

— tournures de phrases obscures et compliquées,
— manque de mise en contexte.

Dans un article présentant le discours juridique, Tiersma (2008) fait des observations similaires sur le discours juridique du Royaume-Uni. Il identifie également des phrases longues et complexes, ainsi qu'un lexique abondant comme étant caractéristiques. Il ajoute qu'une caractéristique supplémentaire de ce discours est la présence de nombreux synonymes et quasi-synonymes, notamment en anglais juridique. Cette caractéristique se manifeste par la présence de doublets juridiques qui sont des tautologies comme *aid and abet* ou *null and void* (Richard, 2015; Humbley, 2016).

2.2.7.2 Politique

Le discours de l'accès aux droits est également issu d'un discours qui serait politique. Il existe bien sûr une multitude de définitions du discours politique qui sont plus ou moins restrictives. La définition suivante semble convenir pour notre étude.

> *[political discourse is] concerned with formal/informal political contexts and political actors ((Graber, 1981)); with, that is, inter alia, politicians, political institutions, governments, political media, and political supporters operating in political environments to achieve political goals.* (Wilson, 2005, p. 1)

Effectivement, l'ensemble des législations notamment d'aide sociale et d'immigration sont prises dans des enjeux politiques, puisqu'elles peuvent être écrites et mises en place par des acteurs politiques. Ces acteurs peuvent ainsi influencer des décisions linguistiques. C'est le cas au Royaume-Uni, avec la création du dispositif *Universal Credit*, qui rassemble plusieurs aides en un dispositif unique. Son introduction, ainsi que le choix de cette dénomination, répond à des enjeux politiques (Morris, 2016; Bozio et Sultan, 2021). À ce titre, il peut s'agir d'un discours politique qui intervient dans le discours de l'accès aux droits. Il conviendra donc dans notre étude de mettre en lumière les résultats obtenus avec le contexte politique de production.

2.2.7.3 Administrative

La langue de l'administration peut être définie comme la langue produite par :

> les services de l'État, les services publics, les institutions régaliennes (justice, police), les finances publiques, les collectivités territoriales, et, plus globalement, les organisations qui assurent une mission de service public. (Krieg-Planque, 2020)

Les acteurs sont mis en avant dans cette définition aux côtés de la notion de service public. Farmer (1995, p. 2) propose une autre définition de la langue de l'administration :

> *the way in which public administration information is arranged is the language of public administration.*

Dans cette définition ce n'est pas les acteurs qui sont mis en avant, mais les systèmes d'information. La langue de l'administration est donc produite par des acteurs multiples, pour transmettre des informations, dans le but d'assurer des services publics. La langue de l'administration rassemble un grand nombre de genres différents. Communication interne, textes de références, textes d'informations destinées aux usagers, formulaires, textes intégrés à des dispositifs techniques tels que les simulateurs, sont autant de genres de textes diffusés par les administrations qui relèvent de cette langue de spécialité. Les sujets abordés par l'administration sont également très divers (transport, impôts, énergie, etc.). Parmi l'ensemble de ces genres administratifs, le discours numérique de l'accès aux droits est un genre vulgarisé du discours administratif.

2.2.7.4 Institutionnelle

Il existe plusieurs définitions du discours institutionnel qui sont plus ou moins précises. De manière générale, il s'agit des « discours « autorisés » dans un milieu donné » (Oger et Ollivier-Yaniv, 2003 : 7). Toutefois, dans le cadre de notre étude, nous nous intéressons également à une définition plus restreinte qui est la suivante :

> Nous entendons par « discours institutionnel », au sens strict, le discours produit officiellement par un énonciateur singulier ou collectif qui occupe une position juridiquement inscrite dans l'appareil d'État, qu'il soit fonctionnaire ou représentant politique. (Oger et Ollivier-Yaniv, 2003 : 7)

En ce sens, le discours de l'accès aux droits s'apparente à un discours institutionnel. Comme l'indique (Maingueneau, 2021a), les administrations bénéficient d'un « monopole énonciatif ». Krieg-Planque et Oger (2010, p. 1) indiquent que l'analyse du discours institutionnel :

> porte un intérêt spécifique à la question de la légitimation des institutions, en insistant sur le système de contraintes qui s'exerce sur l'énonciation.

Nous avons dans le chapitre 1 exploré l'ensemble des contraintes et idéologies qui sous-tendent l'accès aux droits. Nous pourrons donc éclairer nos résultats en étudiant le fonctionnement de ces systèmes de contraintes. Ces définitions indiquent que le discours de l'accès aux droits fait partie des discours institutionnels. Il existe bien sûr plusieurs genres de discours institutionnels qui déploient chacun des stratégies linguistiques ainsi qu'une phraséologie et une terminologie qui leur est propre.

L'influence des langues de spécialités juridique, politique, administrative et institutionnelle est intégrée à l'analyse du discours de l'accès aux droits dans les chapitres 5 et 6 consacrés aux résultats de notre analyse.

2.3 Un discours vulgarisé

Nous venons de présenter les caractéristiques de spécialisation du discours de l'accès aux droits. Toutefois, il s'agit également d'un discours vulgarisé qui est produit par des professionnels à l'intention de personnes profanes possédant un niveau d'expertise varié. Dans cette partie, nous nous intéresserons à la vulgarisation et aux divers outils utilisés pour rédiger un texte vulgarisé tels que le langage simple et clair et les outils rédactionnels.

2.3.1 Qu'est-ce que la vulgarisation ?

La vulgarisation est « une activité de diffusion vers l'extérieur de connaissances déjà produite à l'intérieur d'une communauté » (Authier, 1982, p. 34). Sa définition peut être parfois épineuse, car il existe plusieurs degrés de vulgarisation d'un texte, et il est alors difficile d'identifier un texte vulgarisé d'un texte « purement » spécialisé. Afin d'identifier plus clairement un texte vulgarisé, des chercheurs l'ont qualifié de discours second (Authier, 1982). De nombreuses définitions présentent ce paramètre comme essentiel. Le discours de vulgarisation :

> se présente comme un discours intermédiaire sur l'éventail des discours de transmission des connaissances [et il] est par essence un lieu hétérogène considéré comme un discours second brassant des discours sources. (Reboul-Touré, 2004, p. 197)

Cette définition met en avant les notions de discours source, et de discours second, qui sont centrales dans la définition d'un texte vulgarisé.

La vulgarisation est essentielle dans la communication entre experts et non-experts. Quand elle a pour objectif de former des experts, il s'agit alors d'un type de discours particulier : le discours pédagogique ou didactique. Dans d'autres cas, elle permet simplement aux non-experts d'acquérir les connaissances nécessaires pour appréhender et comprendre un domaine et une communauté discursive. La vulgarisation est un sujet essentiel dans les domaines scientifiques, puisqu'elle permet au grand public d'accéder à des informations spécialisées. Wolton (1997) indique également que la vulgarisation

permet une émancipation du grand public, qui peut grâce à elle se renseigner indépendamment et accéder à des savoirs et à des informations préalablement inatteignables.

Il existe donc des formes diversifiées de discours vulgarisés : semi-vulgarisation, discours pédagogique, discours médiatique, etc. Toutes les formes de vulgarisation sont caractérisées par un travail d'adaptation linguistique. Ce travail d'adaptation linguistique est souvent un travail de reformulation et d'explicitation. Il est ainsi possible d'identifier certaines « marques » (Reboul-Touré, 2004) d'adaptation du discours. Ce travail est notamment visible autour de termes qui peuvent être opaques pour le locuteur non expert. Certaines études ont montré que les termes présentaient un fonctionnement discursif spécifique dans les discours vulgarisés (Delavigne, 2003). Ils ont donc une place et un comportement particuliers au sein de ce discours. Nous nous attacherons à mettre en lumière ces particularités dans notre travail de recherche, et notamment dans les chapitres 5 et 6, bien qu'il existe bien d'autres particularités linguistiques dans la vulgarisation.

2.3.2 Le discours de l'accès aux droits est-il vulgarisé ?

Le discours étudié est issu d'un acte de vulgarisation, c'est-à-dire d'une activité de diffusion des informations vers un public non professionnel. Il peut s'agir d'un discours second, qui reformule et adapte des textes principalement juridiques et administratifs. Comme nous l'avons défini dans la section 1.1, l'accès aux droits indique que les usagers bénéficient d'informations concernant leurs droits, afin d'être en mesure de les faire valoir. Il s'agit donc d'une activité de transmission d'un savoir ou d'une information. Dans le discours étudié, la communauté de non-experts est celle des usagers. Les usagers ne sont pas une communauté dotée d'un savoir homogène. Chacun dispose d'une connaissance plus ou moins approfondie des administrations, comme nous l'avons montré dans la section 2.2.4. Il est cependant possible d'identifier un groupe de professionnels qui diffuse une connaissance interne vers un public dit « externe ». L'information devant être accessible au plus grand nombre, il est important de vulgariser l'information comme si la personne ne disposait d'aucune connaissance préalable. La vulgarisation des informations administratives est un sujet central dans les rapports entre l'administration et les usagers. D'après certaines recherches (Muracciole et Massé, 2018), les administrations détiennent le monopole de la diffusion de l'information. En effet, ce sont les interlocuteurs principaux pour les usagers, mais aussi ceux qui prennent les décisions concernant leurs droits. De plus, ce sont eux qui sont responsables de la mise à disposition des informations pour la réalisation de ces démarches. Il est tout de même important de noter que les informations sont disponibles dans les textes juridiques sur le site `legifrance.gouv.fr`. Cependant, ces informations relèvent du discours spécialisé juridique et administratif, pour lequel les usagers ne sont pas formés. Il relève

donc de la responsabilité de l'État de les rendre accessibles à tous en les vulgarisant. La vulgarisation n'est pas le seul moyen de rendre les informations accessibles. Il est bien sûr important que les usagers saisissent les informations administratives dans leur complexité, mais la simplification des démarches administratives et le langage clair peuvent aider à garantir la compréhension des informations et, dans le même temps, l'accès aux droits. Il existe plusieurs stratégies de vulgarisation. Nous nous concentrons dans les sections suivantes sur les stratégies de simplification et le langage clair.

2.3.3 La simplification et le langage clair

Les outils pour la vulgarisation et la transmission des informations administratives sont divers. En France comme au Royaume-Uni, on retrouve des efforts de vulgarisation et de simplification afin d'améliorer la communication entre l'administration et les administrés. Ces efforts de vulgarisation passent notamment par la simplification des textes, et parfois même des démarches. En France, un grand nombre de dispositifs ont été mis en place pour réaliser cette simplification. Parmi ces dispositifs, nous pouvons citer la Commission pour la simplification des formalités, la Commission pour la simplification administrative (COSLA), la Direction générale de la modernisation de l'État, la Direction interministérielle de la transformation publique (Romain et al., 2022) et la Direction de l'information légale et administrative. Chacune de ces entités a mis en place divers outils : glossaires, fiches, logiciels, etc., qui ont pu apporter leur contribution à la diffusion des informations des administrations. Le domaine intéresse également la recherche, à l'instar des chercheurs de l'université de Louvain qui ont créé la plateforme AMesure pour aider à la rédaction administrative (François et al., 2018). Il est également important de noter que la vulgarisation et la simplification du discours administratif ne sauraient se limiter à des enjeux linguistiques. Le discours administratif est en grande partie garant de la relation entre l'administration et les usagers. En effet, cette simplification répond à deux objectifs : l'exercice des droits humains, et l'efficacité de l'administration (Krieg-Planque, 2020). Le sujet est donc pris d'enjeux sociaux.

La vulgarisation est un facteur important de la diffusion d'informations relatives aux démarches administratives, qui utilise divers procédés afin de garantir l'accès aux informations spécialisées au plus grand nombre. Krieg-Planque (2020) étudie le langage administratif et indique que les stratégies de simplification sont une forme de vulgarisation :

> elles visent à simplifier l'expression, non pas parce que la simplicité serait en soi une qualité, mais parce qu'un type de simplification donné semble plus adapté à un public donné (Krieg-Planque, 2020, p. 25)

Dans cette définition, nous pouvons relever la notion que la simplification sous-entend d'adapter le texte à un certain public. À ce titre, le recours au langage clair et simple est un outil précieux. Le langage clair et simple utilise des caractéristiques textuelles afin de permettre au lecteur de comprendre un texte à la première lecture (Rossetti et al., 2020). Il consiste à mettre en place des règles de rédaction afin de simplifier la transmission de l'information. Il s'agit de porter une attention particulière à certains points, décrits par exemple par Roy (2013, p. 983).

> Ses efforts visent notamment :
> — l'adaptation au lectorat ;
> — la structure générale de l'information (choix de l'information à transmettre, hiérarchisation de l'information, etc.) ;
> — la terminologie (ton employé, choix des mots, explication des termes juridiques, etc.) ;
> — la syntaxe (longueur des phrases, élimination des ambiguïtés, etc.) ;
> — la mise en page du document (mise en relief de l'information, utilisation de graphiques et tableaux, etc.).

Le but est donc de s'assurer que les textes sont compréhensibles par un public donné. Le langage clair répond à des règles qui forment un guide de style. Une simplification du langage ne signifie par une simplification du domaine. En effet, le sujet présenté reste un sujet complexe et les termes spécialisés ne sont pas effacés pour autant. L'usager doit comprendre les informations dans toute leur complexité. Il s'agit toujours d'un discours spécialisé (Adler, 2012). Mais une attention particulière est portée à la clarté du texte, pour que le lecteur ait le moins d'efforts à faire au niveau de sa compréhension et retienne le plus d'informations.

Le style clair en langue de spécialité dans le discours de l'accès aux droits est également un sujet qui intéresse la recherche en linguistique. Le langage simple et clair est notamment étudié dans le domaine médical. Gledhill et al. (2019) étudient les résumés Cochrane en langage clair qui permettent aux experts et aux non-experts de comprendre les résultats de recherche afin d'aider à la prise de décisions médicales. Dans le domaine des finances, Rossetti et al. (2020) s'intéressent au langage clair comme outil de compréhension pour les personnes âgées. Ils évaluent les potentialités de l'utilisation du langage clair et simple pour les textes en ligne. Ces recherches montrent l'intérêt croissant pour ce domaine et la pertinence des recherches entreprises sur ce sujet.

2.3.4 Stratégies de simplification pour le discours de l'accès aux droits

Les gouvernements britanniques et français ont tous deux affiché la volonté de produire des textes les plus clairs possibles à destination des usagers.

2.3.4.1 Royaume-Uni

Au Royaume-Uni, cela se traduit par l'utilisation du *plain language* et par un guide de style très complet mis à disposition des usagers sur le site `gov.uk`[5]. Cela est très clairement indiqué par le slogan choisi pour le site `gov.uk` que l'on retrouve sur la page d'accueil : "*Simpler, clearer, faster*". Un podcast officiel publié par le Government Digital Service en 2019 est même consacré au « *Clear writing* » (Service, Service). Le gouvernement semble également prendre en compte le support utilisé pour transmettre les informations : la page web. En effet, les instructions soulignent les différences entre la lecture sur support papier et la lecture d'une page web et demandent d'adapter le style en conséquence. Ce guide de style et les informations publiées sur les sites sont gérés par le GDS, ce qui montre qu'une attention particulière est portée au numérique. Le site `nidirect.gov.uk` suit le même guide de style et les mêmes recommandations que le site `gov.uk`. Sur les sites de certaines municipalités, le guide mentionne l'utilisation de score de lisibilité Flesch Kincaid (Kincaid et al., 1975), comme le site de la municipalité de Richmond[6], qui en plus d'indiquer l'utilisation du langage simple et clair pour l'ensemble de son site, recommande d'utiliser un outil basé sur le score de lisibilité Flesch Kincaid. Il s'agit d'un score utilisé pour évaluer la lisibilité des informations sur les sites web, notamment dans le domaine de la santé (Brigo et al., 2015; Ellimoottil et al., 2012; Jayaratne et al., 2014). Enfin, il existe sur la plupart des sites d'accès aux droits des rubriques permettant à l'usager de signaler des erreurs ou des possibilités d'amélioration. La simplification des démarches est également une préoccupation importante. Comme nous l'avons évoqué dans la section 1.1.3, en contexte britannique, un des premiers efforts de simplification introduits par le Royaume-Uni est le prélèvement à la source dès 1945. Le Royaume-Uni a ensuite mis en place des efforts de simplification à la fin des années 1970 puis au début des années 2000. La dernière vague de simplification de l'administration a été opérée avec la transformation numérique. Un programme de simplification[7] a été mis en œuvre par le gouvernement entre les années 2005 et 2010. Le rapport final[8] de ce programme de simplification mentionne des recherches notamment concernant l'*administrative burden*, que nous avons défini dans la section 1.2.3. Plus récemment, la simplification des démarches s'est caractérisée par la création d'un dispositif d'aide unique, rassemblant plusieurs dispositifs : *Universal Credit*. Le

5. disponible à l'adresse suivante : `https://www.gov.uk/guidance/style-guide/a-to-z-of-gov-uk-style`, consulté le 20/10/2023

6. `https://www.richmond.gov.uk/style_guide`, consulté le 27/10/2023

7. « *Simplification plans* »

8. disponible à l'adresse suivante : `gov.uk/government/publications/simplifying-regulation-and-administration-of-businesses-final-report` consulté le 20/10/2023

Royaume-Uni a également simplifié ses lois concernant l'immigration [9]. Une consultation concernant la simplification des impôts sur le revenu a été organisée entre mars et juin 2023 [10].

Le Royaume-Uni montre donc un intérêt à simplifier en même temps le discours de l'accès aux droits, mais également les démarches et la législation qui lui correspond.

2.3.4.2 France

En France, les premiers efforts de simplification ont lieu entre 1970-1980 puis, au début des années 2000, de nombreux outils sont mis en place (COSLA, Lara, etc.). La diffusion des informations sur les sites est gérée par la direction de l'information légale et administrative. La page Missions du service public (de la DILA) [11] concernant l'information administrative ne mentionne pas le langage clair. Le but énoncé est de « faciliter et simplifier l'accès à l'information administrative et aux services en ligne ». Il semble que la France adopte une approche différente de celle du Royaume-Uni et souhaite simplifier les démarches et l'accès aux services en ligne, mais n'affiche pas clairement une volonté d'utiliser un langage clair défini. Cela est renforcé par l'absence de guide de style disponible sur le site `service-public.fr`. Il existe sans doute des guides rédactionnelles disponibles pour les rédacteurs, mais ceux-ci ne semblent pas être mis à disposition des usagers. Nous avons toutefois trouvé certains documents qui affichent une volonté de simplification du langage, sans mentionner clairement le langage clair et simple. C'est le cas du guide pratique « Vaincre la phobie administrative grâce aux sciences comportementales » qui est rédigé, comme son titre l'indique, par une équipe de chercheurs, qui « s'appuie sur les connaissances en sciences comportementales ». Ce document bénéficie de multiples sources scientifiques concernant les sciences comportementales, bien qu'il soit principalement consacré à la simplification du langage. Par ailleurs, les recommandations en matière de simplification, elles, ne sont que très peu étayées par des recherches. C'est une observation qui est faite par le chercheur Bertrand Labasse (2022) également concernant les guides de styles administratifs, qui sont généralement plutôt basés sur l'expérience que sur des connaissances scientifiques. C'est le cas des fiches et vidéos *Simplifier le langage administratif* publiées en juin 2021 par la Direction interministérielle de la transformation publique. D'après le chercheur Bertrand Labasse (2022)), ces fiches, et une partie des guides rédactionnels francophones, ont un déficit théorique et ne font pas état des recherches récentes dans le domaine. Elles

9. `lawcom.gov.uk/project/simplifying-the-immigration-rules/`, consulté le 20/10/2023

10. gov.uk/government/consultations/simplifying-and-modernising-hmrcs-income-tax-services-through-the-tax-administration-framework, consulté le 20/10/2023

11. disponible à l'adresse suivante : `www.service-public.fr/P10010` consulté le 20/10/2023

font cependant état d'une ouverture sur les sciences comportementales, en assortissant les conseils de rédaction à des concepts en science comportementale :

> LE SAVIEZ-VOUS ? Présenter trop d'options à une personne peut provoquer ce qu'on appelle une paralysie du choix (DITP, fiche n° 1 : 3)

Enfin, la France met en place certains sites et services de simplification. En 2015, un outil de simulation en ligne (`mes-aides.gouv.fr`) permet aux usagers de disposer d'une estimation globale des différentes aides auxquelles ils peuvent prétendre en fonction de leur situation (d'Etat à la réforme de l'Etat et à la simplification et Le Secrétariat générale pour la modernisation de l'action publique, 2015). Le site Service Public + [12] est mis en place afin de permettre à un usager de signaler un document qu'il trouve trop compliqué dans le but de faciliter la simplification des documents. Cette plateforme montre que l'État cherche à améliorer la communication avec les usagers. Il semble donc que le langage simple et clair, même s'il est connu des institutions diffusant l'information administrative, n'est pas la préoccupation principale pour le site `service-public.fr`. Toutefois, cette observation est à nuancer car le plan « Parlez-vous français ? » annoncé en 2024 semble vouloir remettre le langage clair au cœur du mouvement de simplification.

La France semble également intéressée par la simplification des démarches. On compte 6 lois de simplification entre 1998 et 2020 et pas loin de 350 mesures entre 2006 et 2015 (Gélédan, 2016). Cette stratégie est un objectif essentiel pour le gouvernement. La loi la plus récente est la loi d'accélération et de simplification de l'action publique promulguée en décembre 2020. Cette loi ne concerne qu'un petit nombre de démarches, pour la plupart destinées aux entreprises. Pour les usagers, elle permet de : simplifier la fourniture de justificatif de domicile et faciliter l'ouverture et la gestion du livret d'épargne populaire et d'utiliser des chèques énergies dans les logements pour personnes âgées.

L'idée de simplifier le discours administratif, mais aussi les démarches est donc bien installée chez les deux gouvernements. Les approches britanniques et françaises sont différentes. Il est difficile de trouver des traces de ces simplifications pour le Royaume-Uni, car elles ne font pas l'objet de promotion et restent le plus souvent en interne. Cependant, les guides de style et autres instructions de simplification sont disponibles sur le site `gov.uk`. En revanche en France, dans les années 2000 on retrouve une volonté de promouvoir les actions de simplifications, notamment en nommant des personnalités dans le Comité de Simplification du Langage administratif (COSLA), ou en organisant le prix « Soyons Clairs », mais les guides de rédaction ne font pas l'objet de promotion.

12. disponible à l'adresse suivante : `plus.transformation.gouv.fr`, consulté le 20/10/2023

En plus du langage simple et clair qui a été proposé comme outil de simplifications, il existe d'autres types d'outils d'aide à la rédaction claire.

2.3.4.3 Autres outils

Dans cette partie, nous décrivons trois outils différents utilisés pour la simplification des documents administratifs.

Tout d'abord, le logiciel **Lara** (Logiciel d'Aide à la Rédaction Administrative) est un logiciel créé par le ministère de la Fonction publique. Il est publié pour la première fois en 2002. C'est un logiciel libre de droits, dont le code source est accessible. Il s'agit d'un plug-in de Word qui est compatible avec les versions 95, 98, 2000 et XP. Le logiciel est destiné aux agents de la Fonction publique ayant pour mission de communiquer avec les usagers. Sa fonction première est de simplifier le langage administratif. Le logiciel s'intègre à Word et détecte les termes et expressions difficiles à comprendre. Il donne ensuite quelques recommandations sur une reformulation plus simple à comprendre. Le logiciel n'est désormais plus disponible ni utilisable, puisqu'il nécessite d'utiliser une version obsolète de Word.

La plateforme **AMesure** (François et al., 2018) est un outil d'aide à la simplification des documents administratifs mis en place par l'Université de Louvain [13]. L'outil a été créé pour aider les rédacteurs de textes administratifs à rendre leurs écrits plus lisibles, en appliquant une formule de lisibilité spécifiquement adaptée aux textes administratifs. Elle identifie également les « éléments lexicaux et syntaxiques difficiles du texte » (François et al., 2014, p. 478). Cette plateforme permet donc d'analyser des textes administratifs et de mettre en avant de segments reconnus comme difficiles à comprendre, en prenant en compte trois critères :

— difficulté syntaxique
— difficulté lexicale
— difficulté textuelle

Cet outil permet aux rédacteurs, comme le faisait l'outil LARA, d'identifier les segments présentant des difficultés de compréhension potentiels afin de les simplifier.

Enfin, l'association **Citizen Advice Bureau** a également mis en place un outil similaire afin d'aider ses rédacteurs (Johnson, 2018). Il s'agit d'une extension personnalisée sur Google Doc qui permet de s'assurer que le guide de style en place est bien respecté. Cet outil permet, en outre, d'identifier les segments à reprendre dans le texte et donne également des suggestions afin de les corriger, en accord avec les normes de rédaction mises en place. Le site utilise du *Plain language* pour l'ensemble de ses communications.

13. Disponible à cette adresse : `cental.uclouvain.be/amesure/`, consulté le 20/10/2023

L'outil est donc calibré pour respecter ces normes. Le code est disponible sur GitHub [14].

Ces trois outils possèdent des caractéristiques et des objectifs similaires : proposer un appui informatique à la simplification des textes. Ils ne constituent que des exemples parmi une multitude d'outils, mais ce sont des exemples représentatifs d'une volonté de rendre la simplification des documents administratifs systématique.

2.4 Synthèse

Ce deuxième chapitre introduit les concepts de discours, de spécialisation et de vulgarisation afin de situer notre recherche. Ces différents concepts, ainsi que les théories et les approches qui cherchent à les structurer, décrire et situer, que nous avons présentés tout au long de ce chapitre, permettent de prendre en considération le statut du discours étudié dans cette thèse, à savoir le discours numérique d'accès aux droits, et ses caractéristiques dans les différents espaces discursifs qu'il occupe.

La première partie propose un aperçu du concept de discours et de la discipline de l'analyse du discours. Il s'agit d'une discipline aux courants multiples, qui consiste à étudier la langue dans son usage en prenant en compte son contexte, et qui se rapproche de la terminologie textuelle (Condamines, 2005). Nous avons ensuite exploré les différentes définitions de la notion de discours, central à notre thèse. Un discours est une production étudiée dans son ensemble (Schiffrin et al., 2015) et dans son contexte (Bathia, 2004). Nous nous intéressons donc au phénomène de co-construction du texte, du contexte et du co-texte. Les genres de discours font également partie de notre cadre d'analyse. Il s'agit de productions discursives stables (Delavigne, 2022), qui sont des pratiques qui, de nos jours, intègrent les évolutions numériques. Dans nos travaux, les genres de discours sont des régularités linguistiques mises en œuvre pour réaliser (ou satisfaire) un rituel d'interaction dans un espace numérique. À ce titre, les discours de l'accès aux droits s'apparentent à des discours procéduraux.

Nous avons ensuite présenté le concept de discours spécialisé, qui est une utilisation spécifique de la langue afin de transmettre des connaissances spécialisées (Lerat, 1995). Elle est parlée entre experts afin de communiquer de façon claire et concise (Rouleau, 1995) dans un domaine en particulier, mais elle peut être utilisée par des locuteurs possédant différents niveaux d'expertise (Gledhill et Kübler, 2016). Les locuteurs d'un discours spécialisé font partie d'une communauté de discours (Swales, 2016), car ils présentent un certain nombre de points communs, notamment des objectifs et l'utilisation des moyens et normes linguistiques adaptés à la réalisation de ces objectifs. La notion de communauté de discours et de discours spécialisé nous a menée à question-

14. `github.com/Nyzl/1i`, consulté le 20/10/2023

ner la notion d'expertise dans le domaine de l'accès aux droits. Il existe différents types d'expertises, dans plusieurs domaines sous-jacents tels que le droit ou l'administration. La communauté de discours est donc composée d'un nombre important d'acteurs issus des services publics, mais également d'organisations militantes, ou encore des usagers eux-mêmes qui disposent de niveaux de compétences variables. Les expériences des usagers constituent des types d'expertise uniques, différents des expertises proposées par les professionnels du domaine, qui ne disposent pas du même type de connaissances (ou sur un sujet restreint). Nous avons également mis en avant une expertise de groupe, qui émerge de groupes de partage de connaissance entre usagers et qui permettent parfois d'obtenir des connaissances uniques et pointues de certaines démarches. Les littéracies adminitratives et numériques sont indispensables dans la compréhension et l'utilisation de ce discours. Il s'agit, entre autres, de compétences fonctionnelles, communicatives, structurelles et procédurales à acquérir de la part des usagers afin d'interagir avec les membres de la communauté de discours et d'accéder à ses droits. Le discours spécialisé de l'accès aux droits est à l'intersection de plusieurs langues de spécialité bien documentées : les langues de spécialité juridique, politique, administrative et institutionnelle. Chacune des caractéristiques du discours étudié sera mise en regard des études réalisées sur l'ensemble de ces langues de spécialité.

Dans un second temps, nous nous sommes intéressée à la vulgarisation du discours. La vulgarisation est une activité de diffusion d'un discours (Authier, 1982) afin qu'il soit compris par le grand public. Il s'agit pour certains d'un discours intermédiaire (Reboul-Touré, 2004) essentiel dans la communication entre experts et non-experts. Elle existe sous diverses formes suivant son objectif : semi-vulgarisation, discours pédagogique, discours médiatique, etc. Le discours étudié est un acte de vulgarisation singulier, qui a pour but de diffuser des informations aux usagers. Toutefois, le caractère particulier de l'expertise dans ce domaine lui confère une place particulère dans le champ de la vulgarisation. Le discours de l'accès aux droits fait l'objet de stratégies de simplification, dont les principaux acteurs sont les instances de l'administration, notamment en France la DILA, et au Royaume-Uni le GDS. Leur statut implique une relation hiérarchique entre un discours primaire, composé des textes législatifs et administratifs officiels, et un discours second, vulgarisé, représenté par le discours de l'accès aux droits numérique. En France, une loi concernant cette simplification a été promulguée en décembre 2020. Au Royaume-Uni, les procédés de simplification ne font pas l'objet de lois, mais sont gérés en interne par le GDS.

La Figure 2.2 propose une visualisation de la place du discours expert et du discours vulgarisé, qui forment le discours de l'accès aux droits.

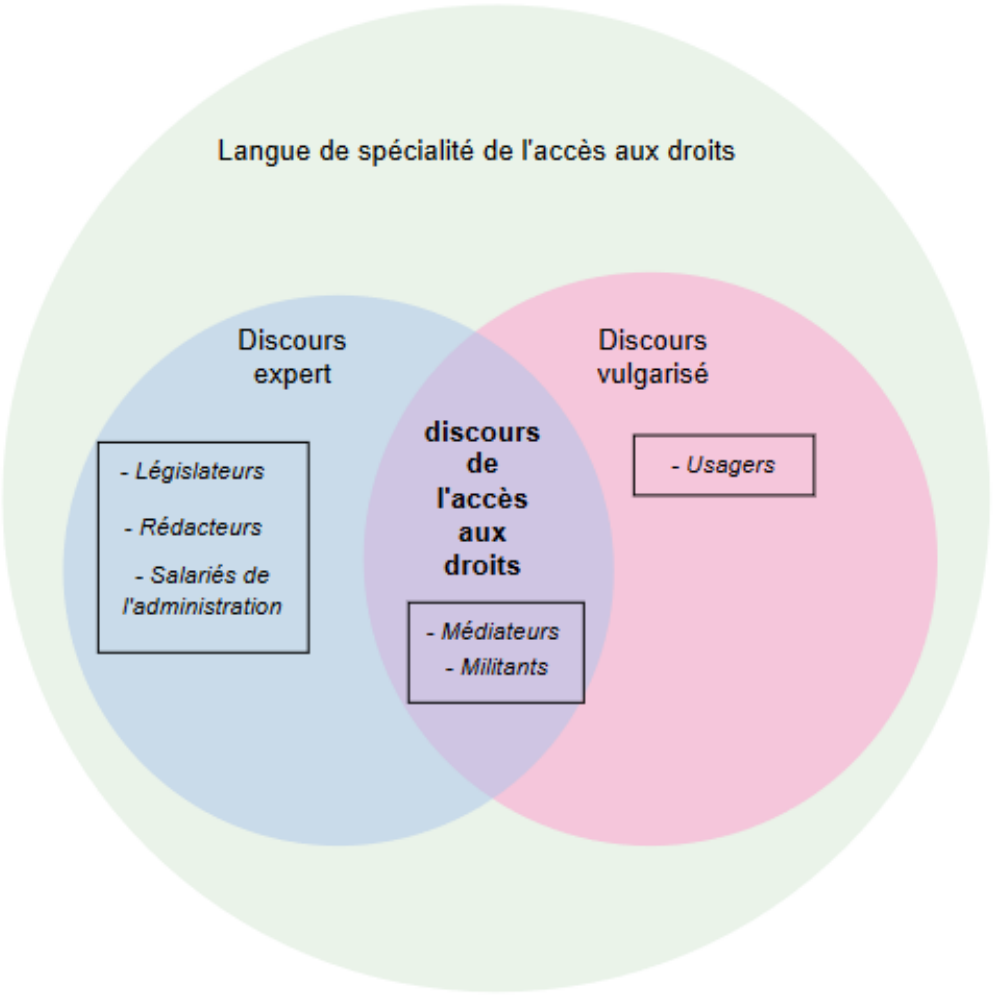


FIGURE 2.2 – Illustration de la spécialisation et de la vulgarisation du discours de l'accès aux droits

L'ensemble de ces discours font partie des langues de spécialités, et montrent différents degrés de spécialisation, traduits par des phénomènes linguistiques spécifiques à la vulgarisation et à la spécialisation. Ces phénomènes sont étudiés dans le chapitre 5.

Ces réflexions nous amènent à envisager un discours protéiforme, complexe, qui se place en même temps dans le champ du spécialisé et du vulgarisé, et qui mobilise des compétences et des connaissances multiples.

Deuxième partie

Outils méthodologiques et théoriques

Chapitre 3

Redéfinir les normes de communication

Avant de présenter l'analyse du discours numérique natif de l'accès aux droits et la méthodologie mise en place pour mener ces analyses, il convient de présenter les ancrages et les repères théoriques et historiques qui soutiennent nos choix d'outils, de méthode et de modèles d'analyse. Ces repères sont également des contextes à la lumière desquelles nous allons pouvoir aboutir à une interprétation des données plus fiables.

Dans un premier temps, il convient d'analyser et de comprendre l'évolution des discours vers le numérique. Dans un second temps, il s'agit de mettre en avant les principaux phénomènes linguistiques présents dans le discours numérique natif, notamment la variation terminologique, les groupes nominaux complexes, la phraséologie, la conditionnalité et la redondance.

3.1 Émergence d'un nouveau discours

Bien que notre recherche n'ait pas pour objectif l'étude de l'histoire du discours numérique, nous souhaitons offrir un aperçu de l'évolution des discours avec l'émergence du numérique. Cela permettra de mieux comprendre les transformations opérées et leurs impacts sur le discours numérique natif tel qu'il existe aujourd'hui.

3.1.1 De l'électronique au numérique

Nous avons étudié dans le chapitre 1 les changements qui ont pu être apportés par l'arrivée du numérique dans l'administration. De la même manière, l'arrivée de l'ordinateur dans un premier temps puis d'internet dans un second temps a induit des transformations dans la définition et l'étude des discours spécialisés dans la société. Ces transformations touchent les normes de communication dans l'ensemble des discours, qu'il s'agisse de discours spécialisés ou de la langue générale, de discours écrits ou oraux.

Dès les années 1990, les chercheurs et chercheuses s'intéressent aux discours qui émergent en ligne, à l'instar de Herring (1992) qui publie dès 1992 un article questionnant la place des femmes dans les discours numériques. Divers domaines se lancent dans l'étude de ces discours sous plusieurs appellations : les discours électroniques, la communication médiée par ordinateur (en anglais *Computer mediated langage*) ou encore les écrits d'écran. Les discours numériques se développent en nombre et présentent des caractéristiques variées. La communication médiée par ordinateur étudie par exemple les discours numériques oralisés qui se développent dans les messageries instantanées, les forums et les SMS. Les différents termes utilisés pour désigner ces discours témoignent également d'une diversité des approches et d'une pluridisciplinarité du sujet. En effet, chaque domaine tente de s'approprier ces nouveaux types de discours.

Les premières études (e.g. Souchier 1996; Herring 1999; Gilmont 2004; Chartier 2005) se concentrent particulièrement sur la matérialité des discours numériques, notamment en s'intéressant à la machine elle-même : l'ordinateur. Les dénominations choisies pour désigner ces types de discours témoignent de cet intérêt, avec par exemple la « communication médiée par ordinateur » qui met en avant l'objet utilisé pour communiquer. En 1996, Emmanuël Souchier décrit ce qu'il appelle « les écrits d'écran », en s'attachant à une composante de la machine (l'écran) comme support. Dans son article, le chercheur questionne la matérialité de l'écrit d'écran, qui est pixélisé :

> Si pour l'écriture le passage de la trace à la forme entérine une première rupture corporelle (le geste est supplanté par la marque métallurgique de la typographie), le passage au pixel traduit une rupture, plus fondamentale encore, à la fois visuelle et conceptuelle. [...] D'une trace inscrite sur un support, nous sommes passés à une trace électronique fugitive qui ne présente plus de matérialité tangible. (Souchier, 1996, p. 108).

L'article dont est extraite cette citation (Souchier, 1996) replace l'apparition de l'écriture électronique dans l'histoire et propose des considérations sémiologiques et philosophiques sur sa matérialité et son rapport à l'humain. Plusieurs spécialistes du domaine cherchent également à comparer les interactions en ligne aux interactions dites IRL [1](c'est-à-dire des interactions en personne) afin de situer ces interactions par rapport à celles déjà existantes (par exemple Topper (1997); Herring (1999)). Ces travaux n'analysent pas le discours par rapport à sa matérialité mais plutôt les pratiques discursives. Rachel Panckhurst, dans un ouvrage consacré au discours électronique, définit l'ordinateur comme un médiateur et indique que « son utilisation modifie ainsi notre discours et notre façon de communiquer avec autrui » (Panckhurst, 2006, p. 1).

1. In Real Life

D'autres se posent la question du support d'écriture et de l'accès aux textes que l'ordinateur et internet ont transformés.

> en l'occurrence l'écran de l'ordinateur, qui fait apparaître face au lecteur les différents types de textes qui, dans le monde de la culture manuscrite et a fortiori de la culture imprimée, étaient distribués entre des objets distincts (Chartier, 2005, p.1)

L'écriture numérique est alors envisagée en opposition avec l'écriture manuscrite en raison de ses différences, notamment en termes de matérialité et d'immatérialité.

Les chercheurs de ce domaine interrogent la validité des interactions en ligne dans différents environnements. En 1997, un ouvrage questionne l'impact des technologies sur les communications en décrivant et en analysant les pratiques d'écritures d'un groupe d'étudiants (Davis et Brewer, 1997). Dans leur ouvrage, les auteurs parlent d'espaces virtuels qui sont comparés aux espaces dits « réels ». Cette analyse se rapproche de l'idéologie du dualisme numérique, qui consiste à faire une différence entre ce qui est virtuel et réel (ou physique). Ce qui se passe en ligne ne serait ainsi pas réel, contrairement aux interactions en personne qui, elles, le seraient (Paveau, 2017).

De nombreux chercheurs et chercheuses comparent l'apparition des textes électroniques à l'apparition de l'imprimerie dans tous les domaines s'intéressant aux langages (e.g. Souchier 1996; Herring 1999; Gilmont 2004; Chartier 2005). Il s'agit de la première vague d'étude du discours numérique que décrivent Garcés-Conejos Blitvich et Bou-Franch (2019), dont l'article retrace l'histoire dans le domaine de l'analyse du discours numérique. Cette première vague est composée d'études linguistiques majoritairement descriptives, entre les années 1990 et 2000.

À partir des années 2000, l'étude des discours numériques est bien installée et représente un sujet dominant en analyse de discours. Les chercheurs s'interrogent sur les méthodes d'analyse mais aussi sur les genres discursifs qui se développent avec le numérique. Il s'agit de la deuxième vague identifiée par Garcés-Conejos Blitvich et Bou-Franch (2019) dans leur article. Cette vague semble se concentrer davantage sur l'analyse des pratiques des utilisateurs. Certaines revues consacrent des numéros à ces problématiques comme *Les Carnets du Cediscor* qui publient en 2004 une revue intitulée « Les Discours de l'Internet » (Mourlhon-Dallies et al., 2004). Entre les années 2000 et 2010, les communautés scientifiques parlent encore largement de *discours électronique* plutôt que de *discours numérique*. Les recherches sur Isidore et Google Scholar confirment ces tendances avec une quasi-absence du terme « discours numériques » dans les articles entre 2000 et 2010[2], tendance qui se confirme également pour les ouvrages, selon Ngram

2. Sur Google Scholar, le premier article comportant le terme « discours numérique » dans son titre paraît en 2010 (Saemmer, 2010). Isidore ne recense aucun document contenant « discours numérique »

Viewer (voir Figure 3.1).

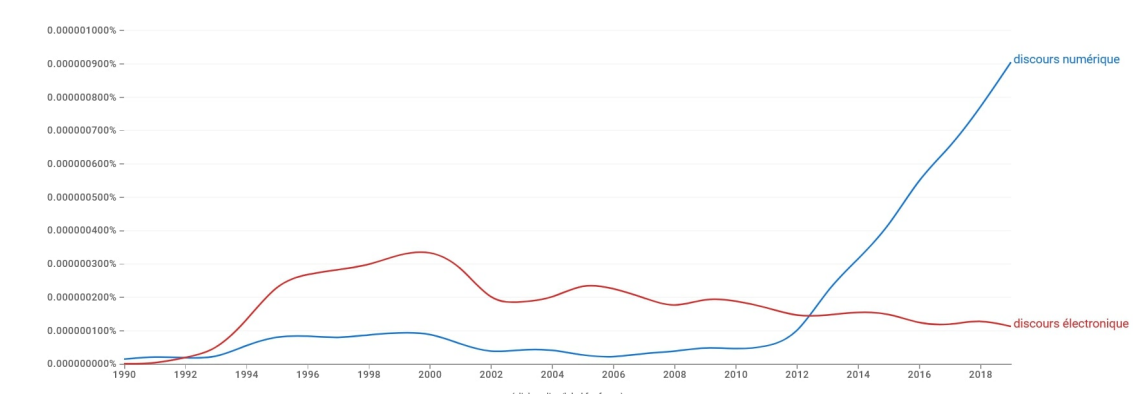


FIGURE 3.1 – Graphique représentant l'évolution des termes **discours numérique** et **discours électronique** dans Ngram Viewer

À partir de 2010, on observe une apparition de ce terme dans les publications scientifiques, puis en 2012, une explosion de son utilisation dans la sphère francophone, notamment avec les ouvrages et publications de Marie-Anne Paveau (Paveau, 2017, 2019). C'est à partir des années 2010 qu'apparaît en France de manière distincte un nouveau champ de recherche qui se focalise sur l'analyse des discours numériques. Aujourd'hui, la matérialité des discours n'est plus au centre des préoccupations et les principaux travaux ne font plus de distinction entre un discours qui serait « virtuel », c'est-à-dire produit en ligne, et « réel ». Dans leur article, Garcés-Conejos Blitvich et Bou-Franch (2019) identifient une troisième vague qui consiste à étudier les discours numériques au-delà de leurs médias, dans leur multimodalité. Dans cette thèse, nous nous inscrivons dans cette troisième vague, qui prend en compte les pratiques langagières et la dimension interactive du discours, tout en se plaçant dans une approche post-dualiste, comme définie par Paveau (2015); Vial (2014). Cela signifie que nous ne pensons pas qu'il soit utile de différencier le monde virtuel et le monde réel. Les informations et communications en ligne ont des conséquences tangibles à l'instar des communications administratives. Les discours et interactions en ligne sont tout aussi réels que les interactions hors ligne et il est important de reconnaître que les espaces discursifs numériques en ligne font partie intégrante de notre société.

3.1.2 Nouveaux paradigmes de communication

Les nombreux espaces en ligne introduisent de nouveaux paradigmes de communication, produisant ainsi une variété de discours numériques. Les spécialistes des discours

dans le titre pour cette période.

numériques s'intéressent à ces divers espaces comme les réseaux sociaux (Paveau, 2013; Bibié-Emerit, 2016), les blogs (Vicari, 2021) ou encore les jeux vidéo (Krichane et Rochat, 2018). Notre étude se concentre sur un espace en particulier : le site web.

3.1.2.1 Les sites web

Un site web est un ensemble de pages web liées entre elles comprenant une page d'accueil et hébergé au même endroit (Lavoie et Nielsen, 1999). Il est généralement édité par une seule entité ou organisation. Cette dernière peut proposer un guide de style linguistique (destiné aux éditeurs et éditrices du site) et une identité visuelle. Une page web est un ensemble d'informations composé d'une ou plusieurs ressources web, destinée à être consultée comme une entité et désignée par un identifiant unique (Lavoie et Nielsen, 1999). Dans les sites sélectionnés pour notre étude, chaque page est mise en forme en HTML (Hypertext Markup Language), un langage de balisage permettant au webmaster de hiérarchiser et de mettre en forme les contenus. Il s'agit du format le plus courant pour mettre en forme les sites. Les contenus discursifs doivent donc être pensés pour correspondre aux normes du site. Dans notre étude, nous nous intéressons à un type de site particulier, qui a vu le jour avec le web 2.0 : les sites dynamiques.

Un site web est dynamique lorsque les pages qu'il contient sont générées à la demande par le serveur selon les requêtes de l'utilisateur. Il est donc nécessaire pour l'utilisateur d'interagir avec la page pour accéder à certains contenus. Le contenu des sites dynamiques est variable, car il dépend donc des choix faits par l'utilisateur. À l'inverse, un site statique est visible tel qu'il a été conçu et n'utilise pas de base de données. Il peut cependant avoir certaines animations (musiques, vidéos, menus…). Une autre caractéristique des sites web est leur mise à jour qui les rend dynamiques dans le temps et parfois même instables. Nous reviendrons sur cette caractéristique dans la section 6.2.3, caractéristique que nous avons soulignée dans la section 4.2.5.

Les sites web ont déjà fait l'objet de terrains de recherche dans plusieurs domaines spécialisés. Les linguistes travaillant sur la langue du tourisme se sont par exemple largement penchés sur le rôle des discours sur les sites web relevant de ce secteur notamment pour la traduction (Vaupot, 2020; Vicari, 2021; Hallett et Kaplan-Weinger, 2010). En linguistique de corpus, certains chercheurs ont étudié des sites web à partir de corpus récoltés automatiquement (Fairon, 2000).

Les pages d'accueil constituent des objets d'étude en particulier. Elles sont parfois d'ailleurs qualifiées de microgenres (Bonhomme, 2015). Elles sont un moyen d'accéder à l'ensemble de l'arborescence du site web et regroupent généralement les caractéristiques du site. Il s'agit d'une « carte d'identité du site » (Silletti, 2016, p. 55).

De par leurs traits génériques et stables, les pages d'accueil de sites web étudiés dans

cette thèse cumulent le même type d'informations, et appartiennent, par conséquent, au genre « site web informatif » et à l'hypergenre « site institutionnel » (Bonhomme, 2015, p. 43). Le site web est l'un des nombreux espaces où se développe le discours numérique. Dans le cadre de cette thèse, nous nous concentrons spécifiquement sur le discours issu de sites web pour des raisons liées au domaine comme nous l'avons présenté dans le chapitre 1, mais également pour des raisons méthodologiques qui sont exposées dans le chapitre 4.

3.2 Le discours numérique

3.2.1 Définition

L'analyse de discours et les sciences du langage en général se sont récemment intéressées aux corpus numériques natifs en tant qu'objet d'étude. Ces discours, produits et consultables dans un environnement numérique, possèdent certaines caractéristiques communes. Le discours étudié est considéré en analyse des discours numériques comme une production numérique native, un concept institué par Marie-Anne Paveau, afin de décrire ces nouveaux types d'écrits.

> On appelle natives les productions élaborées en ligne, dans les espaces d'écriture et avec les outils proposés par internet, et non portées après numérisation d'espaces scripturaux et éditoriaux prénumériques aux espaces numériques connectés. (Paveau, 2017, p. 27)

Cette définition ainsi que les travaux de la chercheuse ont mis en lumière l'importance d'étudier les discours numériques selon de nouveaux paramètres. Les discours les plus étudiés jusqu'à présent sont ceux produits sur les réseaux sociaux (Bibié-Emerit, 2016; Ghliss, 2019; Paveau, 2013). Les études sur les discours des réseaux sociaux ont été les premières à se développer et à gagner de l'importance car ces derniers interrogent de façon radicale nos habitudes discursives. En outre, comme nous l'avons mentionné dans la section 3.1.1, la notion de discours numérique commence à apparaître dans les années 2010 en même temps que l'avènement des réseaux sociaux. D'autres formes de discours numérique existent depuis de nombreuses années, à l'instar des sites web et, dans une moindre mesure, des forums. Les problématiques de recherche qui y sont liées sont donc différentes.

Contrairement aux réseaux sociaux qui ont connu une explosion de popularité, le développement des sites web s'est fait progressivement. À titre d'exemple, le site `service-public.fr` a été mis en place en 2000, puis n'a cessé d'évoluer au cours des vingt dernières années. La discursivité d'un site web peut sembler moins radicale-

ment transformée par le numérique, car la dimension interactive est moins présente. Sur les sites choisis pour cette étude, l'utilisateur ne peut pas interagir avec les autres utilisateurs, mais il doit interagir avec la page web pour naviguer. Les textes présents sont bien pensés et conçus pour une diffusion des sites web, à l'aide des outils numériques, au même titre que les discours des réseaux sociaux. Ces discours répondent ainsi à de nouvelles normes des sites web.

3.2.2 Caractéristiques

Dans cette section, nous présentons les diverses caractéristiques des nombreux genres de discours numérique. Certaines caractéristiques du site web sont partagées par l'ensemble des discours numériques, tandis que d'autres sont réservées au genre du site web. Nous allons partir d'une définition générale du discours numérique pour aller à une définition spécifique du discours numérique des sites web.

Marie-Anne Paveau distingue six traits définitoires des discours numériques : composition, délinéarisation, augmentation, relationalité, investigabilité et imprévisibilité (Paveau, 2015). La délinéarisation est un trait particulièrement caractéristique, qui décrit la capacité à reconfigurer le texte grâce aux liens hypertextuels, ce qui modifie les dispositifs de cohésion textuelle (Mayeur et Paveau, 2020). Ce concept fait également référence aux éléments interactifs et à la création d'un parcours de lecture des usagers. Une des caractéristiques du discours numérique est la lecture partielle de l'ensemble des pages et la navigation. D'après Vitali-Rosati :

> La navigation est la métaphore utilisée pour décrire les mouvements dans l'espace du web. (Vitali-Rosati, 2014, p. 78)

Les « mouvements » font référence à la capacité qu'ont les utilisateurs à se déplacer de page en page et de texte en texte sur le web. Cette caractéristique est particulièrement présente sur les sites web qui sont composés de plusieurs pages organisées hiérarchiquement. Sur les sites web dynamiques, certains éléments permettent de cacher ou de dévoiler du texte. Ces éléments peuvent être dépliés par le lecteur selon ses exigences de lecture, la construction des pages n'étant pas fixe mais modulable (Chartier, 2005; Silletti, 2016). Les utilisateurs bénéficient alors de parcours de lecture enrichis (Silletti, 2016).

Une autre caractéristique du discours numérique qui s'applique particulièrement sur les sites web est définie par Maingueneau (2016). Il indique que les utilisateurs ne sont pas face à un texte, mais plutôt à « une mosaïque plus ou moins dense de modules hétérogènes » (Maingueneau, 2016, p. 14). Ces modules sont en fait les différentes parties de la page web, qui peuvent être : les bannières, la barre de recherche, des photos, des articles,

des titres, des vidéos, etc. Il reprend alors le concept de délinéarisation du discours en parlant de « scansions d'Internet » (Maingueneau, 2016, p. 4). En effet, une page n'est jamais visible dans son intégralité car l'utilisateur n'a accès qu'à des fragments du texte. Il doit donc composer avec ces modules, les relier entre eux pour progresser à travers cette mosaïque textuelle, créant ainsi son propre parcours discursif. Ces caractéristiques façonnent les discours présents sur les sites web et en font un objet d'étude particulier. Les hyperliens, qui permettent à l'utilisateur de se déplacer entre les pages et de construire son parcours de lecture, sont également un élément central du discours.

3.2.3 Discours hypertextualisé

Comme nous l'avons déjà indiqué dans la section 3.2.1, le discours numérique est caractérisé par la présence de technosignes. Parmi ces technosignes, l'hyperlien occupe une place importante, notamment dans les discours issus de sites web. Le *Dictionnaire fondamental de l'informatique et de l'Internet* publié par l'Université de Montréal en donne la définition suivante :

> Énoncé contenant une adresse qui est placé dans une PAGE pour permettre à un INTERNAUTE d'atteindre une autre PAGE ou un emplacement spécifique dans celle-ci. (L'Homme, 2009)

Il s'agit donc d'un objet dynamique permettant de se déplacer dans une page ou sur un site. Cette définition est spécialisée dans le domaine de l'informatique et rédigée pour des professionnels du domaine. D'autres chercheurs en analyse de discours ont également défini les hyperliens, comme Souchier et al. (2003, p. 10), qui parlent de « signes passeur », qui « donnent accès aux multiples modalités du texte ». Ces signes comprennent les boutons, menus déroulants, etc. L'hyperlien est donc un type de signe passeur.

Simon (2015, p. 1) s'intéresse au discours hypertextualisé et indique que le discours hypertextualisé est un « discours augmenté par la présence de liens hypertextes renvoyant à d'autres discours ». Comme dans la définition issue du dictionnaire spécialisé en informatique, l'importance est mise sur la capacité de l'hyperlien à mener l'utilisateur vers d'autres discours. Cette définition permet d'envisager un « ensemble interdiscursif plus large, contenant des points de rencontre concrétisés par des hyperliens » (Simon, 2018, p. 2). Dans ce cadre, les liens hypertextes se réfèrent à un cadre théorique instauré par Bakhtine (1995) avec la notion du dialogisme. Le dialogisme est « le lien que tout énoncé entretient avec les énoncés produits antérieurement ainsi qu'avec les énoncés à venir que pourraient produire ses destinataires » (Charaudeau et Maingueneau, 2002, p. 175). Dans le cas des liens hypertextes, cette relation se matérialise par un dialogisme

avec les pages vers lesquels les liens redirigent l'utilisateur qui peuvent être des discours sources ou complémentaires. Ce dialogisme pourrait se manifester par des marques linguistiques (Authier, 1982). Certains liens permettent de se déplacer sur la même page, en reliant les différentes parties de cette structure polymorphe.

Les relations interdiscursives des énoncés sont également matérialisées par les hyperliens. C'est ce que nous apprend Simon (2015, p. 6), en indiquant que « l'hypertextualité numérique est une traduction technologique des phénomènes d'intertextualité ». L'intertextualité est un phénomène présent dans tous types de textes. Il est parfois explicité par des renvois ou des citations, et parfois implicite. L'hyperlien peut être une manifestation technologique des phénomènes d'intertextualité présents sur les sites web. Cette intertextualité peut être interne au site quand les hyperliens aident l'usager à structurer son parcours, ou externe quand elle renvoie vers des sites externes. Dans le cas des textes de vulgarisation, certains chercheurs observent que le lien peut servir à « décomplexifier » le discours, selon les intérêts de lecture de l'utilisateur (Silletti, 2016). Autrement dit, l'utilisateur peut approfondir ses connaissances sur certains sujets en cliquant sur les liens pour accéder à des informations plus précises.

Il existe plusieurs termes pour qualifier les différents types de discours qui accompagnent l'hyperlien. L'hyperlien est tout d'abord un signe passeur, accompagné d'un *discours d'escorte*. La notion de discours d'escorte désigne initialement le paratexte (Genette, 1982) puis elle est actualisée par Simon (2015). Dans les discours numériques, il s'agit du discours qui entoure et compose l'URL. Il s'agit du fragment de texte cliquable qui est aussi appelé *discours d'accompagnement* par Paveau (2016). La mise en œuvre concrète de l'hyperlien et de cette intertextualité est bien sûr le clic. L'utilisateur est ensuite redirigé vers une page qui dispose également d'un discours appelé le *discours hors écran* (Simon, 2015). Le discours hors écran est un concept qui décrit le discours auquel accédera l'utilisateur s'il clique sur l'hyperlien (Paveau, 2016). Il n'apparaît pas sur l'écran tant que l'utilisateur n'a pas cliqué et est donc initialement « hors écran ».

Plus récemment, des chercheurs en journalisme se sont intéressés au rôle des hyperliens dans l'accès aux articles. Une étude en particulier établit une typologie du « bon lien » et du « mauvais lien » (De Maeyer et Holton, 2016). Cette étude met en avant l'importance des liens, qui peuvent être bénéfiques pour enrichir le contenu, ou renforcer la fiabilité du discours. Cependant, la fonction du lien doit être prise en compte afin d'évaluer sa pertinence. D'autres études choisissent de se concentrer sur le rôle de l'hyperlien sur les réseaux sociaux. Jakob (2022) s'intéresse aux fonctions des liens sur Twitter. Il identifie deux types de liens : informatif et argumentatif. Les résultats de cette étude montrent également que le contexte social et culturel, notamment le pays de provenance des tweets, influence le type de lien utilisé par les internautes. Eisenlauer (2013)

consacre un ouvrage à l'hypertexte sur le réseau social Facebook.

Dans toutes ces recherches, les chercheurs et chercheuses font référence à l'utilisateur qui choisit ou non de cliquer sur le lien afin d'accéder au contenu.

> *Hypertext readers choose their individual reading paths through a multiplicity of hyperlinked textual fragments and thus become partly involved in the production of a text.* (Eisenlauer, 2013, p. 28)

Plutôt qu'un fil discursif délinéarisé, certains chercheurs parlent d'un fil discursif multilinéaire, car le lien hypertexte permet de créer plusieurs parcours discursifs. Storrer différencie des réseaux d'hyperliens dits « fermés », c'est-à-dire les liens qui ne redirigent que sur le site en interne et qui forment ainsi une série de texte finie, et les réseaux d'hyperliens dits « ouverts », qui eux créent une structure de textes dynamique en redirigeant sur différents sites (Storrer (2000) cité par Eisenlauer (2013)).

Dans le cas du discours de l'accès aux droits, la structure du discours a tendance à former un réseau d'hyperliens fermé, car les sites web de l'accès aux droits redirigent les utilisateurs majoritairement en interne (voir la section 6.2.2 dans le chapitre 6). Cacchiani (2018), dans son analyse des pages concernant le Copyright sur le site web `gov.uk`, décrit le discours numérique de l'accès aux droits comme imbriqué et extensible. En effet, la structure hiérarchique du site crée des pages imbriquées, tandis que les liens créent un discours extensible pour l'utilisateur.

L'étude des liens dans le corpus créé pour cette étude permet de mettre en avant les caractéristiques des liens hypertextes dans le discours numérique de l'accès aux droits.

3.3 Spécificités linguistiques

Dans le but de caractériser le discours numérique natif d'accès aux droits, nous avons cherché à identifier ses traits les plus saillants. Ces traits semblent se manifester à trois niveaux :

- lexical, par la présence de la variation terminologique, des groupes nominaux complexes et des unités phraséologiques spécifiques,
- syntaxique, par la présence des structures conditionnelles et
- stylistique, par la présence du phénomène de redondance.

Dans les sections suivantes, nous présentons les outils théoriques de la linguistique sur lesquels reposent nos analyses présentées dans les chapitres 5 et 6 consacrés aux résultats.

3.3.1 Variation terminologique

L'étude de la variation terminologique s'intéresse aux différentes dénominations employées pour désigner un même concept dans les discours spécialisés et dans un domaine en particulier. Freixa en donne la définition suivante :

> *the phenomenon in which one and the same concept has different denominations ; this is not just any formal variation (variation between a term and a periphrasis, or a definition, for example), but is restricted to variation among different denominations, i.e., lexicalised, forms, with a minimum of stability and consensus among the users of units in a specialised domain* (Freixa, 2006, p. 1)

Les premières recherches en terminologie menées par des figures pionnières comme Wüster avaient pour objectifs de créer une norme et de standardiser les langues de spécialités et leur terminologie. Ainsi, la variation, même si elle était prise en compte, était vue non pas comme un sujet de recherche à part entière, mais comme un phénomène à réduire (Humbley et Picton, 2017). Par ailleurs, Pecman (2014) démontre que ces besoins et choix, qui ont fortement marqué la pratique de la terminologie en tant que science, ont mené à une conclusion hâtive, voire naïve, que la variation et l'instabilité conséquente des termes nuisent à la communication et ne peuvent remplir aucun rôle utile en langue de spécialité. Or, c'est grâce à ce phénomène que les connaissances peuvent être remises en question et les nouvelles propositions et découvertes voir le jour (Pecman 2012, 2014, 2018). Si autrefois la variation se trouvait reléguée au second plan, aujourd'hui elle constitue une problématique terminologique amplement étudiée. Le phénomène de variation est étudié sous différents angles comme celui de la néologie (e.g. Pecman 2014 Humbley 2018), celui de la vulgarisation (Delavigne, 2017; Ledouble, 2023) ou encore celui de l'évolution diachronique (Dury, 2019, 2021; Dury et Picton, 2009). Il a très tôt été étudié dans le contexte de la transmission du savoir spécialisé aux non-experts. Cabré, dès 1999, explique que le locuteur ou rédacteur adapte la terminologie à la situation de communication (Cabré, 1999, p. 76). Il peut chercher à rendre un terme plus compréhensible en utilisant des synonymes. Cette théorie est reprise par Freixa (2006) dans un article qui réalise une typologie du phénomène de variation. Lorsque la variation dépend du degré de spécialisation du discours, il s'agit alors d'une variation verticale (appelée aussi diastratique ou socioculturelle). Dans la typologie de Freixa (2006), une variation peut être discursive si elle répond à des besoins stylistiques ou relatifs à l'expression des concepts. D'après l'auteure, les différences dans la conceptualisation mènent à des variations cognitives. Ces variations peuvent être dues à des concepts imprécis ou à des idéologies différentes. L'auteure remarque qu'un terme opaque est plus souvent utilisé entre experts car il a une certaine portée sociale et un certain pouvoir. En effet, connaître,

comprendre et savoir utiliser les termes opaques permet de créer un sens d'appartenance à une communauté, en fermant au demeurant les portes à tous ceux pour qui les connaissances cristallisées par ces termes restent inaccessibles. Ce phénomène ne serait pas observable dans un discours de vulgarisation. Au contraire, l'utilisation de variants aurait pour but « d'aider le lecteur à saisir les limites du concept » (Freixa, 2006, p. 59).

La variation terminologique peut également être liée à une différence de point de vue. C'est ce que Freixa appelle la variation cognitive. Ces différences de point de vue et l'utilisation de la variation qui lui correspond ne sont pas toujours conscientisées par les experts (Condamines, 2010).

Il apparaît donc que le phénomène de variation est lié à la compréhension et à l'expression dans un domaine spécialisé. Pecman (2014) indique que les variants terminologiques sont utilisés pour reconfigurer les connaissances en proposant de nouveaux termes qui seront d'abord instables et donc caractérisés par la variation. Les experts utilisent les variants pour exprimer leur expérience, les transférer dans un système linguistique et donc, pour les comprendre. La variation en terminologie est un sujet riche, qui répond à des besoins d'expression, mais aussi à des contraintes.

Le discours numérique de l'accès aux droits relevant d'un domaine spécialisé et ayant des marques de vulgarisation, ce phénomène y est très présent. À l'aide du cadre théorique posé dans la présente section, nous analysons le phénomène de variation dans la section 5.1.3 du chapitre 5.

3.3.2 Groupes nominaux complexes

En langues de spécialité, les groupes nominaux complexes font l'objet de nombreuses recherches. Ils sont porteurs de connaissances et renvoient à des concepts spécialisés (Kübler et al., 2017). Ils permettent également un compactage de l'information, ce qui est particulièrement adapté aux langues de spécialité qui ont pour objectifs de communiquer de façon efficace (Kübler et al., 2016).

Maniez s'intéresse aux difficultés que posent les groupes nominaux complexes aux apprenants de l'anglais médical (Maniez et Thoiron, 2004; Maniez, 2007). Ses recherches indiquent plusieurs facteurs affectant la compréhension. Les différents éléments des groupes nominaux issus du discours spécialisé sont apposés, et les apprenants pourraient avoir des difficultés à identifier les liens entre les mots. Cette conclusion se rapproche de celle faite par Kübler et al. (2016) qui indique que les étudiants présentent des difficultés à identifier le noyau du terme. Ce phénomène est accentué car ces groupes nominaux complexes se manifestent de différentes manières en français et en anglais. En anglais, on retrouve des groupes nominaux avec prémodification tandis qu'en français on utilise la post-modification avec des groupes prépositionnels (Maniez et Thoiron, 2004).

La prévalence des groupes nominaux complexes en anglais de spécialité en fait donc un sujet de recherche productif dans le domaine de la terminologie. Les recherches de Pecman (2004) indiquent que les groupes nominaux complexes sont utilisés pour former de nouveaux termes. Cette observation se vérifie dans le discours étudié. Dans le type de discours administratif étudié, les dispositifs d'aide se formalisent souvent en tant que groupes nominaux complexes, qui sont des termes complexes. Nous avons observé une distinction entre les termes simples (**allocation**, *benefits*), les termes composés (**allocation chômage**, *Student visa*) et les termes complexes (**allocation d'aide au retour à l'emploi**, *New style jobseeker's allowance*). Sur le plan de la dynamique lexicale, ces termes ne se comportent pas de la même manière. Les termes complexes sont souvent compactés par des sigles ou des acronymes et font l'objet d'une variation terminologique plus importante que les termes composés ou simples. Nous distinguons donc les termes composés simples de type NN (en anglais *Child benefit* et en français **allocation chômage**) ou Adj N (en anglais *Universal Credit* et en français **allocations familiales**) des termes composés complexes de type NN's N pour l'anglais, avec par exemple *Jobseeker's allowance*, ou encore N N NN's N avec *New Style Jobseeker's allowance* et pour le français des termes avec postmodification prépositionnelle comme la **prestation partagée d'éducation de l'enfant** (N Adj prep N prep N) ou la **prestation d'accueil du jeune enfant** (N prep N prep Adj N). Il existe d'autres types de groupes nominaux complexes dans le discours qui ne désignent pas des dispositifs d'aides juridiques et qui n'ont pas forcément le statut de terme. Nous retrouvons par exemple **absence de réponse** ou **durée maximale de versement** en français et *13 week assessment phase* en anglais.

3.3.3 Approche ontologique de la terminologie

Pour compléter l'approche terminologique des groupes nominaux complexes, nous utilisons des outils théoriques et pratiques issus de l'ontologie et de l'ingénierie des connaissances. Une ontologie est « une conceptualisation d'un domaine, c'est-à-dire une définition formelle (dans un langage compréhensible par un ordinateur) des concepts et de leurs relations décrivant une réalité partagée par une communauté de pratique » (Roche, 2012). Studer et al. (1998) proposent également la définition suivante :

> *An ontology is a formal, explicit specification of a shared conceptualization.*
> (Studer et al., 1998, p. 15)

L'ontologie et la terminologie sont particulièrement liées mais constituent des champs d'études différents. L'ontologie s'intéresse aux dénominations des concepts, tandis que la terminologie s'intéresse aux termes utilisés en discours. D'après Rastier (1995), la terminologie décrit des systèmes notionnels qui correspondent à l'ontologie de domaines

de spécialités. La terminologie décrit effectivement l'usage des termes, mais également les liens entre ces termes à l'aide d'arborescences.

En s'intéressant à l'ontologie en terminologie, on adopte une approche onomasiologique, en partant du concept pour étudier le mot ou les mots utilisés pour le désigner (Humbley, 2018). L'ontologie permet donc de se concentrer sur la dimension conceptuelle, tout en établissant les liens entre les différents concepts afin de décrire le réseau conceptuel. Roche (2008) parle même d'ontoterminologie, qui s'appuie sur une approche onomasiologique et prend en compte les dimensions linguistiques et conceptuelles de la terminologie. Cette approche permet de distinguer le terme et son usage, d'un concept et sa dénomination, tout en intégrant ces deux dimensions.

Dans notre thèse, nous proposons une approche à la fois sémasiologique (en partant des termes pour accéder au concept) et onomasiologique. Dans un premier temps, une approche sémasiologique nous permet de faire un état des lieux des termes utilisés en corpus et à partir de ces termes relevés nous pouvons identifier les concepts clés du domaine. Une fois ces concepts identifiés, nous proposerons une approche onomasiologique en étudiant les concepts identifiés et la variation autour de ces concepts. La combinaison de ces approches nous a permis de construire une ontologie de document de séjour, accompagné d'une liste de variants et de termes vedettes dans la section 5.1.5.

3.3.4 Sigles et acronymes

Les sigles et acronymes occupent une place particulière dans le discours administratif numérique. La siglaison est un procédé néologique (Percebois, 2001) qui peut permettre de simplifier un terme en condensant le concept. De ce fait, les longueurs des termes formés par des groupes nominaux complexes sont réduites par l'utilisation des sigles et d'acronymes. C'est le principe d'économie du langage (Martinet, 1969) qui pousse les locuteurs à simplifier les termes complexes en les réduisant. Ce principe indique que les communications spécialisées tendent à faire évoluer le langage afin de réduire les efforts du locuteur, comme l'indique Vicentini (2003).

> *any non-economical change, which would bring about an excessive cost in terms of efforts and constitute an obstacle to comprehension, will be automatically removed or avoided.* (Vicentini, 2003, p.43)

D'après cette citation, les deux éléments importants sont la facilité de compréhension du discours et le principe d'économie ou de moindre effort. Les acronymes sont particuliers puisqu'il s'agit de créer une nouvelle unité lexicale, prononcée comme un mot à part entière en partant d'une unité lexicale composée d'au moins deux éléments (Luján-García, 2020), tandis que le sigle est épelé. Toutefois, l'utilisation des acronymes

et sigles peut causer des problèmes de compréhension notamment chez les locuteurs non experts, qui n'ont plus accès aux termes initiaux et qui donc ne peuvent pas forcément retracer l'origine du terme. De plus, les termes du domaine étant complexes et composés de plusieurs éléments, ils forment également des acronymes et sigles complexes, c'est-à-dire des acronymes de parfois plus de 3 lettres (e.g. **PrePaRe**, **Paje**).

Nous étudions la place et l'utilisation des acronymes dans notre corpus, en particulier la place des sigles et des acronymes liés aux dispositifs d'aide, dans la section 5.1.4.

3.3.5 Redondance/Répétitions

La répétition et la redondance sont des éléments clés de notre discours. Selon la théorie de la cohésion de Halliday et Hasan (1976), la répétition d'éléments lexicaux est une composante clé de la cohésion du discours, renforçant sa clarté et sa cohérence. Morley (2006) indique que la répétition d'éléments lexicaux, qui peuvent être employés dans des structures phraséologiques reconnaissables, est un élément de cohésion évident. Mahlberg (2006) indique que la répétition mais aussi l'utilisation des synonymes est une stratégie afin de créer des éléments de cohésions dans un texte. Pour le chercheur, la répétition est un phénomène complexe, si bien que l'utilisation, par exemple, de *say* et *repeat* peut dans certains cas être une répétition lexicale de *tell*, d'après le concept de *reiteration* de Halliday et Hasan :

> *the repetition of a lexical item, or the occurrence of a synonym of some kind, in the context of reference ; that is, where the two occurrences have the same referent* (Halliday et Hasan, 1976, p. 318)

La redondance et la répétition sont des caractéristiques essentielles du discours juridique, qui est composé de nombreux préfabriqués. Cette caractéristique est essentielle pour assurer une homogénéité du genre (Rey, 2008, p. 125). En effet, le discours juridique se distingue par un usage particulier de la langue où la syntaxe et le lexique sont soigneusement choisis et souvent répété afin de maintenir une précision et une uniformité nécessaires en droit (Rey, 2008, p. 132).

Le discours numérique est également sujet au phénomène de répétition, et notamment aux phénomènes de reprises des sites web (Pecman, 2014), dans lequel certains textes sont repris à l'identique sur différents sites web ou différentes pages. Sur le site `gov.uk`, certaines recherches ont identifié des phénomènes de répétition ayant pour but la vulgarisation du discours.

> *in this perspective, repetition allows the mediator to come closer to their knowledge dissemination goal* (Cacchiani, 2018, p. 56)

Les sites étudiés présentent donc un phénomène de répétition qu'il convient d'étudier

en corpus, afin d'identifier son rôle dans la compréhension du discours. Ce phénomène de répétition est double. Tout d'abord, le discours administratif et juridique présente des répétitions liées à ses besoins de précisions et d'uniformité. Dans un second temps, le caractère numérique et donc délinéarisé du discours sous-tend un phénomène de reprise caractéristique des sites web.

3.3.6 Unités phraséologiques

La terminologie est une aide précieuse pour la compréhension d'un domaine, et notamment de ses principaux concepts. Gledhill et Kübler (2016) abordent les difficultés de compréhension d'une langue de spécialité dans l'enseignement supérieur. Ils suggèrent que la terminologie n'est pas le seul problème et que les étudiants rencontrent des problèmes avec la phraséologie, malgré leur connaissance de la terminologie. La connaissance de la terminologie ne suffit donc pas à maîtriser un domaine, et la phraséologie ne doit pas être négligée. Cette affirmation peut être étendue à toute personne souhaitant maîtriser une langue de spécialité. Si l'on considère les usagers comme des « apprenants » de la langue de l'administration, ils pourraient rencontrer les mêmes difficultés. Il est donc essentiel de ne pas délaisser la phraséologie, qui est tout aussi importante que la terminologie. La phraséologie est un champ d'études vaste, qui possède une multitude de champs d'application et de définitions. Cela peut être observé dans les différentes approches proposées. Pour notre analyse, nous retiendrons plusieurs définitions comme celle de Cowie (1994) reprise par Granger et Paquot (2008), qui définit la phraséologie comme « the study of the grammatical structure, meaning and use of word combinations ». Cette définition permet un champ d'études vaste : structure, sens et restriction combinatoire. Il existe de nombreuses autres définitions qui proposent une approche similaire. Plusieurs linguistes intègrent la dimension lexicale à leurs définitions comme Gries (2008, p. 5) qui définit un phraséologisme comme « the co-occurrence of a form or a lemma of a lexical item and any other kind of linguistic element ». Cela implique que l'unité étudiée possède au moins un élément lexical.

Il existe de nombreuses catégorisations des unités étudiées en phraséologie. Notre objectif n'est pas de réaliser une catégorisation des unités phraséologiques du discours, mais de proposer une caractérisation du discours. Nous tenterons donc de prendre en compte le plus d'unités possible, en définissant les unités phraséologiques comme des « patterns of co-occurrence » (Gries, 2008, p. 10). Adopter l'approche phraséologique en parallèle de l'approche terminologique nous permet d'apporter un regard sur les structures avec lesquelles s'emploient les termes de l'accès aux droits qui ont été peu étudiés. En effet, identifier et étudier les unités phraséologiques importantes pourrait permettre de mieux comprendre les rouages d'un tel discours.

Notre étude se basera également sur l'*idiom principle* de Sinclair.

> *The principle of idiom is that a language user has available to him a large number of semi-preconstructed phrases that constitute single choices.* (Sinclair, 1991, p. 110)

Les locuteurs d'une langue ont donc à leur disposition un nombre limité de choix lors de la formulation d'une phrase. Il semble important de comprendre les mécanismes de ces choix dans le discours de l'accès aux droits, qui a justement une visée de diffusion de l'information et de vulgarisation. Le locuteur qui ne connaît pas bien la langue de spécialité aura moins de choix qu'un locuteur initié. Ses lacunes peuvent non seulement poser des problèmes de compréhension, mais aussi des problèmes de rédaction pour s'adresser aux services publics. De plus, cette approche phraséologique nous semble particulièrement adaptée au type de discours étudié, le discours administratif, qui est basé sur des préfabriqués issus de plusieurs domaines connexes (juridique, politiques, etc.), comme l'indique Anderson :

> *(…) the administrative register relies to a large extent on prefabrication.* Anderson (2003, p. 58)

Il sera donc indispensable à la caractérisation linguistique d'étudier la phraséologie du discours de l'accès aux droits.

Nous allons maintenant présenter la méthodologie utilisée pour réaliser nos analyses.

Chapitre 4

Méthode

4.1 Choix du sujet

En guise d'introduction à ce chapitre sur les méthodes utilisées dans le cadre de cette thèse, et avant de présenter en détail les outils méthodologiques et les étapes qui ont jalonné notre étude, nous souhaitons expliquer l'origine de notre intérêt pour l'étude du discours administratif de l'accès aux droits, ainsi que la manière dont le choix des sources ayant servi à composer notre corpus s'est affirmé au fil du temps.

L'idée d'étudier le discours de l'accès aux droits numériques, et donc d'en constituer un corpus spécialisé, nous est venue lors de rencontres avec les usagers du centre social et culturel 13 Pour Tous dans le 13e arrondissement de Paris dans lequel nous intervenons depuis 2020 en tant qu'écrivain public. Avant d'exercer ce rôle, ce discours ne nous était pas inconnu car comme toute personne résidant en France, nous ne pouvions passer outre des démarches administratives. Toutefois, notre éducation ainsi que nos compétences linguistiques et numériques nous avaient jusque-là permis de réaliser les démarches administratives sans grandes difficultés. L'observation de ces usagers et des difficultés auxquelles ils devaient faire face ont éveillé notre curiosité quant à l'importance de l'étude de ce discours.

À partir de longues conversations avec eux, dont le but était de leur permettre de régler leurs problèmes administratifs et de réaliser certaines démarches, il est apparu clairement que le discours administratif était un élément central de la vie d'un grand nombre de personnes ; et que contrairement à la majorité des personnes vivant en France qui n'étaient confrontées au discours administratif que de façon plus ponctuelle, l'administration et son discours les suivraient toute leur vie, ou du moins jusqu'au moment où ils pourraient parvenir à un changement de leur situation administrative qui les rendraient moins dépendants de celle-ci.

Nous avons donc commencé à nous interroger sur le discours administratif et en par-

ticulier sur la transition numérique qui avait eu lieu ces dernières années et qui s'était intensifiée en 2020. Le discours administratif numérique constituait un sujet d'étude riche et central dans l'étude des langues de spécialité. L'idée d'étudier sa transition entre discours non numérique et discours numérique en comparant des discours pré-numérique aux discours numériques natifs fut brièvement abordée. Cependant, il était clair que la transition avait déjà eu lieu et qu'il était urgent de s'intéresser à ce discours tel qu'il existe aujourd'hui. En effet, au moment où nous avons commencé notre étude, en octobre 2020, peu d'études semblaient s'intéresser au discours de l'accès aux droits, aussi bien dans la communauté scientifique, que du côté des instances gouvernementales. Depuis le COSLA, dont les activités avaient cessé presque 15 ans auparavant, les campagnes de vulgarisation ou les informations concernant le discours administratif numérique étaient devenu rares et malgré quelques tentatives de le remettre sur le devant de la scène, que nous avons évoquée dans le chapitre 1, cela reste aujourd'hui un sujet timide[1]. Nous avons cependant remarqué une orientation forte du gouvernement français : la volonté d'assurer la diffusion d'informations par le biais de divers sites web. Dans le même temps, nous nous sommes intéressée à ce phénomène au Royaume-Uni, où nous avions également découvert le fonctionnement de l'administration et acquis de l'expérience lors d'un emploi dans l'assistanat administratif. Le paysage britannique, très similaire, est pourtant le fruit d'une évolution et d'une stratégie discursive complètement différente évoqué dans le chapitre 1, section 1.1. Il nous a ainsi paru être un terrain de comparaison intéressant car il a pris néanmoins une forme très similaire au discours administratif numérique en France, tout en émanant d'une idéologie et d'une histoire différente. De plus, l'approche comparée nous permettait de mettre en avant les caractéristiques relatives non seulement à la langue de spécialité, mais aussi à la traduction. Nous avons donc choisi de constituer un corpus spécialisé bilingue comparable.

Nous avons ensuite réalisé une étude du paysage documentaire britannique et français, présentée dans la section 1.3. Cette étude nous a amenée à considérer le site web comme terrain principal de diffusion de l'information administrative. En effet, nous nous sommes rendu compte que les sites web étaient devenus le lieu principal de communication entre l'administration et ses citoyens. Pour obtenir des informations administratives, les sites web officiels des gouvernements ou des organisations de service public sont les lieux les plus prolifiques et fiables. Nous avons donc choisi les sites web administratifs proposant des informations pour l'accès aux droits.

Il existe un grand nombre de sites web relayant de l'information administrative, notamment des sites publics et privés. Toutefois, l'ensemble des sites web privés, à l'instar

1. Les administrations semblent reprendre régulièrement le sujet en main pour réaliser des campagnes d'information ou de simplification, mais aucune de ces campagnes n'a pour l'instant mené à des tranformations significatives.

du site `mes-allocs.fr`, recopie et retravaille les documents mis à disposition sur les sites gouvernementaux officiels qui sont libres de droits. De plus, ils ne reflètent pas les politiques discursives du gouvernement. Nous avons donc choisi d'étudier les sites web officiels des gouvernements britanniques et français, respectivement `gov.uk` et `service-public.fr`. L'histoire de l'évolution de ces sites est retracée dans le chapitre 1 de cette thèse.

Dans ce chapitre, nous allons présenter la méthodologie mise en place pour réaliser notre étude. Nous présenterons dans un premier temps le protocole de création d'un corpus spécialisé comparable, puis nous introduirons le protocole de création d'un second corpus composé de parcours utilisateur. Enfin, nous présenterons les différents outils et démarches utilisés dans la réalisation de ces travaux de recherche.

4.2 Linguistique de corpus

Pour notre étude du discours numérique de l'accès aux droits, nous utilisons plusieurs corpus ainsi que des logiciels qui permettent de les exploiter. Nous nous positionnons donc en tant qu'analyste de discours, comme nous l'avons expliqué dans le chapitre 2, activité que nous pratiquons avec l'aide de la linguistique de corpus.

4.2.1 Essor de la linguistique de corpus : ouverture des possibilités

La linguistique de corpus fait son apparition dans les années 1960. Elle est développée initialement pour répondre au besoin d'enseigner l'anglais comme langue étrangère. Dès ses débuts, il s'agit d'une discipline empirique puisqu'elle est d'abord développée comme une méthodologie d'observation du langage (Teubert, 2009). Le numérique occupe non seulement une place importante dans le développement de la linguistique de corpus, mais l'essor et les progrès de la linguistique de corpus sont aussi étroitement liés au progrès de l'informatique et du numérique. La discipline se développe ainsi rapidement dans l'ère du web 1.0 entre 1989 et 2005. Pour les chercheurs et chercheuses étudiant des textes, le développement de l'informatique représente un véritable tournant :

> « Although manualanalysis has made an important contribution over the centuries, especially in lexicography, it was the availability of digital computers from the middle of the 20th century which brought about a radical change in text-based scholarship » (Kennedy, 2014, p. 5)

Deux étapes cruciales permettent l'évolution de cette discipline : l'invention du scanner et des logiciels OCR qui permettent de rendre un grand nombre de textes (autrefois disponibles sur support papier uniquement) disponibles en ligne au format numérique,

puis plus tard les débuts de l'écriture numérique (Bonelli, 2010). L'ouvrage canonique de Sinclair (1991) intitulé *Corpus, Concordance, Collocation* met en avant le rôle important qu'a joué l'informatique dans le développement de la linguistique de corpus et de ses principes. Les possibilités offertes par l'informatique permettent une démocratisation de l'utilisation des corpus dans plusieurs domaines.

> *Over the last three decades the compilation and analysis of corpora stored in computerized databases has led to a new scholarly enterprise known as corpus linguistics.* (Kennedy, 2014, p. 1)

La chercheuse Elena Bonelli va même plus loin en affirmant que ce n'est pas le contexte linguistique qui a mené au développement des corpus, mais bien le contexte technologique[2] (Bonelli, 2010, p. 15). L'apparition de la linguistique de corpus est donc intrinsèquement liée au numérique et à l'informatique en général.

Depuis la publication de l'ouvrage de Sinclair (1991) dans les années 1990, mais aussi depuis la généralisation de l'utilisation du numérique dans la recherche, les possibilités offertes par le numérique n'ont fait que s'accroître. La linguistique de corpus a bénéficié de nombreuses innovations. Pour Sinclair et ses contemporains, la possibilité d'accéder à des corpus de grande taille et de faire des recherches systématiques avait révolutionné la façon dont le langage pouvait être envisagé. La méthode *bottom-up*, qui est une démarche inductive dans laquelle les hypothèses et observations émergent de la lecture et de l'analyse du corpus (Azzopardi, 2010, p. 5), permettait d'étudier le langage sous un nouvel angle.

Aujourd'hui, les capacités informatiques ont largement dépassé nos attentes et le développement des algorithmes révolutionne de façon encore plus radicale les sciences du langage. La taille des corpus ne cesse d'augmenter, mais des problématiques nouvelles liées à leur exploitation sont apparues également. Bien que les linguistes disposent désormais des ressources de taille gigantesque, la situation est loin d'être simple ou facilitée pour eux. Certes les corpus sont plus nombreux et de taille plus importante, mais les nouveaux discours ne cessent d'émerger et les linguistes sont régulièrement amenés à devoir construire pour les besoins de leurs recherches leur propre corpus ou des corpus ad hoc. Par ailleurs, les textes numériques prolifèrent, mais leur format ne les rend pas tous facilement transformables en format explorable et exploitable par les outils d'analyse de la linguistique de corpus. Ces transformations peuvent être coûteuses en temps, bien que les applications de plus en plus performantes aient permis d'automatiser cette tâche. Les restrictions dues aux droits d'auteur rendent les textes compilés en corpus interdits de diffusion, freinant le partage des corpus entre linguistes.

2. Citation originale : "It was not the linguistic climate but the technological one that stimulated the development of corpora." (Bonelli, 2010, p. 15)

Un des tout premiers corpus de grande taille connus est le *Brown corpus* compilé dans les années 60, pour l'anglais, avec approximativement 1 million de mots. Le *British National Corpus*, un corpus de référence constitué sur plusieurs années entre 1980 et 1990, atteint une taille de 100 millions de mots grâce à une collecte de textes issus de différents genres. Un corpus plus récent, le *English Web Corpus (enTenTen)* (Jakubíček et al., 2013), collecté entre 2019 et 2021, rassemble, lui, 36 milliards de mots. C'est une différence de taille considérable, qui montre l'étendue des avancées technologiques de ces 50 (et notamment 30) dernières années. La collecte de ces corpus sur un temps relativement court est possible car de nombreux processus sont désormais automatisés, et que les données sont plus accessibles et exploitables.

De nos jours, les corpus sont un outil incontournable dans plusieurs domaines. Comme nous l'avons dit, les outils numériques permettent de récolter des corpus de grande taille. Ces données sont récoltables facilement en ligne grâce à divers outils. Elles peuvent ensuite être préparées et compilées en fichier afin de réaliser des analyses. Le numérique permet de les explorer de façon systématique, ce qui est également une grande avancée car des logiciels permettent de faire en quelques secondes des analyses qui prenaient parfois plusieurs jours. Les corpus sont donc des objets d'étude numériques créés et compilés afin de faire des analyses linguistiques. En 2024, la linguistique de corpus est incontournable dans l'ensemble des sciences du langage. Néanmoins, comme nous l'avons souligné, de nouvelles problématiques émergent, concernant notamment la confidentialité des informations et la démarche globale vers le partage des corpus et des données. En outre, d'autres questions se posent concernant les systèmes de traduction automatique neuronale (TAN) et les grands modèles linguistiques (Large Language Models (LLM)) avec des systèmes d'apprentissage neuronaux s'appuyant sur les textes et données. Cependant, l'avènement de la linguistique de corpus, dans sa multiplicité de pratiques, est bien installé.

4.2.2 Approche adoptée

Il existe plusieurs approches pour réaliser une analyse de corpus. La première, nommée l'approche *top-down*, terme qui décrit une approche descendante, consiste à établir un cadre d'analyse avant de commencer l'exploration du corpus. Dans l'approche *top-down*, le chercheur arrive avec une théorie qu'il applique à ses données. La seconde approche consiste à observer le corpus et à fonder le cadre d'analyse sur les données observées. Sa dénomination, *bottom-up*, décrit une approche inductive dans laquelle le chercheur se laisse guider par le corpus. Comme l'explique Azzopardi (2010), la différence entre ces deux approches vient de la place donnée au corpus. Dans l'approche descendante, l'hypothèse et les analyses sont faites, le corpus existe pour confirmer ou

infirmer ces hypothèses et analyses. Pour Mayaffre, le corpus peut-être l'« observatoire d'une théorie » dans le cas de l'approche *top-down* ou encore « un observé dynamique qui permet de décrire et d'élaborer des modèles » (Mayaffre, 2005, p. 6) dans le cas de l'approche *bottom-up*.

Dans notre étude, nous utilisons une approche majoritairement *bottom-up*. Nous avons bien sûr un cadre théorique et méthodologique, qui est celui de l'analyse du discours numérique et des langues de spécialité. Toutefois, l'ensemble des analyses réalisées ont été faites de façon inductive. Nous avons d'abord observé le corpus, puis nous l'avons exploré de façon systématique grâce aux logiciels d'exploration de corpus UAM Corpus Tools et Sketch Engine, que nous présenterons dans la section 4.6.1.

Notre méthode de recherche s'approche de ce fait de celle de Millot (2019), à ceci près que nous n'adoptons pas une approche diachronique, mais une approche synchronique, tout en prenant en compte que l'extraction des données des sites web a été effectuée à un moment précis, mais que le discours contenu sur ces sites a été rédigé sur plusieurs années, si bien que notre analyse des données s'effectue en partie en diachronie courte (Picton, 2018). Pour aboutir à une analyse avec des résultats probants, Millot met en avant que la recherche en langue de spécialité a intérêt non seulement à se faire en deux temps :

> La méthode de caractérisation repose tout d'abord sur une approche quantitative fondée sur le concept de mot-clé que nous utilisons dans une perspective diachronique. Elle repose ensuite sur une analyse qualitative à travers l'étude de l'histoire du domaine et l'évolution des définitions de termes ayant survécu à l'évolution.

Nous avons donc pris en compte, auprès des dimensions linguistiques et discursives, la dimension sociohistorique. Pour l'étude du genre spécialisé et de la langue de spécialité numérique de l'accès aux droits, cette méthode s'est avérée indispensable.

Nous avons créé deux corpus, un corpus que nous appelons **corpus classique**, présenté dans la section 4.3.1 et un corpus appelé **corpus pilote** ou **corpus parcours**, présenté dans la section 4.5, qui a été collecté et élaboré après une analyse pilote du corpus classique. Pour le **corpus classique**, nous utilisons un système d'annotations automatiques par catégories syntaxiques proposé par le logiciel Sketch Engine. Ce sont des cadres d'analyse et des segmentations que nous avons apportés à nos données avant leur analyse. Pour le **corpus parcours**, certains cadres d'analyse ont été créés spécifiquement, après une analyse pilote des données afin de prendre en compte les particularités des corpus numériques. Nous avons notamment créé et mise en œuvre une méthode d'annotation manuelle de ce parcours présenté dans la section 4.5.3.

4.2.3 Définition d'un corpus

D'après Sinclair, cité par Habert (2000, p. 1), un corpus est « une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques et extralinguistiques explicites pour servir d'échantillon d'emplois déterminés d'une langue ». De nos jours, les corpus sont stockés et utilisés sous forme numérique. Ce sont des données collectées pour servir d'échantillon. Il est donc nécessaire que les textes choisis soient représentatifs du discours étudié. La taille du corpus est notamment un élément important, qui dépend des critères d'analyse que nous définirons dans la section 4.3.1. Les corpus sont également des textes exploitables par une machine (McEnery et Wilson, 2001, p. 32)[3]. Ils sont préparés et encodés afin d'être lus par des logiciels.

Les objectifs et caractéristiques d'un corpus sont différents de celui d'un texte. Dans la Figure 4.1, Bonelli (2010) illustre la différence entre un texte et un corpus.

Table 2.1 A qualitative comparison of a text versus a corpus

A text	*A corpus*
read whole	read fragmented
read horizontally	read vertically
read for content	read for formal patterning
read as a unique event	read for repeated events
read as as individual act of will	read as a sample of social practice
instance of *parole*	gives insights into *langue*
coherent communicative event	not a coherent communicative event

FIGURE 4.1 – Tableau montrant les différences entre texte et corpus (Bonelli, 2010)

La différence principale entre un corpus et un texte est qu'un corpus est exploré de façon non linéaire et verticale, tandis qu'un texte est lu de façon linéaire et horizontale. Un texte est également un élément cohérent, qui forme un tout.

Cependant, même si texte et corpus sont épistémologiquement différents, ils forment une continuité. En effet, Gatto (2014) définit les corpus comme une série de textes de grande taille, ce qui ne permettrait pas une exploration manuelle. Un corpus se différencie d'un texte par sa taille et ses objectifs de recherche, comme l'illustre cette définition également :

> *(...) a corpus designed for linguistic analysis is normally a systematic, planned and structured compilation of text. » (Kennedy, 2014, p. 4).*

Ces deux chercheurs indiquent qu'un corpus est formé de plusieurs textes, qui, une fois rassemblés, forment un ensemble de données langagières qui est explorable de façon non linéaire. Un travail de préparation est donc nécessaire pour faire des analyses en corpus. Tout d'abord, le corpus est récolté, puis nettoyé. Un grand nombre d'informations et

3. Citation originale : "machine-readable text" (McEnery et Wilson, 2001, p. 32)

de textes jugés comme étant du bruit sont enlevés. La mise en forme est également effacée. En effet, comme l'explique Schäfer et al. (2013), dans les corpus issus du web, il est parfois difficile de distinguer le texte utile au linguiste du bruit du corpus. Cela vient de l'encodage et du format original du fichier, qui intègre des balises et du code d'information. Pour obtenir un corpus exploitable, l'encodage et le format des fichiers originaux sont modifiés et nettoyés[4] . Les textes peuvent ensuite être annotés (manuellement ou automatiquement) avant d'être exploités par un logiciel.

Les données collectées dans le cadre d'un corpus « sont sélectionnées et organisées selon des critères linguistiques et extralinguistiques explicites pour servir d'échantillon d'emplois déterminés d'une langue » (Sinclair, repris par Habert (2000)). Il est donc nécessaire que les textes choisis soient représentatifs du domaine étudié. Pour cela, il convient d'établir des critères de sélection clairs, afin de proposer une étude adaptée (Biber, 1993). Il est donc essentiel de préparer la création d'un corpus en définissant les critères de sélection des textes en fonction des objectifs scientifiques de la recherche, avant la récolte des données. Un protocole doit ensuite être mis en place afin de récolter les données et de les préparer. Enfin, suivant l'approche choisie (*top-down* ou *bottom-up*), le cadre d'analyse doit être rédigé et implémenté.

Nous pensons qu'il est important de mettre en avant la place du numérique dans la définition d'un corpus. En effet, comme nous l'avons montré dans la section 4.2.1, les corpus tels qu'ils existent aujourd'hui sont intrinsèquement liés avec le développement des outils informatiques et numériques. Nous avons donc élaboré la définition suivante du corpus :

> Le corpus est un objet numérique, modifié et préparé pour être exploité par une machine. (Bouchet, 2023)

Cette définition permet de mettre en avant le caractère numérique du corpus. Elle montre également la place qu'occupe la préparation des données dans l'élaboration d'un corpus.

4.2.4 Une méthode adaptée à l'étude des langues de spécialité

Nous avons présenté dans les chapitres 2 et 3 la définition des langues de spécialité, ainsi que la position du discours étudié par rapport aux langues de spécialité. Pour rappel, nous étudions un discours spécialisé, qui regroupe plusieurs langues de spécialité,

4. Bien que ces suppressions soient nécessaires, elles ont été remises en cause dans certains cas puisque la structure des documents se trouve encodée dans la mise en forme, empêchant ainsi de prendre en compte la dimension textuelle et la structure interne des textes dont certaines analyses ne peuvent se passer (comme par exemple celle de la variation terminologique rhétorique au sein de l'unité textuelle, e.g. Pecman 2014).

notamment les domaines spécialisés du droit et de l'administration. Notre méthodologie doit donc être adaptée au type de discours étudié. De nos jours, les corpus sont incontournables en sciences du langage, et particulièrement dans l'étude des discours spécialisés. Les discours spécialisés font état de la communication entre experts ou entre experts et profanes, dans l'objectif de transmettre des connaissances spécialisées (Lerat, 1995, p. 20). Les corpus généraux, portant sur une langue dans son ensemble, ont pour but de décrire le système du langage tandis que les corpus spécialisés s'intéressent à un domaine en particulier. La langue de spécialité, comme la langue générale, se réalisant dans un acte de communication, l'observation de son utilisation (dans les ressources disponibles collectées en corpus) est le seul moyen de l'étudier. Il est ainsi possible d'analyser une interaction ou un texte en particulier afin d'étudier, par exemple, les moyens utilisés pour transmettre une information ou la cohérence d'un texte.

Dans l'objectif où nous cherchons à caractériser un discours, il semble alors pertinent de s'intéresser au discours dans son ensemble. Les corpus spécialisés ont été des outils précieux qui ont permis de décrire un grand nombre de langues de spécialité, et aidé à la création de dictionnaires spécialisés (Bowker, 2010). Lorsqu'ils sont de grande taille, ils permettent de réaliser des analyses terminologiques et phraséologiques (O'Keeffe, 2010).

Les corpus spécialisés sont adaptés à l'étude des langues de spécialité, notamment parce qu'ils permettent d'étudier des données qualitativement, en prenant en compte leur contexte de production (Flowerdew, 2010). Les corpus de petite taille en langue de spécialité sont également très précieux. Certaines catégories grammaticales, comme les pronoms, peuvent être étudiées dans un petit corpus, car ils sont plus fréquents. De plus, il est plus facile de lier les données langagières à leur contexte et de réaliser ainsi une analyse du discours située dans le texte.

La linguistique de corpus est donc une méthode adaptée à l'étude des langues de spécialité car elle permet de les étudier dans leur globalité, mais aussi en contexte. Les modalités de construction du corpus doivent être définies par le chercheur selon ses objectifs de recherche. Il est donc nécessaire d'avoir une bonne compréhension du domaine et du discours étudié afin de poser un cadre d'analyse adapté.

4.2.5 Les corpus issus du web dynamique

Le corpus est numérique par essence, mais depuis quelques années, on voit apparaître un nouveau type de corpus : les corpus compilés à partir de textes numériques natifs (Paveau, 2017). Il ne s'agit plus d'étudier des textes numérisés puis mis en ligne, mais bien des textes qui ont été écrits directement en ligne ou sur un ordinateur connecté à internet dans un environnement dynamique. Les corpus numériques prennent donc une

nouvelle dimension car ils sont désormais doublement numériques : par leur source, puis par leur encodage.

Au début des années 2000, on retrouve beaucoup d'interrogations sur le web considéré en tant que corpus. Certains chercheurs considèrent que le web ne peut être un corpus, à l'image de Sinclair (2005a) qui indique qu'Internet ne constitue pas un corpus car il n'a pas été créé dans une perspective linguistique (cité par Gatto 2014, p. 39). Mais depuis, l'utilisation du web dans toutes les sphères de la vie, et particulièrement dans la communication, s'est imposée. Le web est alors devenu un terrain d'étude pour les chercheurs en sciences du langage, que ce soit pour collecter des corpus dits conventionnels (composées de données sélectionnées et préparées) ou pour utiliser le web dans son ensemble en tant que corpus (Gatto, 2014). On retrouve deux types d'utilisation du web dans les corpus : le « Web as Corpus », qui consiste à considérer le web en tant que corpus dans son ensemble, en l'explorant à l'aide de moteur de recherche par exemple, et le « Web for Corpus », qui consiste à récupérer des données langagières sur le web pour former des corpus (Fairon, 2000). Dans le cadre de cette thèse, nous nous situons plus particulièrement dans la démarche « Web for Corpus » car nous avons collecté un corpus à partir du web.

La création de corpus issus de textes disponibles sur le web demande un certain nombre d'étapes de transformation. L'utilisation de données de sites web en ligne pose tout d'abord la question de la récupération des données. En effet, les textes sont généralement mis en forme par du code, ou parfois par plusieurs couches de code. Nous proposons ici un protocole générique d'exploitation des textes numériques natifs inspiré des protocoles pour l'étude des discours numériques natifs existants et mis à disposition par les différentes équipes de recherches s'étant intéressées à ces types de textes (Gatto, 2014; Baroni et Ueyama, 2006; Fairon, 2000; Cheok et al., 2022; Jakubíček et al., 2013) :

1. Identification des pages à collecter dans le corpus à l'aide des URLs.
2. Collecte automatique, manuelle, ou semi-automatique des textes. La collecte automatique est effectuée par un logiciel s'appelant un web crawler, qui se déplace de page en page. Ce logiciel permet parfois d'éliminer la première étape d'identification des pages, car il « choisit » les pages à retenir, à partir de critères définis par les chercheurs.
3. Nettoyage des textes, notamment suppression des éléments de codes. Cette étape est réalisée automatiquement. Il existe un certain nombre de méthodes et de logiciels qui sont généralement adaptés aux objectifs de l'étude.
4. Identification des éléments pertinents. Il s'agit d'un enjeu important des textes récupérés automatiquement sur le web. En effet, une grande partie des textes récupérés en ligne sont en fait du bruit. Il s'agit d'identifier les éléments utiles à l'ana-

lyse des données. Pour cela, les chercheurs utilisent différents outils. HTMLCleaner (Girardi, 2007) et jusText (Pomikálek, 2011) sont deux outils largement utilisés dans la communauté. Ils permettent d'éliminer les textes ne formant pas de phrases, et donc d'éliminer les menus et autres éléments de navigation, afin de ne garder que le contenu des pages. Ces outils utilisent des règles, ce qui leur permet de rejeter « les textes qui ne possèdent pas une grande proportion de tokens correspondant à des mots grammaticaux » (Jakubíček et al., 2013, p. 1)[5]. Il est également possible de se baser sur les balises HTML afin de ne sélectionner que le texte utile, mais cela n'est possible que sur des données possédant une structure HTML similaire. Cela concerne généralement les données issues d'un même site ou domaine.

5. Déduplication des textes. Il existe sur le web un phénomène de redondance, ce qui signifie que les textes sont parfois multipliés sur plusieurs sites ou plusieurs pages. Il s'agit d'un effet de reprise communément observable sur les sites web (Pecman, 2018). Cela cause bien sûr des problèmes dans l'analyse du corpus. Afin de pallier ce problème, il est parfois nécessaire d'utiliser un algorithme permettant d'effacer ces textes redondants. L'outil *Onion* (Pomikálek, 2011) analyse par exemple les textes par paragraphe et efface les textes en double. Cela est très utile dans la construction de corpus portant sur la langue générale, comme le corpus TenTen.

Certaines étapes de ce protocole générique peuvent être interchangées ou omises complètement suivant les objectifs de la recherche. Nous avons jugé que la déduplication n'est pas une étape qui est toujours pertinente pour les corpus spécialisés, et pour notre étude notamment. En effet, les textes utilisés ont été soigneusement sélectionnés, et les redondances présentes dans le texte ne font pas état d'un phénomène de copiage des contenus caractéristiques d'internet, mais plutôt d'une stratégie de communication adoptée par l'éditeur des sites étudiés. Il est donc important d'adapter les étapes de récupération des données décrites plus haut au corpus étudié.

Plusieurs chercheurs recommandent également de séparer les sections de textes « non-informatives » et « paratextuels » (Gatto, 2014). Pour nous, il s'agit d'une pratique contestable dans certains cas. En effet, il est parfois erroné de définir les menus, les en-têtes ou les pieds de page comme « non-informatifs ». Ils font désormais partie intégrante de notre expérience sur internet et peuvent être porteurs d'informations comme le montrent plusieurs études (Perea, 2012; Pinède et al., 2011).

Ce protocole est donc précieux pour l'étude des discours numériques natifs, mais il

5. Citation originale : "material that does not have a high proportion of tokens that are the grammar words of the language"

est parfois incomplet. Nous estimons qu'il se concentre sur la conformité des données pour les outils existants, en passant les textes originaux à travers diverses étapes de préparations, sans se demander si certains éléments constitutifs de ce discours ne sont pas perdus lors de cette mise en conformité des données.

Les corpus numériques portent notamment sur des objets nouveaux et complexes. De nombreuses recherches s'intéressent aux interactions entre internautes en ligne, qu'on appelle la communication médiée par ordinateur et qui est définie comme « le processus par lequel les utilisateurs créent, échangent et comprennent l'information via l'utilisation de systèmes de télécommunication en réseau permettant l'encodage, la transmission et le décodage de messages » (Lombart et Fairon, 2020, p. 1). Ces corpus collectés sous diverses formes sont issus des supports qui ont des caractéristiques différentes : blog, forums, SMS, messagerie instantanée et même messagerie multimodale à l'instar de Snapchat, où l'usager envoie une messagerie éphémère sous forme de vidéo, couplée d'un texte ou d'un message audio, ou encore de musique.

Ces corpus peuvent donc être considérés comme complexes. Ils sont parfois collectés et analysés pour leur particularité, à l'image de l'étude des Tweets dans leur environnement natif (Paveau, 2013), de l'étude des posts Facebook (Bibié-Emerit, 2016) ou encore des multiples études concernant les forums (par exemple Eshkol-Taravella et Grabar (2016); Lombart et Fairon (2020). Ces données langagières sont interrogées en conservant les caractéristiques des plateformes desquelles elles proviennent, en complétant l'analyse du corpus avec une analyse du contexte. Certains chercheurs s'intéressent aux communications médiées par ordinateur dans leur ensemble et collectent des corpus aux sources variées, comme le projet Communication Médiée par les Réseaux CoMeRe (Wigham et Poudat, 2020). Les données de ce type de corpus doivent être standardisées afin de pouvoir être comparées aux autres. Notre recherche s'inscrit dans ces deux approches : l'étude des discours pour leurs spécificités et dans leur ensemble.

Les terrains numériques sont vastes, et les approches dans l'élaboration et l'analyse des corpus numériques sont multiples. Nous ne saurions donc couvrir l'ensemble des problématiques et recherches liées aux corpus numériques. Cependant, nous pouvons observer un intérêt grandissant dans les communautés scientifiques pour les corpus numériques et notre étude s'inscrit dans ce mouvement. La communication numérique, quelle que soit sa forme, est désormais incontournable et nous souhaitons développer un corpus et un cadre d'analyse adapté à notre objet, qui le décrit dans sa complexité et dans sa multimodalité.

4.3 Création d'un corpus spécialisé

4.3.1 Définition des critères pour la constitution du corpus

Pour rappel, nous avons défini le corpus de la façon suivante : « une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques et extralinguistiques explicites pour servir d'échantillon d'emplois déterminés d'une langue » Sinclair, cité par Habert (2000). À cette définition, nous avons ajouté un élément qui le caractérise intrinsèquement : sa nature, qui est numérique (Bouchet, 2023), discuté en 4.2. En effet, de nos jours, les corpus sont principalement utilisés sous forme numérique.

Afin de créer notre corpus, nous avons ainsi mis en place des critères de sélection (Biber, 1993). Bien que dans les études sur la méthodologie de création de corpus, la nécessité d'avoir des critères prédéfinis en fonction des objectifs scientifiques soit mise en évidence, il est plus rarement précisé que la création d'un corpus est aussi une démarche cyclique. En effet, l'ensemble des paramètres ne peuvent être définis préalablement. Nous avons donc créé un corpus pilote. Cette étude pilote nous a permis de sélectionner des critères adaptés à notre étude et à nos objectifs.

Notre travail de recherche a pour but de décrire un discours, tel qu'il est reçu par les différents acteurs du domaine, et notamment le public cible qui est l'usager. Nous nous sommes appuyée sur la phrase suivante pour choisir nos matériaux :

> *The linguist aims to describe language use rather than identify linguistic universals.* (Bonelli, 2010, p. 15)

C'est ainsi que nous avons décidé d'étudier les textes tels qu'ils sont le plus souvent consultés par les usagers. Nous avons donc orienté nos recherches vers les sites web publiés par les instances gouvernementales officielles. Ce critère est lié à notre objectif pour cette recherche : réaliser avec ce corpus des explorations terminologiques, phraséologiques et discursives afin de caractériser le discours de l'accès aux droits.

Afin que le discours soit représentatif d'un discours gouvernemental officiel, nous avons choisi de n'inclure que des sites publiés par des organisations officiellement instituées de missions de services publics. Cela inclut des organisations qui font partie du gouvernement ou qui sont investies d'une mission de service public. Sur ces sites, nous avons sélectionné des thèmes relatifs à l'accès aux droits qui sont les suivants :

— Aides sociales
— Droit au séjour
— Citoyenneté

Sur chacun des sites sélectionnés, que nous allons présenter dans la section suivante, nous avons retenu le contenu des pages web. En effet, il s'agit de la partie de la page où se

trouve le discours que nous souhaitons étudier. Nous aborderons dans la section 4.5.1 la conception d'un second corpus qui nous a permis d'étudier d'autres parties du discours. La sélection des trois thèmes mentionnés ci-dessous a de surcroît permis d'instaurer une hiérarchisation de notre corpus, se traduisant par la création de trois sous-corpus distincts, chacun étant correspondant à l'un des thèmes. Pour chaque thème, plusieurs sous-corpus ont été élaborés, chacun d'eux provenant d'un site web distinct.

La taille du corpus dépend également de la finalité de l'étude et le critère supplémentaire que nous avons respecté est sa représentativité du discours étudié : notre corpus constitue ainsi un échantillon représentatif de la totalité du discours étudié. Il a été donc nécessaire de définir ses limites, mais aussi son organisation hiérarchique (Biber, 1993).

Pour résumé, notre corpus provient principalement de sites web gouvernementaux (ou d'organismes officiels institués d'une mission de service public) et il se compose de la partie informative, également connue sous le nom de contenu, extraite de diverses pages de ces sites. Cette sélection de contenu nous a permis de mener des analyses terminologiques, phraséologiques et discursives, tandis qu'un autre corpus a été utilisé pour examiner des aspects spécifiques du discours numérique que nous étudions.

4.3.2 Choix des sites

Pour le corpus britannique, nous avons choisi le site officiel de l'administration britannique accessible sous le nom de `gov.uk` qui rassemble l'ensemble des informations administratives mises à disposition des citoyens par l'administration britannique. Il est décrit comme étant « le meilleur endroit pour trouver des informations et des services gouvernementaux »[6] sur la page d'accueil du site. Ceci en fait donc la principale source pour accéder au type de discours que nous souhaitions analyser. De plus, il est possible de réaliser un grand nombre de démarches directement sur le site `gov.uk`. Nous l'avons donc choisi comme première source de collecte de corpus.

Les mêmes recherches et conclusions ont été faites pour le site officiel de l'administration française, accessible sous le nom de domaine `service-public.fr`. Il est décrit comme étant « le site officiel d'information et de démarches administratives ». Il est également possible de faire certaines démarches à partir du site web. Ce site rassemble les informations sur presque toutes les démarches administratives disponibles. À noter que ce site présente également un volet « professionnel » réservé aux démarches administratives professionnelles, que nous avons écarté car ils ne font pas partie de notre champ d'étude.

Ces deux sites web rassemblent des informations sur un très grand nombre de sujets. En prenant l'intégralité du site, nous aurions récolté un corpus bien trop conséquent pour

6. Traduction du texte original : The best place to find government services and information

effectuer des analyses qualitatives pertinentes. Nous avons donc choisi une thématique administrative, afin de restreindre le corpus.

En plus de cela, nous avons sélectionné des sites pertinents pour chaque thème de notre corpus, que nous allons à présent introduire.

4.3.2.1 Thématique aide sociale

Pour notre mémoire de master, nous avions choisi de récolter un corpus sur le thème de l'aide sociale (ou *benefits* en anglais). C'est à partir de ce corpus récolté pour le mémoire de master que nous avons constitué le sous-corpus correspondant à la thématique aide sociale. Une fois le sujet choisi, nous avons réalisé que plusieurs organismes responsables de ces aides sociales diffusaient également des informations sur le sujet. Nous avons donc fait une sélection des organismes correspondant aux critères suivants :

— Organismes responsables d'un service public publiant des informations dans le but d'aider les usagers à accéder à leurs droits.

Pour le corpus français sur la thématique aide sociale, nous avons donc sélectionné les sites suivants :

— Le site officiel de l'administration française (`service-public.fr`), qui présente l'ensemble des démarches sur le sujet au niveau national.
— Le site officiel de la Caisse des allocations familiales (`caf.fr`), sur lequel des informations sur les aides dont la Caf est responsable sont disponibles. À noter que la Caf est un organisme privé responsable d'un service public.
— Le site officiel du Crous de Paris, qui propose des informations sur les aides disponibles pour les étudiant.es.
— Le site officiel de Pôle emploi[7], sur lequel on trouve des informations sur les aides dont cet organisme est responsable. À noter que Pôle emploi est un organisme privé responsable d'un service public.
— Le site officiel de la Ville de Paris, qui propose des informations sur les aides disponibles pour les habitants de la ville de Paris.

Le corpus français sur le thème des aides sociales (CORPUS AIDESOC-FR) est donc composé de quatre sites web. Ces sites ne sont pas les seuls à proposer des informations administratives officielles, mais ils sont représentatifs des différents types de sites disponibles : un site gouvernemental d'accès aux droits (`service-public.fr`), un site local (`paris.fr`), un site d'un organisme privé responsable d'un service public (`caf.fr`) et un site qui s'adresse à un certain type d'usager (`crous.fr`).

7. Depuis le 1er janvier 2024, Pôle emploi est devenu France Travail. Toutefois, ce corpus ayant été recolté avant ce changement d'appellation, nous continuons d'utiliser l'appellation Pôle emploi dans l'ensemble de notre étude.

Pour le corpus britannique, nous avons procédé de façon similaire afin de sélectionner des sites qui représentent le paysage documentaire du discours administratif numérique concernant les aides sociales. Nous avons donc sélectionné les sites suivants :

— le site officiel de l'administration britannique (`gov.uk`), qui présente l'ensemble des démarches sur le sujet au niveau national, mais sur lequel il est également possible de réaliser ces démarches.
— le site `understanduniversalcredit.gov.uk` qui est réservé aux usagers bénéficiaires ou souhaitant bénéficier de l'*Universal Credit*, une aide sociale transversale qui a remplacé un grand nombre de dispositifs auparavant indépendants.
— les sites des municipalités de Bornemouth, Bristol et Falkirk. Les sites des municipalités proposent des informations administratives concernant les aides sociales locales, mais aussi nationales.
— Le site de l'organisme Citizen Advice (`citizensadvice.org.uk`). Même s'il ne s'agit pas d'un organisme directement sous la tutelle du gouvernement, il est largement financé par l'État et réalise une mission de service public. De plus, au Royaume-Uni, l'accès aux droits passe en grande partie par les associations et le bénévolat.

Le corpus anglais sur le thème des aides sociales (CORPUS BEN-AN) est donc composé de six sites web, deux sites nationaux, trois sites locaux et un site d'un organisme indépendant.

4.3.2.2 Thématique droit au séjour

Afin d'étendre notre corpus, nous avons choisi lors de notre première année de thèse d'autres thématiques administratives. La seconde thématique choisie concerne les démarches administratives des étrangers, qu'ils soient résidents ou non en France et notamment le droit au séjour. En effet, cette thématique nous a paru essentielle car les étrangers sont souvent les plus fréquemment confrontés au discours administratif. Nous avons donc fait une sélection des organismes correspondant aux critères suivants : sites présentant des informations nécessaires à la réalisation de démarches administratives réservées aux personnes n'ayant pas la nationalité française et publiée par des instances officielles. Pour le corpus français, les sites suivants ont été sélectionnés :

— Le site officiel de l'administration française (`service-public.fr`), qui présente l'ensemble des démarches sur le sujet au niveau national.
— Le site *Welcome To France* (`welcometofrance.com`), édité par l'agence publique Business France, qui est un « Service d'information à destination des talents étrangers et de leur famille, en mobilité professionnelle en France ».

— Le site de la préfecture de Paris, qui propose des informations concernant les démarches à réaliser pour l'obtention des diverses autorisations de séjour.

Le corpus français sur le thème du droit au séjour (CORPUS ETR-FR) est donc composé de trois sites web.

Ces deux sites ne sont pas les seuls à présenter des informations concernant les démarches administratives réservées aux étrangers. Cependant, ils sont représentatifs des sites disponibles, à savoir trois sites édités par trois organismes publics : une agence, la direction de l'information légale et administrative et une préfecture. Le corpus français sur le thème du droit au séjour est donc composé de 3 sites web. Pour le corpus britannique, le site suivant a été sélectionné :

— Le site officiel de l'administration britannique (`gov.uk`), qui présente l'ensemble des démarches sur le sujet au niveau national, mais sur lequel il est également possible de réaliser ces démarches.

Il n'existe pas d'autres sites publics présentant des informations administratives concernant les étrangers au Royaume-Uni. Le corpus anglais sur le thème du droit au séjour (CORPUS IMM-AN) est donc composé d'un seul site web.

4.3.2.3 Thématique citoyenneté

Enfin, nous avons choisi une dernière thématique afin de compléter notre corpus. L'étude de trois thèmes différents nous permet de couvrir la majorité des usagers susceptibles de consulter le discours administratif numérique. Le dernier thème choisi concerne les démarches réalisées par les citoyens et certains résidents longue durée : élections, état civil, permis de conduire, etc. Nous l'avons appelé citoyenneté. Pour ce thème, nous avons inclus un plus grand nombre de sites de municipalités, car ce sont souvent elles qui sont responsables de ces démarches.

Pour le corpus français, nous avons sélectionné les sites suivants :

— Le site officiel de l'administration française (`service-public.fr`), qui présente un bon nombre des démarches sur le sujet au niveau national, mais qui n'indique pas toujours les particularités locales des démarches.

— Les sites officiels des villes de Marseille, Paris et Rennes, qui présentent des informations générales, mais aussi locales, sur ces démarches.

L'inclusion des sites de différentes municipalités nous permet de présenter un échantillon représentatif des informations au niveau national et local. Les villages et petites villes ne sont pas représentés car ils ne donnent pas, ou peu, d'informations administratives sur leurs sites web. Le corpus français sur le thème de la citoyenneté (CORPUS CIT-FR) est donc composé de quatre sites web.

Pour le corpus britannique, nous avons sélectionné les sites suivants :

— Le site officiel de l'administration britannique (`gov.uk`), qui présente une grande partie des démarches sur le sujet au niveau national et sur lequel il est également possible de réaliser ces démarches.
— Les sites officiels des villes de Bristol, Falkirk et York, qui présentent des informations générales, mais aussi locales sur ce type de démarches.

Comme pour le corpus français, l'inclusion des différentes municipalités nous permet de présenter un échantillon représentatif des informations au niveau national et local. Le corpus anglais sur le thème de la citoyenneté (CORPUS CIT-AN) est donc composé de quatre sites. Pour diffuser le discours administratif numérique, les organismes publics ont choisi comme canaux principaux les sites web.

Certains de ces sites ont depuis été transformés en application mobile. Cependant, il ne s'agit que d'une adaptation du site web pour le téléphone. Il ne nous a donc pas paru utile d'inclure les applications dans notre étude, car les textes proposés sont les mêmes.

4.3.3 Création d'un corpus exploitable à partir d'un site web

Dans la section 4.3.1, nous avons défini les différents critères de sélection du corpus que nous avons constitué pour notre étude. À partir de ces critères, nous avons identifié les sites qui ont été utilisés. Nous avons également identifié les objectifs de notre étude : l'exploration terminologique, phraséologique et discursive du corpus afin de caractériser le discours numérique de l'accès aux droits. Cette section est consacrée à la présentation des critères d'exploitabilité du corpus. En effet, le corpus étant un objet numérique, il est nécessaire qu'il soit « lisible par la machine » (McEnery et Wilson, 2001, p. 32).

Comme nous l'avons exposé dans la section précédente, nous avons jugé que la partie contenu des sites était la plus pertinente pour notre étude. Cependant, le site web étant composé de différentes pages HTML organisées en arborescence, il est nécessaire de réaliser un travail d'extraction afin d'obtenir un corpus exploitable.

4.3.3.1 HTML, un outil d'aide pour le chercheur en linguistique

Comme nous l'avons expliqué, l'ensemble des pages sont disponibles sous forme de fichier HTML. Le langage HTML est un langage de balisage, permettant de mettre en forme une page web, en y intégrant des éléments typographiques, des hyperliens, des éléments interactifs, etc. Les balises HTML sont donc généralement utilisées par les rédacteurs web et permettent de structurer les pages web, mais également de donner des informations aux autres rédacteurs quant au rôle des différentes parties du site. En effet, chaque site web possède des conventions de mises en page qui ne sont généralement pas disponibles aux personnes externes. Toutefois, le code HTML étant consultable, il est

souvent facile de comprendre ces conventions. Nous avons donc étudié le code HTML de chaque site pour identifier la balise indiquant le contenu de la page. La Figure 4.2 présente une capture d'écran d'une page du site `service-public.fr`.



FIGURE 4.2 – Illustration du contenu d'une page du site `service-public.fr`

Sur cette page, un cadre rouge identifie le contenu principal de la page, celui qui doit être extrait afin d'être intégré au corpus. Ce contenu a été identifié grâce à une balise HTML, qui le désigne comme étant le texte « principal ». La Table 4.1 représente ensuite une version simplifiée du code HTML de la page.

<main class="main" id="main" role="main" lang="fr>
Accueil >Papiers - Citoyenneté - Élections >Actes d'état civil >Acte de naissance : demande de copie intégrale ou d'extrait La sélection d'une langue déclenchera automatiquement la traduction du contenu de la page. Acte de naissance : demande de copie intégrale ou d'extrait Vérifié le 11 janvier 2023 - Direction de l'information légale et administrative (Première ministre) Vous avez besoin d'un acte de naissance et vous voulez savoir comment faire la demande ? Nous vous guidons dans votre démarche. Elle varie en fonction du lieu de naissance de la personne concernée par l'acte (en France ou à l'étranger). Dans tous les cas, la demande d'un acte de naissance est gratuite. Quelle est votre situation ? Vous demandez votre propre acte de naissance Vous demandez l'acte de naissance d'une autre personne Accéder aux informations générales sans renseigner ma situation
</main>

TABLE 4.1 – Représentation simplifiée du texte balisé comme contenu par le code HTML

Dans la table 4.1, le texte compris entre les deux balises indiquées en rouge relève du contenu principal de la page.

L'utilisation de ces balises afin d'identifier le contenu nous permet de recourir à une méthode systématique et automatique d'extraction, qui garantit un corpus représentatif et diminue la subjectivité dans la sélection des textes récoltés. Chaque site web possède des conventions spécifiques de mise en page et de balisage, ce qui signifie que les balises entourant le contenu principal sont différentes pour chaque site. La Table 4.2 est un récapitulatif des balises qui nous ont servies à construire le corpus dans les sites britanniques.

Site	Balise
bcpcouncil.gov.uk	<main id="main"> </main>
bristol.gov.uk	<main id="main-content" class="main-content"> </main>
cityoflondon.gov.uk	
falkirk.gov.uk	<article id="page-content"> </article>
york.gov.uk	<main class="site-main"> </main>
gov.uk	<main id="content" role="main"> </main>
understandinguniversalcredit.gov.uk	<main id="main" class="site-main"> </main>

TABLE 4.2 – Balises HTML indiquant le contenu sur les pages des sites britanniques

Toutes les balises présentées dans ce tableau contiennent le mot *main*, à l'exception du site `falkirk.gov.uk` qui identifie le contenu principal par l'attribut « page-content ».

La Table 4.3 propose un récapitulatif de l'ensemble des balises pour les sites français.

Site	Balise
service-public.fr	<main class="main" id="main" role="main" lang="fr"> </main>
caf.fr	<main role="main"> </main>
crous-paris.fr	<main style="page_color :var(color-greenwater)"> </main>
paris.fr	<main class="paris-main" id="content" data-page-type="pages"> </main>
marseille.fr	
rennes.fr	<main class="clearfix main" role="main"> </main>
welcometofrance.fr	<section id="content" class="cell columns" role="main"> </section>
prefecturedepolice.interieur.gouv.fr	N/a

TABLE 4.3 – Balises HTML indiquant le contenu sur les pages des sites français

Toutes les balises présentées dans ce tableau contiennent le mot *main*, à l'exception de `marseille.fr` qui identifie le contenu principal par l'attribut « content ». Un seul site n'a pas fait l'objet d'une analyse de son code HTML, il s'agit du site officiel de la préfecture de Paris. Ce site étant rattaché au site du ministère de l'Intérieur, il bénéficie de mesures de sécurité supplémentaires qui rendent impossible son exploration automatique. Nous avons donc téléchargé les pages manuellement. Le site propose une

icône « imprimer » qui permet d'obtenir la page en PDF. Une fois le PDF récupéré, il est converti au format TXT et le contenu est sélectionné manuellement.

Comme nous l'avons observé pour notre étude, le code HTML peut être un outil intéressant pour la construction de corpus issu de sites web dynamiques. Il permet de mettre en place une méthode automatique et systématique d'extraction, mais également de respecter la structure interne des textes mise en place par les rédacteurs.

4.3.3.2 Protocole d'extraction

Afin d'obtenir un corpus stable, nettoyé et représentatif du discours étudié, nous avons établi un protocole de récolte du corpus en cinq étapes.

1. Téléchargement en utilisant un script [8] des sites :
   Les sites sélectionnés sont téléchargés automatiquement grâce à l'utilisation d'un script. Si cela n'est pas possible, le site est téléchargé manuellement page par page.
2. Tri des pages :
   Les pages qui concernent notre thème ou sous-thème (voir section 4.3.2) sont récupérées, tandis que les pages hors thèmes sont supprimées.
3. Récupération de la balise :
   Les pages sont étudiées afin d'identifier la balise indiquant le contenu principal de la page. L'ensemble des balises récupérées sont consultables dans les tableaux 4.2 et 4.3.
4. Extraction du texte grâce à un script :
   Le but du script est de récupérer le texte correspondant au contenu au format TXT, dans un fichier unique encodé en UTF-8 pour l'ensemble des pages d'un site. Pour cela, il est nécessaire d'indiquer la balise recherchée (indiquée dans les Tables 4.2 et 4.3). Une fois la balise renseignée, nous pouvons lancer le script. Le script crée tout d'abord un fichier texte qui pourra accueillir l'ensemble des contenus, et qui devient un des fichiers composant le corpus. Le script parcourt ensuite chaque page HTML du site, et supprime certaines balises parasites, il cherche ensuite à identifier la balise indiquée. Une fois qu'elle est identifiée, il supprime les balises HTML et copie le texte au format TXT, dans le fichier créé préalablement. Chaque texte sélectionné est donc copié dans le fichier texte, qui devient alors un des textes du corpus.
5. Nettoyage du texte :
   Le texte est parcouru et nettoyé manuellement en cas d'erreurs d'extraction. Des

8. téléchargement par aspiration des pages

expressions régulières sont utilisées pour corriger les anomalies typographiques, ainsi que quelques sauts de lignes.

Une fois l'ensemble des étapes du protocole réalisées pour chaque site et chaque sous-thème, nous obtenons le corpus final.

4.3.4 Description du corpus spécialisé dit classique

Nous avons réalisé deux grands corpus spécialisés dans le domaine de l'accès aux droits. Un corpus en français, ainsi qu'un corpus en anglais britannique. Chacun de ces corpus est divisé en trois sous-parties correspondant à trois thématiques : aide sociale, droit au séjour et citoyenneté. La Table 4.4 propose un récapitulatif de ces deux corpus.

	Français (FR)	**Anglais (R-U)**
Aide Sociale	638 741 mots — 4 sites	211 831 mots — 6 sites
Citoyenneté	526 539 mots — 4 sites	303 163 mots — 4 sites
Droit au séjour	446 431 mots — 3 sites	124 564 mots — 1 site

TABLE 4.4 – Récapitulatif des deux corpus dits classiques

Il existe une grande disparité entre les deux corpus, le corpus français étant beaucoup plus fourni que le corpus britannique. Cette différence est due à plusieurs facteurs. Au Royaume-Uni, la communication est centralisée autour d'un site : `gov.uk` et d'une organisation, le Government Digital Service. La France, quant à elle, dispose de multiples acteurs dans la diffusion de l'accès aux droits. L'organisme principal est la Direction de l'information légale et administrative (DILA) qui est responsable de la diffusion d'informations administratives, notamment sur le site `service-public.fr`. Il existe cependant d'autres entités externes au gouvernement responsables d'une certaine partie de la communication. Le site de la caisse d'allocations familiales (`caf.fr`), par exemple, diffuse des informations concernant les aides sociales qu'elle propose. La préfecture de chaque département propose des informations personnalisées sur les procédures de demande de titres de séjour. Il existe également d'autres initiatives de diffusion des informations concernant l'accès aux droits, comme le site web *Welcome to France*, un service d'information bilingue (français et anglais) à destination des étrangers et de leur famille.

Cette disparité s'explique également par le système administratif des deux pays. Les démarches sont moins nombreuses au Royaume-Uni, qui a construit une structure plus centralisée et qui maintenant propose par exemple un dispositif d'aide sociale quasi unique : *Universal credit*, qui rassemble l'allocation logement, familiale et même dans certains cas des allocations liées au travail. La France a pris une direction différente car

elle compte différents organismes suivant les types de services : aide sociale, droit au séjour, au logement, etc. Chacun de ces organismes propose de multiples aides, à l'image des aides au logement, qui se déclinent en trois dispositifs : l'aide personnalisée au logement (Apl), l'allocation de logement familiale (Alf) et l'allocation de logement sociale (Als).

Ces différences ont pour résultat de créer un corpus de texte plus conséquent en France et plus concis au Royaume-Uni. Ces deux corpus reflètent donc des réalités discursives différentes, qui permettront de caractériser le discours de l'accès aux droits dans chacun des pays.

4.4 De nouvelles méthodes d'analyse

Nous avons étudié dans la section précédente l'évolution de la linguistique de corpus et sa définition afin de caractériser un discours issu notamment des sites web dynamiques. Toutefois, les particularités de ce discours, mis en lumière dans la section 3.2, impliquent de se tourner vers de nouvelles méthodes d'analyse. Dans cette section, nous allons présenter des cadres théoriques et méthodologiques qui nous permettront d'étudier les caractéristiques de ce discours. Nous nous intéresserons aux parcours utilisateur, à leur origine et à leur place en analyse de discours numérique, avant de nous tourner vers la discipline de la rédactologie et la théorie des mouvements rhétoriques pour étudier l'organisation rhétorique du discours.

4.4.1 Parcours utilisateur

Le parcours utilisateur est un concept utilisé dans le domaine du web design, notamment en marketing et en ergonomie afin d'orienter l'utilisateur d'un site marchand vers l'achat. Un parcours utilisateur est créé par l'utilisateur lorsque celui-ci se déplace de page en page.

> Un parcours sur le Web s'apparente à un parcours de lecture et d'action dans le cadre d'une navigation hypertextuelle. (Beauvisage, 2004, p. 24)

Dans cette définition, Beauvisage (2004) met en lumière les principales composantes d'un parcours utilisateur : les liens hypertextes qui permettent de naviguer de page en page. Ils sont mis en action par le clic de l'utilisateur, qui déclenche l'ouverture d'une nouvelle page. Les deux composantes sont donc la lecture, puis l'action (le clic).

Pour les spécialistes de ces domaines, les parcours utilisateur servent à identifier le profil des clients potentiels et leur motivation afin d'adapter au mieux leurs services. Il est important de noter que peu de recherches s'intéressent aux parcours utilisateur. Il

s'agit plutôt d'un concept fonctionnel, que l'on retrouve sur de nombreux blogs de marketing. Une seule thèse en sciences du langage semble s'être toutefois intéressée à ce concept Beauvisage (2004), dans le domaine du web sémantique, et nous en avons extrait la définition ci-dessus. Certains chercheurs en marketing s'intéressent à ce concept ou le mentionnent dans leur recherche, mais très peu d'articles en font la mention principale (Sabri et al., 2021). Certains articles font également état du parcours utilisateur dans le cadre de services publics. C'est le cas de Clot (2017) qui s'intéresse à l'expérience utilisateur afin d'améliorer l'accès aux services des bibliothèques.

L'équivalent de ce concept dans le monde anglophone est *User experience map*. Il s'agit d'un outil qui permet de visualiser l'expérience utilisateur afin d'améliorer le site web ou la plateforme. Il est également utilisé dans le domaine de la sécurité informatique (Li et al., 2017).

Pour les besoins de notre étude, le parcours utilisateur fait référence aux différentes pages visitées par l'utilisateur dans l'ordre choisi par ce dernier. La Figure 4.3 propose une illustration d'un parcours utilisateur sur le site `service-public.fr`.

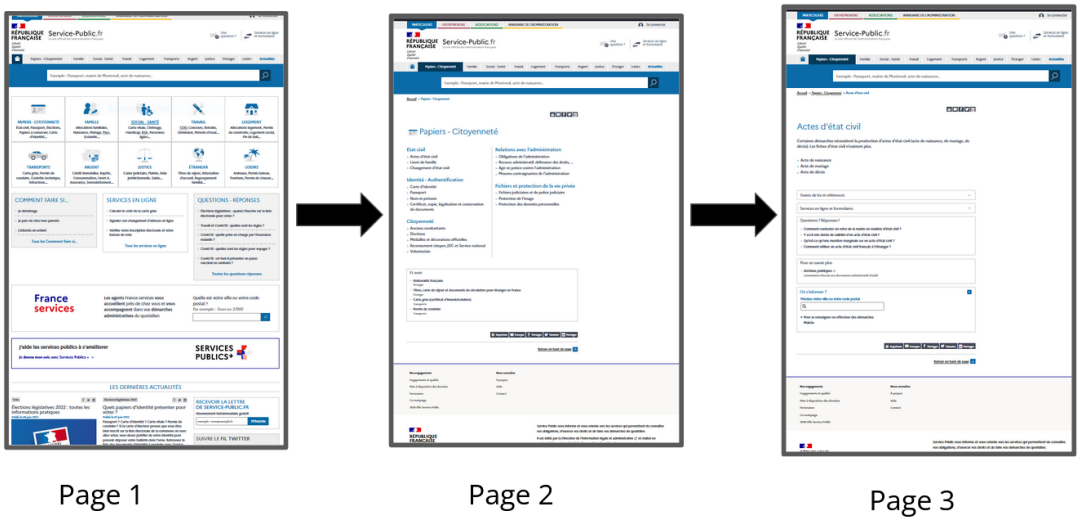


FIGURE 4.3 – Illustration d'un parcours utilisateur sur le site `service-public.fr`, consulté le 10 septembre 2021

Nous pouvons observer trois pages qui se succèdent, que l'utilisateur a visitées. Il accède à ces pages en cliquant sur un hyperlien qui ouvre une nouvelle page. Ces hyperliens lui permettent de se déplacer dans l'arborescence du site. Nous voulons dans cette étude nous réapproprier ce concept afin d'envisager le parcours utilisateur en tant qu'outil d'analyse du discours numérique natif. En effet, certains analystes du discours numérique font référence à la façon dont le discours prend forme avec la navigation.

> L'espace virtuel est un espace qui se structure au fur et à mesure qu'on y navigue. (Vitali-Rosati, 2014, p. 87)

Le parcours utilisateur sera donc une unité d'analyse de notre étude. Ce concept permet ainsi de retracer et d'étudier le parcours discursif.

4.4.2 Les *moves* - analyse de l'organisation rhétorique

Afin d'étudier l'organisation, et plus précisément la structure du discours, nous nous sommes tournée vers les *moves* (ou mouvements rhétoriques). Le concept des *moves* est introduit par Swales (1990) afin d'analyser la structure interne des genres académiques. L'approche consiste à identifier les séquences ou étapes (appelées *moves* et *steps*) d'un article de recherche. Les unités discursives qui forment le texte sont ainsi annotées selon leurs fonctions communicatives. Cette méthode d'analyse des genres permet notamment d'accompagner les personnes non natives dans l'apprentissage de l'anglais académique, en mettant en avant la structure d'un article de recherche (Samraj, 2014).

Les genres présentent des similarités dans le style, le contenu et le public cible (Swales, 1990, p. 58). Ces similarités observées permettent de mettre en avant des régularités dans la structure interne des textes. Chaque unité discursive (ou *move*) possède une fonction qui contribue aux objectifs de communication du genre. Ces unités peuvent être de longueurs variées, mais contiennent toutes au moins une proposition.

L'identification de la structure consiste donc à annoter les *moves* et les *steps*. Dans son ouvrage sur le sujet, Biber et al. (2007) identifie trois étapes d'annotation :

1. comprendre le genre et son objectif rhétorique,
2. observer les fonctions et définir les objectifs,
3. annoter les fonctions ou les thèmes.

Cette approche permet donc d'analyser la place de chaque unité discursive dans le parcours discursif. Nous pouvons noter qu'il est d'abord nécessaire d'obtenir une connaissance générale du genre, de ses caractéristiques externes et de son objectif rhétorique général avant d'annoter les fonctions des différentes unités discursives.

Cette méthode, qui a vu le jour pour décrire le genre académique, a ensuite été utilisée pour décrire la structure générique sous-jacente de plusieurs genres (Moreno et Swales, 2018). Elle pourrait donc servir à décrire la structure du discours de l'accès aux droits dans le cadre de sa caractérisation. Toutefois, comme l'indique Biber et al. (2007) cela sous-entend de connaître les caractéristiques externes de ce genre, que nous avons présenté dans le chapitre 1 de cette thèse. La deuxième étape est donc d'adapter cette méthode à notre genre discursif en particulier, et à notre corpus.

4.4.3 La rédactologie : s'intéresser à la rédaction professionnelle pour comprendre le discours de l'accès aux droits

La rédactologie est une discipline qui intègre les préoccupations de la recherche en langue de spécialité. Un numéro de la revue ASp lui est consacré dès 2002 : *Rédactologie - Situation d'apprentissage*. Les travaux de cette discipline concernent notamment l'apprentissage des langues secondes. Elle est définie ainsi :

> La rédactololologie, étude des conduites scripturales, fait partie d'un conglomérat dynamique composé notamment de la terminologie, de la néologie et de la traductologie et formant, en quelque sorte, un noyau actif au centre d'un hyperespace rédactologique. (Nakbi, 2002, p. 1)

Ce domaine de recherche propose notamment de s'intéresser aux pratiques d'écriture des rédacteurs professionnels, ce que l'auteure appelle « l'étude des conduites scripturales ». Dans le cas de notre discours, il s'agit des personnes en charge de la rédaction des différentes pages des sites web administratifs.

> la rédactologie s'intéresse à la description fine des opérations sous-jacentes au processus d'écriture d'un écrit professionnel, aux caractéristiques de ces écrits et à l'identification des connaissances multidisciplinaires que le rédacteur mobilise tout au long de cette activité. (Beaudet et al., 2016, p. 1)

Ces deux définitions montrent qu'il s'agit d'une discipline interdisciplinaire, qui a pour but d'étudier le processus de création d'un texte spécialisé, mais aussi les caractéristiques de ces textes en utilisant les méthodes mises en place par la terminologie et la traductologie. Cela implique notamment l'utilisation de corpus en adoptant une approche outillée. Cette position nous semble adaptée à nos travaux de recherche dans le discours administratif numérique. Cela est d'autant plus évident que récemment, un ouvrage intitulé *Communication écrite État-citoyens* (Clerc, 2022) apporte de nouveaux éclairages quant au rôle de la rédactologie dans notre domaine : l'accès aux droits.

La transition d'une administration « papier » à une administration numérique a fait l'objet d'un guide en rédactologie intitulé *De la lettre à la page web* publié en 2003. Ce guide se concentre sur la transition numérique au Québec et il donne des conseils pratiques notamment pour la création de pages web destinées aux usagers. Les auteurs proposent des conseils rédactionnels simples et concrets, comme la création de l'interface, des menus ou des titres de page.

Comme Clerc (2022) le fait très pertinemment remarquer, la rédactologie permet d'intégrer les composantes « d'expérience utilisateur » et de parcours utilisateur dans le cas de nos travaux, aux préoccupations plus classiques de l'étude de discours. Cet ouvrage s'intéresse à la relation entre les administrations et leurs usagers dans le cadre

de la transition numérique des gouvernements. Il met en avant de multiples problématiques que nous avons abordées ou que nous aborderons dans cette thèse, en particulier les questions de la stratégie de simplification, mais aussi l'étude des pratiques dites de « lecture-navigation » (Clerc, 2022, p. 2). La « lecture-navigation » fait référence à la consultation d'information écrite sur les sites web, ce qui implique un déplacement de page en page, en utilisant des hyperliens afin d'obtenir l'information désirée.

La rédactologie s'intéressant en même temps aux stratégies de rédaction des professionnels et à la compréhension par les utilisateurs, elle répond aux critères nécessaires pour caractériser le discours étudié dans cette thèse. Elle permet d'aborder les facteurs d'efficacité de la communication entre administrations et citoyens, notamment l'aspect pragmatique de ce type de discours. Nous pensons donc que nos travaux de recherche s'inscrivent partiellement dans cette discipline. L'utilisation des parcours utilisateur et d'un corpus annoté en mouvements rhétoriques à l'appui du code HTML issu des pages web est particulièrement concernée par ces enjeux.

4.4.4 Adaptation à notre corpus

4.4.4.1 Les parcours utilisateur pour l'analyse du discours

L'ensemble de ces méthodes : les parcours utilisateurs et les mouvements rhétoriques sont adaptés à nos objectifs de recherche. Cependant, le discours numérique possède des caractéristiques spécifiques, comme nous l'avons exposé dans la section 3.2. Les méthodes choisies doivent donc être adaptées aux spécificités de notre discours et de nos corpus.

Tout d'abord, le parcours utilisateur est, comme nous l'avons noté dans la section 4.4.1, un outil utilisé en marketing. Nous souhaitons l'utiliser dans la recherche, pour faire des observations et conduire des analyses. Afin de recréer le parcours d'un utilisateur des sites, les professionnels en marketing créent des scénarios suivant des profils d'utilisateur (Bensaïd, 2022). Nous n'avons cependant pas d'informations concernant les profils de nos utilisateurs et il ne nous semble pas adapté aux critères de la recherche de spéculer quant aux types d'utilisateurs qu'il pourrait exister. Nous avons donc décidé d'utiliser l'arborescence du site pour construire ce parcours. En effet, les sites web sont composés de pages classées hiérarchiquement selon une arborescence. Nous nous appuyons donc sur ce classement hiérarchique afin de réaliser nos parcours utilisateur.

Nous avons donc établi deux protocoles de récupération des pages pour les parcours. Ces deux protocoles sont basés sur la terminologie du domaine et l'arborescence.

Le premier protocole consiste à suivre l'arborescence du site en descendant étape par étape dans les menus. Ce protocole a pour but de reconstituer une navigation li-

néaire dans la hiérarchie du site. Elle permet également d'identifier les différentes étapes induites par le menu lors d'une exploration du site.

La Figure 4.4 est une représentation de l'arborescence, et notamment des pages sélectionnées pour un parcours utilisateur.



FIGURE 4.4 – Reconstitution de l'arborescence correspondant à un parcours utilisateur

Les titres correspondent aux titres des pages sélectionnées, tandis que les branches vides correspondent aux autres pages qu'il est possible d'explorer, mais qui n'ont pas été sélectionnées.

Le second protocole consiste à faire une recherche par mot-clé dans la barre de recherche du site et à naviguer entre les pages contenant les informations demandées. La recherche par mot-clé est un autre moyen de se déplacer dans l'arborescence du site, sans évoluer hiérarchiquement. Elle permet de cibler un concept et d'accéder directement aux pages qui les concernent.

Ces deux protocoles sont basés sur la terminologie du domaine et les mots-clés que les usagers sont susceptibles de rechercher. Ils permettent d'observer le discours en contexte autour d'un terme et d'un concept spécifique.

4.4.4.2 Les mouvements rhétoriques sur la page web

Les moves ou mouvements rhétoriques ont peu été utilisés pour caractériser les sites web dynamiques spécialisés. Toutefois, Bathia (2004), dans une étude de genre issue du domaine publicitaire, identifie une évolution des pratiques discursives et intègre des éléments multimodaux dans l'analyse des mouvements rhétoriques de supports de publicité, notamment des images. Il semble donc que l'utilisation initiale de cette théorie puisse s'appliquer à un discours numérique dynamique.

Les mouvements rhétoriques correspondent à des fonctions du texte qui sont intégrées dans sa structure interne et qui décrivent des pratiques de construction du texte interne à une communauté discursive (Bathia, 2004). Nous considérons qu'il est donc

important de se fonder sur des critères externes afin d'identifier ces unités discursives, sans prendre en compte l'interprétation du lecteur.

Afin d'identifier les différents mouvements rhétoriques présents sur les pages web de notre corpus, nous nous sommes donc appuyée sur le code HTML de la page. Celui-ci est composé de balises qui mettent en forme le texte. Nous avons présenté dans la section 4.3.3.1 les utilisations que le chercheur pouvait avoir du code HTML de la page web. Nous avons notamment identifié une certaine partie du texte. Pour cette étude, nous avons généralisé nos observations sur le contenu principal. En effet, les balises contiennent, pour certaines, des informations concernant la fonction du segment qu'elle encode. Ces informations obtenues par le co-texte de la page nous permettent de mettre en lumière les pratiques discursives des rédacteurs web, afin de nous questionner sur leur communauté de discours. Elles permettent également de questionner les objectifs et leur application en discours. Nous avons donc recherché, en plus des balises concernant le contenu principal, d'autres balises donnant des informations sur différentes unités discursives.

Sur le site `gov.uk`, nous avons identifié les fonctions « browse » et « container », qui correspondent respectivement aux fonctions « navigation » et « contenu » de notre schéma. Sur le site `service-public.fr`, les balises ont des attributs nommés « role », qui indiquent le rôle du segment dans la page. On retrouve par exemple les rôles « navigation » et « main », qui correspondent aux fonctions du même nom dans le schéma. L'analyse des balises HTML des deux sites nous a permis de créer différentes catégories de textes présents sur la page web. Cette étude a été complétée par une analyse des textes en ligne en double vue HTML / page web.

L'ensemble de ces pratiques a conduit à la création d'un corpus spécialisé que nous allons présenter dans la section 4.5. L'étude des balises HTML et de différentes fonctions sur les pages sera détaillée dans la section 4.5.3, consacrée au schéma d'annotation créé à partir de ces textes.

4.5 Un corpus réalisé à partir du parcours utilisateur

L'étude de ces nouvelles méthodes, mais aussi l'identification de lacunes dans notre premier corpus nous ont menée à créer un corpus complémentaire pour l'étude d'un discours numérique. Nous avons réalisé un corpus basé sur des parcours utilisateur, que nous avons annoté à partir de l'étude des mouvements rhétoriques.

4.5.1 Mise en place d'une méthodologie complémentaire

La première analyse du corpus classique, décrite dans la section 4.3, a révélé des limitations inhérentes au discours numérique issu de sites web. Le corpus dit classique est composé d'une partie de la page seulement (le *contenu*). Cette sélection était indispensable afin de créer un corpus exploitable et explorable de façon systématique par des logiciels. Toutefois, elle n'est pas entièrement satisfaisante, car elle implique la suppression d'une partie du texte. Il est vrai que le contenu de la page est le plus significatif, car il contient le plus d'informations, mais cela n'empêche pas l'utilisateur d'interagir avec l'ensemble des textes : contenus complémentaires, et autres menus déroulants.

Nous avons donc identifié des limites des outils classiques d'étude de corpus, dans le cadre de l'étude d'un discours numérique diffusé sur un site web dynamique. Avec le logiciel Sketch Engine, nous avons étudié la collocation [pouvoir solliciter]. Celle-ci nous a permis de nous rendre compte des lacunes d'une étude textuelle classique. Par exemple, notre corpus possède un grand nombre de phrases répétées comme représenté dans la Figure 4.5.



FIGURE 4.5 – Résultats pour la collocation de [pouvoir solliciter] sur Sketch Engine

Sketch Engine identifie des collocations grâce à une mesure statistique appelée logDice qui s'appuie sur la fréquence d'utilisation des collocats et des collocations (Rychlý, 2008). Nous avons identifié 72 occurrences de cette phrase exactement identiques. Et il ne s'agit pas de la seule phrase répétée : sur les 582 occurrences de la collocation [pouvoir solliciter], aucune phrase n'est unique. Il est évident que la présence de nombreuses phrases répétées peut fausser la mesure statistique qui identifie la cooccurrence des deux verbes comment étant significative à partir d'un corpus composé d'un grand nombre de phrases répétées. Il est important de noter que la répétition de ces phrases est en partie due au domaine choisi : l'accès aux droits. En effet, le discours administratif est très normé, et les démarches sont souvent les mêmes pour différents profils. Par exemple, la transmission d'un avis d'imposition, dont il est question dans notre exemple, est commune à un très grand nombre de démarches. Il semble donc naturel que la même phrase soit répétée plusieurs fois. Cependant, cette redondance inhérente au discours adminis-

tratif est accentuée par le site web, qui est construit en arborescence. Chaque page est indépendante et requiert donc une réitération des informations. Dans la phrase étudiée sur la Figure 4.5, la fonction de cette répétition est de rediriger l'utilisateur vers un service en ligne.

De plus, comme nous l'analysons dans la section 3.3.5 du chapitre 3 de cette thèse, les répétitions sont un phénomène commun sur le web. Les textes peuvent faire l'objet d'un effet de reprise (Pecman, 2018). Toutefois, cette observation nous a menée à nous interroger sur la validité de telles observations statistiques, sans qu'elle soit mise en regard avec les conditions de navigation réelles de l'utilisateur. En effet, les répétitions n'affectent pas l'utilisateur de la même façon, car il n'a accès qu'à une seule page à la fois. Nous nous sommes donc tournée vers une méthodologie d'analyse de corpus alternative, qui pourrait venir compléter notre étude de corpus dite classique.

Pour ce faire, nous avons étudié les parcours utilisateur, concept que nous avons présenté dans la section 4.4.1. Nous en avons conclu qu'il serait pertinent d'étudier le discours numérique de l'accès aux droits à l'aide de différents parcours utilisateur. L'objectif de l'étude de ces parcours est de mettre en valeur le rôle que jouent les différentes parties du discours présentes sur la page. Pour s'intéresser à ces différentes parties du discours, nous avons décidé d'inclure l'étude des mouvements rhétoriques (voir section 4.4.2) et de les adapter à l'étude d'une page web (voir section 4.4.4.2).

Le second corpus, autrement appelé corpus parcours, est ainsi composé de plusieurs pages, sélectionnées dans un ordre chronologique de navigation, selon les méthodes présentées en section 4.4.4.1. Les différentes fonctions de chacune des parties du discours seront annotées manuellement, à l'aide du code HTML.

4.5.2 Création des parcours utilisateur

Nous avons créé trois parcours utilisateur par corpus, un pour chaque thématique et chaque langue. Comme nous l'avons présenté dans la section 4.4.4.1, nous avons utilisé deux méthodologies différentes pour la création des parcours. Nous avons restreint la création de ces parcours aux sites principaux de chaque pays : `gov.uk` pour le Royaume-Uni et `service-public.fr` pour la France. L'ensemble de ces parcours ont été récoltés en avril 2022. Les sites web étant des objets instables, ces pages, et notamment l'ergonomie des deux sites web, ont été modifiées depuis. Toutefois, la structure de la page et l'ensemble des informations restent très similaires, voire identiques dans certains cas. La première méthodologie utilisée, qui consiste à sélectionner les pages en suivant l'arborescence du site, a mené à la création de deux parcours sur le site `gov.uk`, et deux parcours sur le site `service-public.fr`.

Le premier parcours du site d'accès aux droits britannique concerne la thématique ci-

toyenneté et comporte 5 pages. Il a pour objet le *Birth Certificate* (ou acte de naissance). La Table 4.5 propose un récapitulatif des titres de chaque page. Les pages 1 à 4 montrent une trajectoire descendante dans l'arborescence. La page 5 est une page permettant d'obtenir des informations complémentaires sur l'obtention d'un acte de naissance.

Numéro de page	**Nom de la page**
Page 1	*Welcome to GOV.UK*
Page 2	*Browse Births, deaths, marriages and care*
Page 3	*Certificates, register offices, changes of name or gender*
Page 4	*Order a birth, death, marriage or civil partnership certificate*
Page 5	*Research your family history using the General Register Office*

TABLE 4.5 – Parcours utilisateur du site `gov.uk` concernant l'acte de naissance

Le deuxième parcours du site d'accès aux droits britannique concerne la thématique du droit au séjour et comporte 8 pages. Il a pour objet la naturalisation (*becoming a British citizen*). Il est important d'indiquer que dans le cas de ce deuxième parcours nous avons choisi un « statut » de naturalisation qui est le suivant : « *Apply for citizenship if your spouse is a British citizen* ». Cette particularité n'apparaît pas dans les titres des pages. La Table 4.6 propose un récapitulatif des titres de chaque page.

Numéro de page	**Nom de la page**
Page 1	*Welcome to GOV.UK*
Page 2	*Citizenship and living in the UK*
Page 3	*British citizenship*
Page 4	*Check if you can become a British citizen*
Page 5	*Eligibility and fees*
Page 6	*How to apply*
Page 7	*After you've applied*
Page 8	*After you get your certificate*

TABLE 4.6 – Parcours utilisateur du site `gov.uk` concernant la naturalisation

Le premier parcours du site principal d'accès aux droits français (`service-public.fr`) concerne la thématique citoyenneté et comporte 8 pages. Il a pour objet l'acte de naissance. La Table 4.7 propose un récapitulatif des titres de chaque page. Les pages 1 à 6 montrent une trajectoire descendante dans l'arborescence du site. La page 7 est une page permettant d'obtenir des informations complémentaires sur l'acte de naissance.

Numéro de page	Nom de la page
Page 1	Accueil Particuliers
Page 2	Papiers - Citoyenneté
Page 3	Actes d'état civil
Page 4	Acte de naissance demande de copie intégrale ou d'extrait
Page 5	Demande d'une copie d'un extrait conservé au répertoire civil - etape_1
Page 6	Demande d'une copie d'un extrait conservé au répertoire civil (Formulaire 13485 02)
Page 7	Qu'est-ce qu'une mention marginale sur un acte d'état civil

TABLE 4.7 – Parcours utilisateur du site `service-public.fr` concernant l'acte de naissance

Le deuxième parcours du site d'accès aux droits britannique concerne la thématique droit au séjour et comporte 4 pages. Il a pour objet la naturalisation. Il est important d'indiquer que dans le cas de ce deuxième parcours nous avons choisi un « statut » de naturalisation qui est la naturalisation par mariage, pour avoir un thème identique au parcours britannique. La Table 4.8 propose un récapitulatif des titres de chaque page.

Numéro de page	Nom de la page
Page 1	Accueil Particuliers
Page 2	Étranger - Europe
Page 3	Comment obtenir la nationalité française ?
Page 4	Nationalité française par mariage

TABLE 4.8 – Parcours utilisateur du site `service-public.fr` concernant la naturalisation par mariage

La seconde méthodologie utilisée, qui consiste à sélectionner les pages en faisant une recherche par mot-clé sur le site, a mené à la création d'un parcours sur le site `gov.uk`, et d'un parcours sur le site `service-public.fr`.

Le parcours du site principal d'accès aux droits britannique concerne la thématique aide sociale et comporte 5 pages. Il a pour objet les *Housing Benefits* (ou aides au logement). La Table 4.9 propose un récapitulatif des titres de chaque page.

Numéro de page	Nom de la page
Page 1	*Welcome to GOV.UK*
Page 2	*housing benefits - Search*
Page 3	*Housing Benefit Eligibility*
Page 4	*Appeal a Housing Benefit decision*
Page 5	*Housing Benefit information from your council*

TABLE 4.9 – Parcours utilisateur du site *gov.uk* concernant les aides au logement

Le parcours du site principal d'accès aux droits français concerne la thématique aide sociale et comporte 5 pages. Il a pour objet les allocations logement. La Table 4.10 propose un récapitulatif des titres de chaque page.

Numéro de page	**Nom de la page**
Page 1	Accueil Particuliers
Page 2	Résultat(s) pour : « aide au logement »
Page 3	Aide personnalisée au logement (APL)
Page 4	Quelles sont les différences entre les allocations logement ?
Page 5	Un étudiant peut-il toucher une aide au logement (APL, ALS, ALF) ?

TABLE 4.10 – Parcours utilisateur du site `service-public.fr` concernant les aides au logement

Ces six parcours ont donc été le point de départ pour la création de notre corpus spécialisé spécifiquement adapté pour l'étude du discours numérique, appelé corpus parcours. Bien évidemment, la conception manuelle de ce corpus explique qu'il soit de petite taille. La Table 4.11 ci-dessous propose un récapitulatif de ce corpus.

Parcours	**Anglais**	**Français**
Acte de naissance	4 568 mots	26 350 mots
Aides au logement	5 001 mots	14 166 mots
Naturalisation	7 585 mots	29 051 mots
Total	15 157 mots	69 567 mots

TABLE 4.11 – Tableau récapitulatif du corpus parcours dans les deux langues

4.5.3 Mise en place d'un schéma d'annotation

4.5.3.1 Identification des fonctions des unités discursives

Pour étudier les mouvements rhétoriques et les fonctions du discours sur la page, nous avons décidé d'annoter chaque page selon les fonctions des parties du discours. Pour annoter le discours, nous avons mis en avant trois fonctions principales du texte. Ces fonctions ont été récoltées grâce au code HTML des sites `service-public.fr` et `gov.uk`, mais aussi grâce à l'analyse de l'organisation de la page. Nous avons mis en commun les informations contenues dans les codes HTML des deux sites afin de réaliser ce schéma d'annotation. Les fonctions identifiées sont les suivantes :

— *contenu* :

Il s'agit de la partie précédemment identifiée pour collecter le premier corpus.

— *navigation* :

  Il s'agit de parties du texte qui ont pour fonction principale d'aider l'utilisateur à naviguer sur le site web.

— *technique* :

  Il s'agit de parties du texte qui ont pour fonction principale d'informer l'utilisateur concernant des particularités relatives au site web.

Nous avions déjà identifié les unités correspondant au *contenu* grâce aux balises présentées dans les tableaux 4.2 et 4.3. Ces balises ont donc été réutilisées pour identifier les unités discursives correspondant au *contenu* dans les parcours utilisateur.

Sur le site `gov.uk`, les unités correspondant à la *navigation* ont été identifiées grâce à la balise à l'attribut « role=navigation ». Nous avons également sélectionné des unités ayant une balise contenant l'attribut suivant : « navigation-header ». Enfin, nous avons sélectionné le bas de page, qui est identifié par la balise « footer », mais qui, au regard de la bannière du haut de la page, permet de naviguer sur l'ensemble du site.

Sur le site `service-public.fr`, les unités correspondant à la navigation ont été identifiées grâce à la balise « <nav » et à l'attribut « role=navigation ». Nous avons également sélectionné le bas de page qui est identifié par la balise « footer », mais qui, au regard de la bannière du haut de la page, permet de naviguer sur le site.

La Figure 4.6 est une illustration des différentes parties d'une page du site `gov.uk`. Les parties encadrées en rouge correspondent aux unités de *navigation*, tandis que les parties encadrées en bleu correspondent aux parties *contenu*. Pour chacune de ces parties, nous avons indiqué dans un encadré les différentes balises ou attributs correspondants qui nous ont servis à construire le schéma d'annotation.

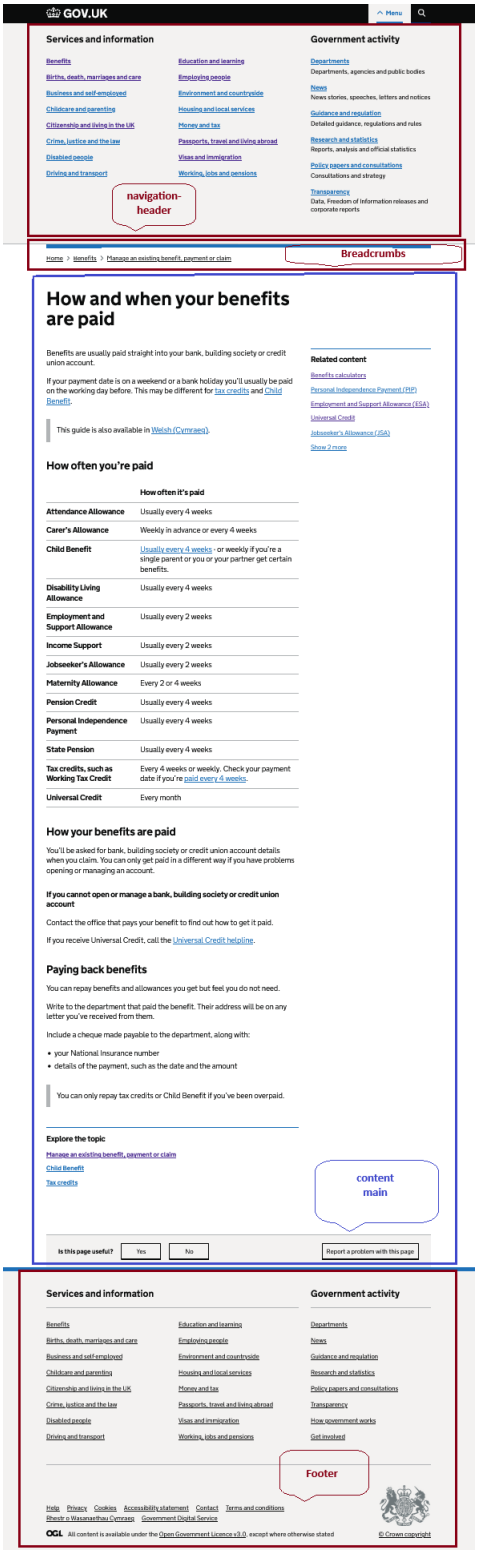


FIGURE 4.6 – Différentes parties d'une page du site `gov.uk`

La Figure 4.7 est une illustration des différentes parties d'une page du site `service-public`.

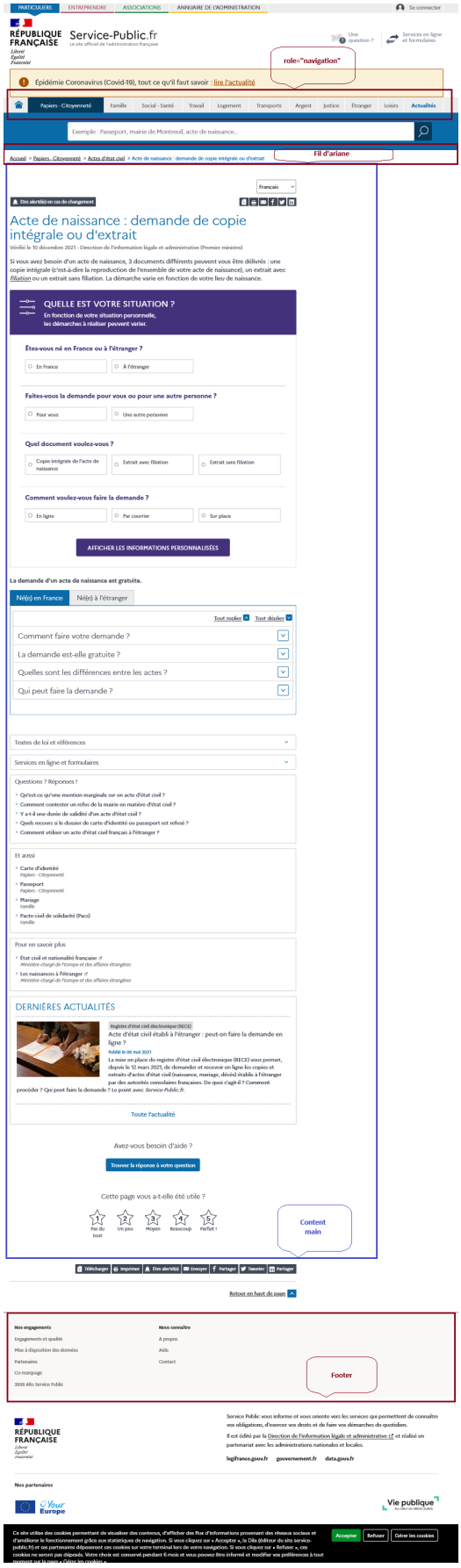


FIGURE 4.7 – Différentes parties d'une page du site `service-public.fr`

Les parties encadrées en rouge correspondent aux unités de *navigation*, tandis que les parties encadrées en bleu correspondent aux parties *contenu*. Pour chacune de ces parties, nous avons indiqué dans un encadré les différentes balises ou attributs correspondants qui nous ont servis à construire le schéma d'annotation.

Nous souhaitons attirer l'attention sur le fait qu'un seul schéma d'annotation a été réalisé pour les deux langues afin de pouvoir comparer les deux corpus. En raison des similitudes marquées entre les deux sites, nous avons fusionné les analyses effectuées sur les deux sites, en tenant compte des deux encodages HTML, pour élaborer ce schéma.

Dans les parties identifiées comme étant du *contenu*, nous en avons identifié différents types. Quatre types ont été annotés :

— *principal* :
Cette partie du texte fournit des informations relatives au thème de la page.
— *complémentaire* :
Cette partie du texte fournit des informations sur des sujets en lien avec le thème de la page et renvoie l'utilisateur vers les pages concernées. Nous avons identifié cette particularité du discours sur le site français `service-public.fr`, qui contient une balise dotée de l'attribut suivant : « role="complementary" ». Nous avons donc identifié qu'il s'agissait de *contenus complémentaires*. Le site `gov.uk` contient le même type de rubrique, que nous avons identifié comme équivalent.
— *actualité* :
Cette partie du texte fournit des informations sur des sujets d'actualité (allocation de rentrée en septembre, statuts des réfugiés pour la guerre en Ukraine au moment du déclenchement de la guerre, etc.) et renvoie sur les pages concernées. Cette partie du *contenu* n'a pas été identifiée par un code particulier, mais par un titre « actualités » sur le site français et « *Featured* » sur le site britannique.
— *amélioration* :
Cette partie du texte permet à l'utilisateur d'identifier des améliorations à réaliser dans le contenu du site. Cette rubrique est systématiquement présente sur les deux sites. Elle n'est pas expressément étiquetée comme telle dans le code source, mais représente un élément stable et essentiel de chaque page.

Les Figures 4.8, 4.9 et 4.10 présentent des unités étiquetées comme étant du *contenu complémentaire* sur les deux sites web concernés par le corpus parcours.

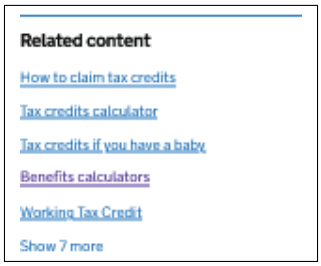


FIGURE 4.8 – *Contenu complémentaire* sur le site `gov.uk`

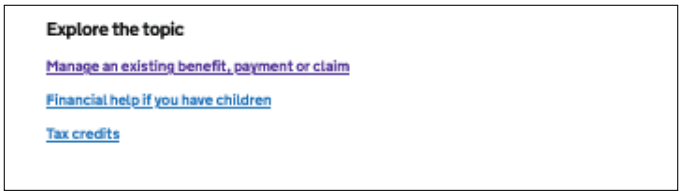


FIGURE 4.9 – *Contenu complémentaire* sur le site `gov.uk`

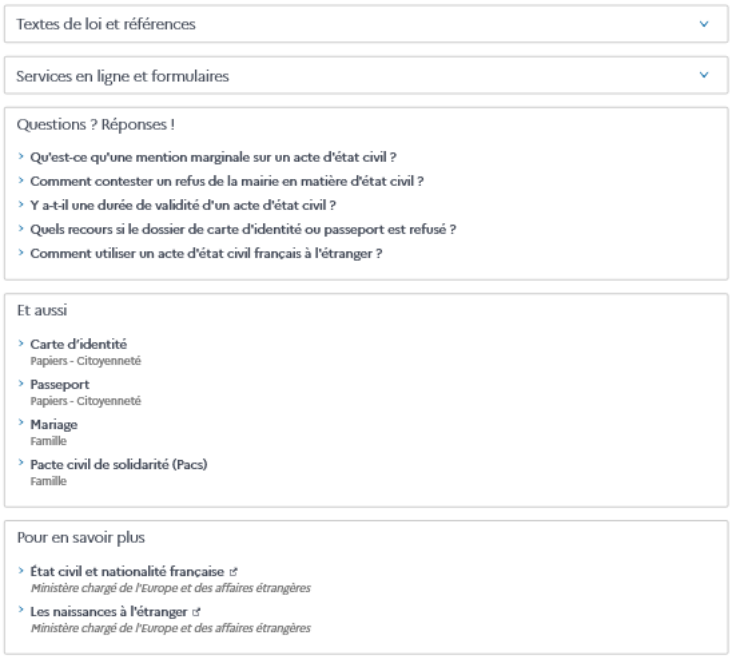


FIGURE 4.10 – *Contenu complémentaire* sur le site `service-public.fr`

Les Figures 4.11 et 4.12 présentent des unités étiquetées comme étant de l'*actualité* sur les deux sites web concernés par le corpus parcours.

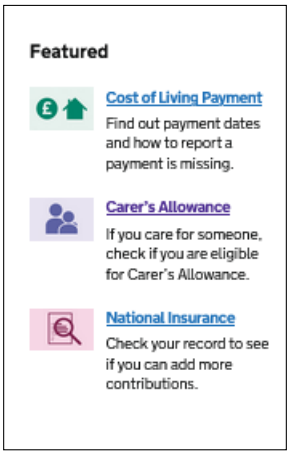


FIGURE 4.11 – Rubrique *actualité* sur le site `gov.uk`



FIGURE 4.12 – Rubrique *actualité* sur le site `service-public.fr`

Une fois les unités sélectionnées comme étant de la *navigation* ou du *contenu*, nous avons procédé au téléchargement. Une fois les pages téléchargées, nous avons rencontré une troisième fonction du discours, qui n'apparaît pas toujours sur les pages. Il s'agit de parties de la page que nous avons étiquetées comme *technique*.

Ces unités discursives sont généralement affichées sous forme de bannières ou de pop-ups. Elles concernent, pour les plus communes, les cookies. Mais elles peuvent également prendre la forme d'aide à la navigation pour utiliser les barres de recherches, ou encore les formulaires.

Les Figures 4.13 et 4.14 représentent une partie de la page étiquetée comme *technique*.



FIGURE 4.13 – Pop-up *technique* d'une page sur le site `gov.uk`



FIGURE 4.14 – Pop-up *technique* d'une page sur le site `service-public.fr`

Il ne s'agit pas des seuls types de discours *technique*. Pour plus de précisions, nous avons introduit une typologie qui divise le discours *technique* en trois catégories :

— *aide* :
Il s'agit d'une catégorie qui englobe l'ensemble du discours dédié à guider l'utilisateur pour naviguer sur la page web. Elle comprend les discours présentés dans les Figures 4.13 et 4.14.

— *problème* :
Cette catégorie concerne l'ensemble des parties destinées à indiquer à l'utilisateur qu'un problème technique a eu lieu sur la page.

— *juridique* :
Il s'agit de l'ensemble des discours juridiques qui concernent uniquement le site web, et non les lois liées aux contenus des pages. Cela comprend la licence des sites, mais également le discours concernant la reproduction du site, etc.

L'ensemble de ces observations et analyses ont mené à la création d'une partie du schéma d'annotation consacrée aux fonctions des unités discursives que nous allons présenter dans la Figure suivante.

La Figure 4.15 présente une partie du schéma d'annotation dédié aux mouvements rhétoriques, caractérisés dans notre cas par des fonctions sur une page web.

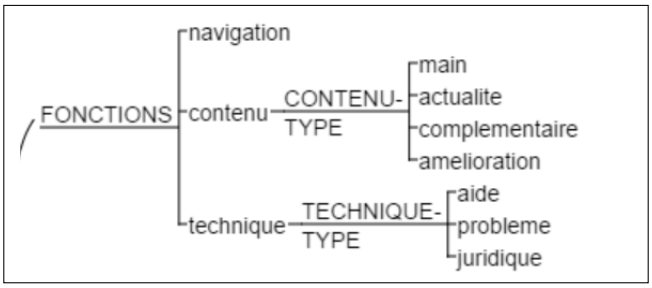


FIGURE 4.15 – Schéma d'annotation - partie fonctions

4.5.4 Intégration des hyperliens

Dans le chapitre 3 de notre thèse, nous avons décrit les particularités du discours numérique. La section 3.2.3 est notamment consacrée à l'importance de l'hyperlien dans la définition de ce discours, en particulier sur les sites web dynamiques. C'est pourquoi nous avons décidé d'intégrer l'annotation des liens à notre corpus parcours.

Les liens sont présents dans l'ensemble de la page et pas seulement dans les parties consacrées à la *navigation*. C'est pour cela qu'il nous a semblé essentiel de les prendre en compte. Nous avons donc annoté chaque hyperlien manuellement, en observant le texte source en situation de navigation. Une typologie a également été mise en place en différenciant deux types de liens : *externe*, redirigeant vers un autre site et interne, redirigeant sur le site visité. Les liens internes ont ensuite été divisés en deux catégories : *page*, s'ils redirigent sur une autre section de la même page, ou *site*, s'ils redirigent sur une autre page du site. La Figure 4.16 illustre la partie du schéma consacré aux hyperliens.

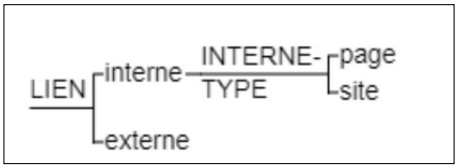


FIGURE 4.16 – Schéma d'annotation - partie hyperliens

4.5.5 Intégration des éléments interactifs

Les éléments interactifs du site web ont également retenu notre attention. En effet, en plus des hyperliens, des menus déroulants et formulaires sont présents sur le site. L'usager peut ainsi cliquer sur la page pour faire apparaître du texte, ou encore cocher des cases afin de faire apparaître du contenu personnalisé.

L'ensemble de ces éléments interactifs ont été annotés afin d'être intégrés à notre étude. La typologie suivante a été utilisée pour classer ces éléments. Dans un premier

temps, nous avons identifié les *formulaires*. Notre schéma n'a pas pour but d'annoter les formulaires administratifs, mais bien les formulaires numériques. Un formulaire est un espace (généralement un rectangle), où l'utilisateur peut transmettre des informations aux serveurs du site web. Ces informations transmises permettent généralement d'accéder à du contenu personnalisé. Nous avons ensuite choisi d'annoter les éléments *éphémères*, qui sont les pop-ups ou informations d'actualité. Enfin, nous avons choisi d'annoter les *menus déroulants*, c'est-à-dire les textes sur lesquels l'utilisateur doit cliquer afin de faire apparaître du texte. La Figure 4.17 illustre la branche concernant ces éléments interactifs dans le schéma d'annotation.

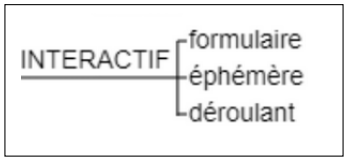


FIGURE 4.17 – Schéma d'annotation - éléments interactifs

4.5.6 Protocole de récolte du corpus parcours

Nous allons présenter dans cette section les différentes étapes mises en place afin de créer le corpus parcours.

1. Sélection des parcours :
   Nous avons tout d'abord sélectionné les parcours, à partir d'une observation du site et d'une connaissance de l'arborescence. Les parcours sélectionnés sont présentés dans la section 4.5.2.
2. Téléchargement des pages :
   Nous avons ensuite réalisé le téléchargement des pages concernées. Chaque page a été téléchargée manuellement sur le site, puis rangée dans des dossiers correspondant au parcours.
3. Identification des balises correspondant aux fonctions du texte :
   Nous avons ensuite étudié le code HTML des pages afin d'identifier les balises correspondant aux fonctions du texte.
4. Création du schéma d'annotation :
   À la suite de l'identification des fonctions, nous avons créé un schéma d'annotation regroupant l'ensemble des fonctions et des éléments identifiés sur la page. Ce schéma d'annotation est présenté dans la Figure 4.18
5. Nettoyage du texte HTML :
   Une fois les pages téléchargées et les balises sélectionnées, nous avons pu nettoyer le texte du code HTML non utile à l'annotation.

6. Annotation manuelle :
   Enfin, nous avons réalisé une annotation manuelle sur le logiciel UAM Corpus Tool, présenté dans la section 4.6.1.

La Figure 4.18 présente le schéma d'annotation dans son ensemble.

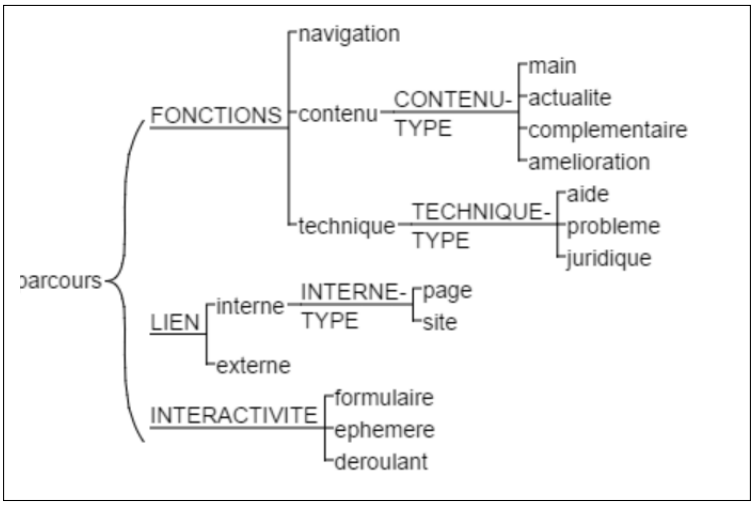


FIGURE 4.18 – Schéma d'annotation complet

4.6 Outils et démarches

4.6.1 Présentation des différents outils utilisés

4.6.1.1 UAM Corpus Tool

Pour réaliser le corpus parcours, nous avons utilisé le logiciel UAM Corpus Tool (version 6) (O'Donnell, 2008). Nous y avons intégré un fichier texte pour chaque page et avons donc créé trois sous-corpus différents. Ce logiciel permet également de faire des annotations manuelles sur le texte. Il était particulièrement adapté car nous souhaitons annoter l'ensemble du texte par segments. L'utilisateur peut manuellement attribuer un tag à un segment de texte ce qui permet d'annoter de longs textes et rend le logiciel facile à manipuler. Nous avons ensuite utilisé la fonction de recherche afin d'isoler les différentes parties annotées et de les extraire. Une fois extraites, nous avons appliqué le script présenté dans la section 4.5.6. Ce script permet d'obtenir un document texte composé des segments annotés voulus.

4.6.1.2 Sketch Engine

Pour explorer les corpus et sous-corpus, nous avons également utilisé le logiciel Sketch Engine (Kilgarriff et al., 2014). Ce logiciel est un gestionnaire de corpus et un lo-

giciel d'analyse textuelle. Il propose divers outils d'exploration et d'analyse. Nous avons utilisé les lignes de concordance pour analyser le corpus, notamment avec le KWIC (*Key-Word In Context*). Nous avons également utilisé les outils de Fréquence qui permettent de classer les mots-clés ou les cooccurrents par fréquence, ainsi que les outils permettant d'identifier les différents collocats. Le logiciel propose un outil d'annotation, mais il n'est possible que d'annoter certains mots ou segments et non l'ensemble du texte.

Troisième partie

Résultats et analyses

Chapitre 5

Caractérisation linguistique du discours numérique natif

Ce chapitre se consacre à la caractérisation linguistique du discours à différents niveaux. La première partie est dédiée à une caractérisation lexicale, où nous examinons la terminologie du domaine, en particulier les termes formés par des groupes nominaux complexes et dans certains cas les termes composés. Nous étudions également les patrons lexico-syntaxiques et la variation terminologique. Dans une deuxième partie, nous réalisons une caractérisation syntaxique en menant une analyse des pronoms les plus utilisés et de leurs collocations. Cette analyse est suivie d'une exploration des schémas conditionnels. Enfin, la troisième partie se concentre sur une caractérisation lexico-syntaxique, stylistique et rhétorique, en explorant la phraséologie employée par certains organismes d'accès aux droits, les phrases répétées et l'expression de la modalité dans le discours.

5.1 Caractérisation lexicale

5.1.1 Groupes nominaux complexes

Les groupes nominaux complexes représentent une grande partie des noms de dispositifs dans notre corpus. Nous avons fait cette observation dans un premier temps dans le sous-corpus sur la thématique aide sociale. En effet, nous avons identifié et décrit un principe de formation des termes relatifs aux dispositifs d'aide.

Dans une étude pilote du corpus réalisée dans le cadre de notre mémoire de master (Bouchet, 2021), nous avons mis en avant et identifié les termes clés du domaine de l'aide sociale. Il s'est avéré que les principaux termes clés de ce domaine étaient des groupes nominaux complexes.

Nous y avons identifié et décrit un principe de formation des termes relatifs aux dispositifs d'aide (Bouchet, 2021 : 71). Ces dispositifs sont souvent complexes et leur dénomination est formée selon une méthode qui nous a parue comme stable. Nous avons ainsi formalisé le schéma suivant :

[noyau] [préposition] [complément1] [complément2] [complément3] [...]

L'identification de ce schéma dans le corpus d'aide sociale nous a servi de base à l'étude de la terminologie faisant référence aux dispositifs administratifs destinés aux usagers. Nous avons ensuite procédé à une analyse par thème, car chacun des thèmes abordés dans notre étude présente une terminologie unique, construite dans le cadre de la législation. Les explorations terminologiques font ensuite l'objet d'une analyse comparative bilingue. Précisons aussi que notre analyse n'exclut pas les termes composés. En effet, comme le discours administratif et les trois thèmes étudiés sont caractérisés par les groupes nominaux complexes, notre analyse se focalise sur ces derniers, mais elle n'exclut pas les quelques termes composés qui ont pu être relevés lors de l'exploration du corpus.

5.1.1.1 Terminologie dans le corpus français sur le thème des aides sociales

L'ensemble des explorations présentées dans cette section ont été faites dans le corpus AIDESOC-FR qui reprend l'ensemble des pages dédiées aux aides sociales.

Afin de faire les recherches en corpus sur Sketch Engine, il est nécessaire de transformer le patron syntaxique relevé en expression CQL[1]. Nous avons donc fait des recherches à partir de l'expression CQL suivante :

[tag="NC.*"][tag="SP"]([tag="NC.*"] | [tag="A.*"])

L'expression CQL correspond au patron syntaxique suivant :

[nom commun][préposition][nom commun] ou [adjectif]

Les explorations sont initialement réalisées à partir d'étiquettes syntaxiques attribuées automatiquement par le logiciel Sketch Engine. Une fois les exemples relevés, nous les explorons manuellement en corpus. C'est pourquoi le nombre d'occurrences relevées dans les tableaux initiaux n'est pas toujours le même que celui relevé ensuite. L'étiquetage automatique n'est pas toujours précis et l'utilisation des majuscules peut fausser le nombre d'occurrences ressorti par le logiciel.

Nous avons donc effectué une requête qui nous a permis de relever les groupes nominaux complexes présentés dans la Table 5.1 ci-dessous.

1. Corpus Query Language

word	Frequency	Freq. per million	% of concordance
allocation de soutien familial	114	151,32248	26,2069
revenu de solidarité active	81	107,5186	18,62069
allocation de rentrée scolaire	63	83,62558	14,48276
bourse sur critères sociaux	32	42,47649	7,35632
allocation de logement sociale	25	33,18475	5,74713

TABLE 5.1 – Dispositifs d'aide financière pour le patron [nom commun][préposition][nom commun] ou [adjectif]

Ces relevés initiaux nous ont confirmé que le patron choisi correspondait à des noms de dispositif et ils nous ont guidés pour faire des explorations. Nous nous sommes également appuyée sur les noyaux suivants pour réaliser nos explorations :

— aide,
— allocation,
— revenu,
— complément,
— prestation,
— bourse.

Une fois ces premiers relevés réalisés, nous avons donc pu observer l'ensemble des termes se rapportant aux dispositifs d'aide. Le patron syntaxique a donc été modifié de la façon suivante :

[**nom noyau**][préposition][nom commun]

5.1.1.1.1 Analyse des patrons lexico-syntaxiques avec le noyau « allocation »

La Table 5.2 présente les résultats du patron avec le noyau **allocation** relevés en corpus par Sketch Engine.

word	Frequency	Freq. per million	% of concordance
allocation de base	157	208,40026	19,67419
allocation de logement	133	176,54289	16,66667
allocation de rentrée	129	171,23333	16,16541
allocation de soutien	119	157,95943	14,91228
allocation d'éducation	91	120,79251	11,40351
allocation de solidarité	41	54,423	5,13784
allocation d'aide	31	41,1491	3,88471
allocation de remplacement	23	30,52997	2,88221
allocation de chômage	13	17,25607	1,62907
allocation de retour	12	15,92868	1,50376
allocation d'activité	10	13,2739	1,25313

TABLE 5.2 – Résultats les plus fréquents pour le patron [**allocation**][préposition][nom commun]

Dans un premier temps, nous allons analyser le second terme de la Table 5.2 : **allocation de logement**. Nous avons d'abord trouvé trois lexèmes fréquemment postposés à ce terme, dont deux compléments adjectivaux (**familiale** et **sociale**) qui entrent dans la création de deux termes concurrents, et une préposition (**à**).

— **allocation de logement familiale** (25 occurrences)
— **allocation de logement sociale** (25 occurrences)
— **allocation de logement à** (57 occurrences)

Nous avons donc exploré plus amplement cette dernière occurrence, qui nous a permis d'identifier deux concurrents supplémentaires :

— **allocation de logement à caractère familial** (28 occurrences sur `caf.fr` uniquement)
— **allocation de logement à caractère social** (29 occurrences sur `caf.fr` uniquement)

Ces deux termes sont issus du site `caf.fr` uniquement, tandis que les occurrences des relevés présentés dans la liste ci-dessous sont issues de plusieurs sites.

— **allocation de logement familiale** (28 occurrences) : 7 occurrences issues de `Service-public.fr`, 21 issus de `caf.fr`
— **allocation de logement sociale** (33 occurrences) 12 occurrences issues de `Service-public.fr`, 21 issus de `caf.fr`

L'analyse de ces relevés nous permet de faire plusieurs observations. Tout d'abord, il existe deux dénominations pour un même dispositif : **allocation de logement à caractère familial** et **allocation de logement familiale** qui sont synonymes ou **allocation de logement à caractère social** et **allocation de logement sociale** qui sont également sy-

nonymes. La dénomination formée à partir du schéma suivant : **allocation de logement à caractère [complément]** n'existe que sur un seul site, tandis que l'autre (**allocation de logement [complément]**) se trouve sur plusieurs sites. Ces termes sont également systématiquement utilisés ensemble sur le site `caf.fr`, comme le montrent les exemples ci-dessous :

(1) Pourl'allocation de logement à caractère familial (Alf) ou l'allocation de logement à caractère social (Als), le montant vous est versé directement. (`caf.fr`)

(2) Si vous payez un loyer ou remboursez un prêt pour votre résidence principale, et si vos ressources sont modestes, vous pouvez bénéficier de l'une des deux aides au logement suivantes : l'allocation de logement familiale (Alf) ou l'allocation de logement sociale (Als). (`caf.fr`)

Nous allons maintenant nous intéresser au premier terme apparaissant dans la Table 5.2 : **allocation de base**. Il s'agit d'une des composantes de la **Prestation d'Accueil Jeune Enfant**, un dispositif d'aide versé par la Caf. Ce terme est parfois accompagné des compléments « de la Paje » ou « de la prestation d'accueil du jeune enfant », afin d'indiquer qu'il fait partie d'un autre dispositif d'aide. Il existe un sigle correspondant (**Ab**) qui apparaît uniquement entre parenthèses à la suite du terme complet.

— **allocation de base** (174 occurrences, dont 40 sur `service-public.fr` et 134 sur `caf.fr`)
— **ab** (14 occurrences sur `caf.fr`)

Le troisième terme relevé est **allocation de rentrée scolaire** qui présente 146 occurrences en corpus, majoritairement sur les sites `caf.fr` et `service-public.fr`. Il s'agit d'une aide versée une fois par an seulement.

— **allocation de rentrée scolaire** (146 occurrences dont 133 occurrences sur `caf.fr`, 18 occurrences sur `service-public.fr` et 5 sur `paris.fr`)

Pour le terme **allocation de soutien**, nous avons relevé plusieurs termes formés à partir de cette base :

— **allocation de soutien aux parents d'enfants handicapés** (15 occurrences sur `paris.fr`)
— **allocation de soutien familial** (153 occurrences dont 30 sur `service-public.fr` et 123 sur `caf.fr`)

Dans les deux cas, il s'agit de dispositifs mis en place par différents organismes, notamment la Ville de Paris et la Caisse d'allocations familiales.

Pour **allocation d'éducation**, le terme correspondant est **allocation d'éducation de l'enfant handicapé**.

— **allocation d'éducation de l'enfant handicapée** (133 occurrences dont 114 occurrences sur `caf.fr`, 19 sur `service-public.fr`)

Le terme **allocation de solidarité** a deux compléments en corpus :

— **allocation de solidarité spécifique** (41 occurrences)

— **allocation de solidarité aux personnes âgées** (27 occurrences)

Ces deux dispositifs ne sont pas proposés par les mêmes organismes. L'**allocation de solidarité spécifique** est proposée par Pôle emploi. Il s'agit d'un dispositif dans le prolongement des allocations chômage. L'**allocation de solidarité aux personnes âgées** est versée par les caisses de retraite. Il est intéressant de noter que deux organismes différents utilisent des processus de dénominations similaires.

Le relevé **allocation d'aide** est plutôt surprenant, car il rassemble deux des noms noyaux présentés dans cette section. Il forme les termes suivants :

— **allocation d'aide au retour à l'emploi** (31 occurrences sur `service-public.fr`)

— **allocation d'aide au retour à l'emploi formation** (6 occurrences, phrase répétée sur `service-public.fr`)

— **allocation d'aide sociale** (1 occurrence, `service-public.fr`)

Les deux premiers dispositifs sont très similaires puisqu'il s'agit dans les deux cas d'une aide destinée aux demandeurs d'emploi. Pour le dernier terme, il n'existe qu'une occurrence en corpus :

(3) (...) recevoir l'une des allocations suivantes (allocation pour adultes handicapés (AAH) ou allocation compensatrice ou allocation d'aide sociale (dite allocation simple).

D'après cette occurrence et ce contexte, il s'agit d'une dénomination alternative. Cet exemple permet d'observer qu'il existe à la fois plusieurs dénominations pour désigner un même dispositif (par exemple, **allocation d'aide sociale** et son variant **allocation simple**, ou **allocation de logement sociale** et son variant **allocation logement à caractère social**) et qu'un grand nombre de dispositifs formés avec le noyau **allocation** sont utilisés de manière concomitante en corpus.

Le relevé **allocation de remplacement** correspond aux termes suivants :

— **allocation de remplacement pour maternité** (11 occurrences, 3 sur `service-public.fr`, 8 sur `caf.fr`)

— **allocation de remplacement pour maternité ou paternité** (10 occurrences, 2 sur `service-public.fr`, 8 sur `caf.fr`)

Le terme **allocation de chômage** a un statut particulier. En effet, ce terme ne correspond pas strictement à un dispositif. Il s'agit d'une utilisation dérivée du terme initial qui est **allocation chômage**. Il peut s'agir également, dans certaines utilisations, d'un variant de **allocation d'aide au retour à l'emploi** qui est le terme juridique.

— **allocation de chômage** (35 occurrences dont 30 sur `caf.fr` et 5 sur `service-public.fr`)
— **allocation chômage** (116 occurrences dont 16 sur `caf.fr` et 100 sur `service-public.fr`)

Nous avons réalisé une étude de ce concept dans la section 5.1.3.1.

Le terme **allocation de retour à l'emploi** (12 occurrences) est également un variant de **allocation d'aide au retour à l'emploi** que nous avons présenté précedemment dans cette section. Il s'agit d'une occurrence dérivée du terme juridique, comme nous l'analysons dans la section 5.1.3.1.

— **allocation de retour à l'emploi** (9 occurrences sur `service-public.fr` et 3 occurrences sur `paris.fr`)

Enfin, le dernier relevé proposé est le suivant : **allocation d'activité partielle** (10 occurrences). Il se trouve majoritairement sur le site `service-public.fr`. Il n'est pas versé par un organisme comme la Caf ou Pôle emploi, mais par un employeur.

— **allocation d'activité partielle** (10 occurrences sur `service-public.fr`)

Nous avons également exploré manuellement le terme **allocation familiale** qui n'apparaît pas dans le tableau car il n'a pas été détecté par le logiciel, mais il nous semble important dans le domaine.

— **allocation familiale**(469 occurrences, tous sites)

Une grande partie des occurrences relevées sont en fait issues de « caisse d'allocations familiales » (197 occurrences) ou encore de « caisse nationale des allocations familiales » (220 occurrences). On trouve donc 469 occurrences faisant référence au dispositif. On ne retrouve pas de variants de cette aide. Elle est attribuée automatiquement et ne pose donc pas de difficulté particulière pour les usagers. Le sigle **Af** est parfois utilisé pour faire référence à ce dispositif. Toutefois, ces occurrences sont uniquement présentes sur le site `caf.fr` et elles sont plutôt rares (24 occurrences du sigle contre 469 du terme complet).

— **Af** (24 occurrences sur `caf.fr`)

5.1.1.1.2 Analyse des patrons lexico-syntaxiques avec le noyau « aide »

La Table 5.3 ci-dessous présente les résultats les plus fréquents pour le patron syntaxiques :

[**aide**][préposition][nom commun]

word	Frequency	Freq. per million	% of concordance
aide au logement	317	420,76481	46,75516
aide à domicile	144	191,13606	21,23894
aide au mérite	52	69,02136	7,66962
aide au retour	38	50,43868	5,60472
aide au recouvrement	38	50,43868	5,60472
aide à mi-taux	14	18,58267	2,0649
aide au permis	10	13,27334	1,47493
aide en cas	7	9,29134	1,03245
aide du FSL	6	7,964	0,88496
aide au fonctionnement	6	7,964	0,88496

TABLE 5.3 – Résultats les plus fréquents pour le patron [**aide**][préposition][nom commun]

Les relevés présentés dans la Table 5.3 mettent en avant un premier dispositif formé à partir du nom noyau **aide** : l'**aide au logement**. Les autres relevés ne sont pas clairement des dispositifs d'aide. Nous avons donc décidé de faire une exploration supplémentaire, afin d'identifier d'autres dispositifs qui pourraient ne pas correspondre au schéma choisi. La Table 5.4 présente les collocats les plus fréquents du terme **aide**.

Collocate	Freq	Coll. freq.	T-score	MI	logDice
personnelle	302	324	17,28613	7,56122	11,25755
au	637	6741	23,92072	4,25907	10,96269
sociale	291	1276	16,68956	5,53013	10,8989
adaptées	159	170	12,54298	7,56616	10,38807
personnelles	91	135	9,46955	7,09364	9,59603
à	422	16124	16,66895	2,40684	9,44483
financière	71	322	8,23755	5,4815	9,16961
Habiter	64	76	7,95312	7,41474	9,1105
personnalisée	59	97	7,61882	6,9454	8,98518

TABLE 5.4 – Collocats relevés automatiquement du noyau **aide**

À partir de ces relevés, nous avons exploré manuellement en corpus les différentes occurrences afin d'identifier les différents dispositifs. Comme précédemment indiqué,

cela signifie que le nombre d'occurrences relevé automatiquement par le logiciel et présenté dans les Tables peut être différent du nombre d'occurrences relevé ensuite manuellement et indiqué dans les commentaires.

Nous avons commencé par le premier collocat identifié, **aide personnelle**, qui nous a permis d'identifier le terme suivant :

— **aide personnelle au logement** (393 occurrences sur `caf.fr`)

Ce terme n'apparaît pas sur les autres sites, notamment sur `service-public.fr` qui est un site général qui recense l'ensemble des aides disponibles. Il ne s'agit pas d'un dispositif, mais d'un hyperonyme rassemblant trois dispositifs d'aide au logement (l'**aide personnalisée au logement**, l'**allocation de logement sociale** et l'**allocation de logement familiale**).

Nous avons ensuite exploré les différents dispositifs présents à partir du relevé **aide au** qui nous a permis de repérer quatre dispositifs. Tout d'abord, nous avons identifié **aide au retour à l'emploi** qui est un variant du terme **allocation d'aide au retour à l'emploi**. Il s'agit d'une utilisation dérivée du terme juridique initial qui a été raccourci, soit par erreur, soit par principe d'économie du langage (Martinet, 1969).

— **aide au retour à l'emploi** (9 occurrences sur `service-public.fr`)

Nous avons ensuite identifié les trois dispositifs suivants :

— **aide au logement** (317 occurrences, tous sites)
— **aide au recouvrement** (68 occurrences, dont 6 occurrences sur `service-public.fr` et 62 occurrences sur `caf.fr`)
— **aide au mérite** (60 occurrences, dont 7 occurrences sur `crous-paris.fr` et 53 occurrences sur `service-public.fr`)

Le premier terme proposé (**aide au logement**) a déjà été relevé grâce au premier patron choisi dans la Table 5.3. Il s'agit d'une catégorie d'aide financière et non d'un dispositif spécifique. Dans certains cas, il s'agit d'un concurrent d'**aide personnelle au logement**.

Le dispositif suivant, **aide au recouvrement**, est un dispositif juridique, tandis que le dernier dispositif relevé pour ce collocat, **aide au mérite**, est attribué par le Crous, un organisme versant des aides sociales destinées spécifiquement aux personnes ayant le statut étudiant.

Nous avons ensuite exploré les compléments des relevés « Aide à », « aide financière » et « aide Habiter ». Ces relevés ne correspondent à aucun dispositif d'aide sociale

en corpus.

Enfin, le dernier terme relevé qui nous a intéressé est le suivant : **aide personnalisée**. Dans le corpus, nous avons identifié deux dispositifs formés à partir de ce noyau :

— **aide personnalisée au logement** (61 occurrences, 12 occurrences sur `service-public.fr`, 46 occurrences sur `caf.fr`, 2 sur occurrences sur `crous-paris.fr`, 1 occurrence sur `paris.fr`)
— **aide personnalisée d'autonomie** (3 occurrences sur `paris.fr`)
— **Apa** (203 occurrences, 117 sur `service-public.fr` et 86 occurrences sur `paris.fr`)

Le premier dispositif (**aide personnalisée au logement**) est versé par la Caf et fait partie de l'ensemble des aides consacrées aux logements. Il s'agit d'un type d'**aide personnelle au logement** ou d'**aide au logement**.

Le second terme relevé est beaucoup moins présent en corpus (3 occurrences), mais dispose d'un acronyme (**Apa**) qui compte 203 occurrences et qui semble donc être le terme utilisé par défaut, soit le terme vedette. Nous abordons les sigles et acronymes en corpus dans la section 5.1.4.

Le noyau **aide** est donc bien utilisé pour former des dénominations de dispositifs. Toutefois, ce noyau semble apparaître dans des contextes très divers et avoir plusieurs sens en corpus, si bien qu'une exploration supplémentaire serait nécessaire pour identifier l'ensemble des dénominations correspondant aux dispositifs d'aide. Nous pouvons remarquer à ce stade que les explorations automatiques en corpus à l'aide de la fonction « collocation » de Sketch Engine permettent d'obtenir des relevés facilitant l'identification d'un nombre de dispositifs. Toutefois, dans la plupart des cas, une exploration plus poussée, semi-automatique ou semi-manuelle, est nécessaire pour identifier l'ensemble des dispositifs.

5.1.1.1.3 Analyse des patrons lexico-syntaxiques avec le noyau « revenu »

Le nom noyau suivant est **revenu** qui est un terme qui prend plusieurs sens dans notre corpus. Le sens le plus utilisé provient du domaine de la finance :

> « ce qu'un particulier reçoit à des titres divers (rémunération, pension, rente, etc.). » (Termium)

Toutefois, nous avons relevé une acception différente lorsqu'il devient un concurrent d'**allocation**. D'après nos explorations terminologiques, il n'existe qu'un seul dispositif utilisant ce noyau : **revenu de solidarité active** avec 142 occurrences. Ce dernier est

beaucoup plus utilisé par son sigle **RSA** qui compte 1 466 occurrences.

— **revenu de solidarité active** (142 occurrences, dont 40 occurrences sur `service-public.fr`, 84 occurrences sur `caf.fr` et 18 occurrences sur `paris.fr`)

— **RSA** (1 466 occurrences, tous sites)

Le choix de cette dénomination (**revenu de solidarité active**) semble s'inspirer de son précurseur le **revenu minimum d'insertion** et reflète l'influence politique dans la formation des noms pour les dispositifs d'aide sociale. En effet, le **revenu de solidarité active** est particulièrement concerné par le caractère politique du discours, car il est sujet à de nombreux débats dans les médias. Le choix de ce noyau isole ce dispositif qui prend alors une place spécifique dans le discours. Il est toutefois intéressant de noter que l'**allocation d'aide au retour à l'emploi** est parfois qualifiée de revenu de substitution.

Ces explorations révèlent les premières observations du phénomène de variation terminologique présent dans le discours. Certains concepts sont instables et disposent de plusieurs variants pour désigner le même dispositif ou le même concept, ce qui peut induire des difficultés de compréhension pour les usagers. Ces difficultés sont causées par la multiplicité des noyaux (aide, allocation, revenu…), par le recours au pléonasme (allocation d'aide) et la profusion d'extensions (de solidarité active, retour à l'emploi, minimum d'insertion).

5.1.1.1.4 Analyse des patrons lexico-syntaxiques avec le noyau « complément »

Nous allons maintenant nous intéresser au nom noyau **complément**. Pour explorer ce nom noyau, nous avons combiné les deux méthodologies utilisées pour faire les relevés précédents : le patron syntaxique et les collocats.

Les Tables 5.6 et 5.5 présentent les relevés qui nous ont permis de faire les analyses autour du noyau **complément**.

word	Frequency	Freq. per million	% of concordance
complément de ressources	58	76,98536	37,9085
complément d'Aeeh	24	31,85601	15,68627
complément d'Asf	20	26,54668	13,0719
complément en fonction	12	15,92801	7,84314
complément de fin	12	15,92801	7,84314
complément pour frais	6	7,964	3,92157
complément à taux	5	6,63667	3,26797
complément de revenu	4	5,30934	2,61438

TABLE 5.5 – Résultats les plus fréquents pour le patron [**complément**][préposition][nom commun]

Collocate	Freq	Coll. freq.	T-score	MI	logDice
familial	126	530	11,17302	7,75518	11,56895
Santé	53	79	7,26817	9,25189	10,90137
éventuel	16	25	3,99312	9,18389	9,26191
libre	19	248	4,29629	6,12148	9,17512
mode	16	190	3,94773	6,25789	9,00706
accordé	12	118	3,42662	6,53007	8,69774
mensuel	13	259	3,52651	5,51138	8,61298

TABLE 5.6 – Collocats relevés automatiquement du nom noyau **complément**

Suivant les résultats présentés dans la Table 5.5, nous avons tout d'abord exploré le dispositif **complément de ressources** (59 occurrences). Il s'agit d'un dispositif spécifique qui est donné en corrélation avec l'**allocation adulte handicapé** et non d'un concept général de complément comme sa dénomination pourrait suggérer. Il ne dispose pas de sigle correspondant.

— **complément de ressources** (59 occurrences, 9 sur `service-public.fr`, 48 sur `caf.fr` et 2 sur `paris.fr`)

Les relevés obtenus grâce au patron syntaxique (Table 5.5) montrent qu'il existe de nombreux dispositifs qui complètent certaines allocations ((**complément d'Aeeh** et **complément d'Asf** par exemple). Toutefois, ces relevés ne nous permettent pas d'identifier des dispositifs formés à partir de **complément**. Nous nous sommes donc basée sur la Table 5.6 afin de faire les relevés suivants.

Le premier exemple de collocat associé au nom noyau **complément** est **complément familial** avec 126 occurrences. Il s'agit d'une prestation versée en complément des **allocations familiales** selon les revenus de l'usager. Dans ce cas, le complément « **familial** » indique une parenté entre les deux aides.

— **complément familial** (126 occurrences, dont 23 occurrences sur `service-public.fr` et 103 occurrences sur `caf.fr`)

Le deuxième exemple est **complément Santé Paris**, avec 53 occurrences. Il s'agit d'une aide financière versée par la ville de Paris aux séniors et aux personnes en situation de handicap afin de souscrire à une mutuelle.

— **complément Santé Paris** (53 occurrences sur `paris.fr`)

Pour ce qui est du terme relevé **complément libre**, il est utilisé pour former les deux dénominations suivantes :

— **complément libre choix de mode de garde** (11 occurrences)

— **complément libre choix d'activité** (8 occurrences)

Ces deux dispositifs sont en lien avec l'**allocation familiale** et la **prestation d'accueil du jeune enfant** mais restent des aides indépendantes.

Nous avons également trouvé des occurrences de **complément mode de garde**, avec 16 occurrences. Il s'agit d'une variation terminologique pour le dispositif suivant : **complément libre choix de mode de garde**, formé dans un principe d'économie du langage dans le but de simplifier son appellation. Le sigle **CMG** est aussi présent.

— **complément mode de garde** (16 occurrences sur `caf.fr`)
— **CMG** (196 occurrences dont 109 sur `service-public.fr` et 87 sur `caf.fr`)

L'utilisation du noyau **complément** peut donc avoir deux significations en corpus. Tout d'abord, il peut s'agir d'un complément d'un dispositif existant. Ensuite, il peut s'agir d'un dispositif à part entière, comme dans **complément libre choix de mode de garde**. La différence d'utilisation des noyaux est, dans ce cas, idéologique. Dans le cas des dispositifs formés à partir de **complément**, on considère qu'il s'agit d'une aide additionnelle destinée à financer une partie d'un service et non de le financer entièrement. Cela peut être la différence avec le dispositif formé à partir du noyau **revenu** qui, lui, est destiné à représenter la majeure partie des revenus de l'usager.

5.1.1.1.5 Analyse des patrons lexico-syntaxiques avec le noyau « prestation »

Pour analyser l'usage du nom noyau **prestation**, nous avons procédé avec la même méthode. Nous avons relevé à la fois les occurrences du patron syntaxique présenté au début de cette section en utilisant le noyau **prestation** et les collocats de **prestation** (Tables 5.7 et 5.8 respectivement).

word	Frequency	Freq. per million	% of concordance
prestation de compensation	80	106,1867	27,39726
prestation d'accueil	59	78,31269	20,20548
prestation de logement	21	27,87401	7,19178
prestations de soutien	18	23,89201	6,16438
prestation de service	18	23,89201	6,16438
prestations en nature	15	19,91001	5,13699
prestation d'aide	14	18,58267	4,79452
prestations de service	10	13,27334	3,42466
prestation de soutien	5	6,63667	1,71233

TABLE 5.7 – Résultats les plus fréquents pour le patron [**prestation**][préposition][nom commun]

Collocate	Freq	Coll. freq.	T-score	MI	logDice
Handicap	342	359	18,438	8,3869	12,12841
familiales	396	1325	19,71026	6,71447	11,86905
partagée	144	144	11,96585	8,45689	11,01005
n’	50	1669	6,39934	3,39597	8,74715
sociales	33	388	5,55234	4,90137	8,73833
Pour	44	1405	6,03045	3,45996	8,66623
suivantes	29	486	5,12833	4,39006	8,49713

TABLE 5.8 – Collocats relevés automatiquement du nom noyau **prestation**

Le premier dispositif que nous avons étudié a été mis en lumière dans la Table 5.7. Il s’agit du premier relevé (**prestation de compensation**) dont le terme complet est **prestation de compensation du handicap** avec 73 occurrences et qui correspond au sigle **PCH**. Le sigle semble beaucoup plus utilisé, notamment sur le site `service-public.fr`.

— **prestation de compensation du handicap** (73 occurrences, 14 sur `service-public.fr`, 24 sur `caf.fr` et 35 sur `paris.fr`)
— **PCH** (333 occurrences, 308 sur `service-public.fr`, 16 occurrences sur `caf.fr` et 9 occurrences sur `paris.fr`)

Le second exemple relevé est **prestation d’accueil** dont le terme complet est **prestation d’accueil du jeune enfant** avec 72 occurrences. Ce terme correspond à l’acronyme **Paje**. Ce dispositif est similaire au dispositif **prestation partagée d’éducation de l’enfant** dont l’acronyme est **PrePare**. Cet acronyme semble déroger les règles de formation, en utilisant plusieurs lettres de chaque mot formant le dispositif. Ils concernent tous les deux les familles avec enfants. Dans les deux cas, les acronymes sont plus utilisés que le terme initial.

— **prestation d’accueil du jeune enfant** (72 occurrences, 18 occurrences sur `service-public.fr`, 54 occurrences sur `caf.fr`)
— **Paje** (170 occurrences, 44 sur `service-public.fr`, 126 sur `caf.fr`)
— **prestation partagée d’éducation de l’enfant** (143 occurrences, 22 occurrences sur `service-public.fr`, 121 occurrences sur `caf.fr`)
— **PreParE** (377 occurrences, 97 sur `service-public.fr` et 280 occurrences sur `caf.fr`)

Dans certains cas, le nom noyau **prestation** entre dans la construction de termes hyperonymes qui désignent des types d’aide financière pour lesquelles il existe plusieurs sous-types ou sous-catégories. C’est notamment le cas avec des termes désignant les aides suivantes :

— **prestations de logement** (21 occurrences)
— **prestations Handicap** (342 occurrences)
— **prestations familiales** (396 occurrences)
— **prestations sociales** (43 occurrences)

Chacun de ces termes hyperonymes désigne un ensemble de dispositifs sur le thème indiqué (logement, handicap, familial, social, etc.). Les **prestations de logement** rassemblent par exemple :

— l'**aide personnalisée au logement**,
— l'**allocation de logement sociale**,
— et l'**allocation de logement familiale**.

Les **prestations logement** peuvent également inclure des aides annexes comme la **prime de déménagement**. Nous avons précédemment identifié un variant de **prestations logement** dans nos recherches : **aides au logement**. Nous retrouvons donc deux concurrents utilisant deux noyaux différents (**aide** et **prestation**). Le terme **aides personnelles au logement** que nous avons également identifié précédemment n'est pas un variant car il désigne une catégorie d'aides plus restreinte.

5.1.1.1.6 Analyse des patrons lexico-syntaxiques avec le noyau « prime »

Nous avons ensuite exploré le nom noyau **prime** grâce au patron et aux collocats relevés automatiquement par Sketch Engine. Les Tables 5.9 et 5.10 ci-dessous présentent l'ensemble de ces relevés. Pour rappel, la disparité entre les relevés de les Tables et nos relevés est due au type d'exploration (automatique ou manuelle) et à la prise en compte de la majuscule.

word	Frequency	Freq. per million	% of concordance
prime d'activité	288	382,27213	73,65729
prime de déménagement	78	103,53204	19,94885
prime d'équipement	10	13,27334	2,55754
prime de Noël	5	6,63667	1,27877
prime au déménagement	4	5,30934	1,02302
prime de précarité	3	3,982	0,76726
prime d'adoption	1	1,32733	0,25575
prime de reprise	1	1,32733	0,25575
Prime de déménagement	1	1,32733	0,25575

TABLE 5.9 – Résultats les plus fréquents pour le patron [**prime**][préposition][nom commun]

Collocate	Freq	Coll. freq.	T-score	MI	logDice
exceptionnelle	270	413	16,41492	9,93771	12,22663
activité	760	3869	27,49274	8,2049	12,19943
déménagement	93	312	9,62208	8,80465	10,77015
d’	775	44577	26,77143	4,70493	9,10839
versée	37	1359	5,93383	5,35201	8,76237
aux	78	4604	8,48426	4,66762	8,72579

TABLE 5.10 – Collocats relevés automatiquement du nom noyau **prime**

La Table 5.10 présente des collocats qui ne font pas partie des dénominations de dispositifs ou qui répètent les résultats de la Table 5.9. Le seul terme relevé unique est **prime exceptionnelle**.

— **prime exceptionnelle** (270 occurrences, dont 234 sur `caf.fr` et 36 sur `caf.fr`)

Une **prime exceptionnelle** est un type de prime qui est versée exceptionnellement au bénéficiaire, elle ne fait pas référence à un dispositif en particulier mais à un type de dispositif.

Le premier terme relevé, **prime d’activité**, présente 760 occurrences et il est présent sur les sites de la Caf et le site officiel de l’administration. Cette aide est attribuée par la Caf à une personne percevant un bas revenu afin de le compléter.

— **prime d’activité** (760 occurrences, 242 sur `service-public.fr` et 518 sur `caf.fr`)

Le deuxième exemple de terme, **prime de déménagement**, présente 89 occurrences et est une aide attribuée par la Caf. Il existe un variant de ce dispositif : la **prime au déménagement** dont on trouve 4 occurrences. Le variant présente une modification au niveau de la préposition et n’apparaît que sur un site (`caf.fr`). Le terme juridique initial est **prime de déménagement** et l’on peut donc supposer qu’il s’agit d’un terme formé par dérivation.

— **prime de déménagement** (89 occurrences dont 21 sur `service-public.fr` et 68 sur `caf.fr`)

— **prime au déménagement** (4 occurrences sur `caf.fr`)

Le troisième terme relevé, **prime d’équipement**, désigne une aide destinée aux lycéens ou aux étudiants afin d’acheter les équipements nécessaires à leurs études. On ne retrouve que 10 occurrences en corpus.

— **prime d’équipement** (10 occurrences sur `service-public.fr`)

Le terme suivant, **prime de précarité**, n'est pas un dispositif d'aide sociale, mais une prime versée par l'employeur à la fin d'un contrat.

— **prime de précarité** (3 occurrences sur `service-public.fr`)

Enfin, la **prime d'adoption** fait partie de la prestation d'accueil du jeune enfant (Paje), tout comme la **prime à la naissance** (80 occurrences), qui n'apparaît pas dans nos relevés en raison du patron syntaxique utilisé dans Sketch Engine ([**prime**][préposition][nom commun]).

— **prime d'adoption** (1 occurrence sur `service-public.fr`)
— **prime à la naissance** (80 occurrences dont 26 sur `service-public.fr` et 54 sur `caf.fr`)

Il est intéressant de noter que le noyau **prime** montre une préférence sémantique pour désigner les aides liées aux situations familiales, sociales et professionnelles transitoires spécifiques. Dans certains cas, comme pour la **prime d'activité** par exemple, il est utilisé comme **complément** pour désigner une aide qui est censée venir compléter financièrement une somme déjà détenue par l'usager (dans ce cas, son salaire).

5.1.1.1.7 Analyse des patrons lexico-syntaxiques avec le noyau « bourse »

Le dernier nom noyau auquel nous nous intéressons est **bourse**. Pour faire nos relevés, nous utilisons, comme pour les noms noyaux précédents, le patron et les collocats. Les résultats sont présentés dans les deux Tables 5.11 et 5.12 ci-dessous.

word	Frequency	Freq. per million	% of concordance
bourse de lycée	36	47,78402	24,82759
bourse sur critères	32	42,47468	22,06897
bourse de collège	17	22,56467	11,72414
bourse des collèges	15	19,91001	10,34483
bourse au mérite	14	18,58267	9,65517
bourses sur critères	5	6,63667	3,44828
bourse d'enseignement	3	3,982	2,06897
bourses d'études	2	2,65467	1,37931
bourse d'été	2	2,65467	1,37931
bourse de fréquentation	2	2,65467	1,37931
bourse sur critère	2	2,65467	1,37931

TABLE 5.11 – Résultats les plus fréquents pour le patron [**bourse**][préposition][nom commun]

Collocate	Freq	Coll. freq.	T-score	MI	logDice
Talents	22	27	4,68737	10,58728	10,72472
sur	47	2921	6,62999	4,92507	8,85762
est-elle	6	49	2,43889	7,85298	8,77761
Erasmus	3	6	1,73022	9,88273	7,92318
2023-2024	3	44	1,7186	7,00826	7,7938
nationale	5	424	2,13564	4,47674	7,63718
est	23	5966	4,13698	2,86374	6,88762

TABLE 5.12 – Collocats relevés automatiquement du nom noyau **bourse**

Le premier terme relevé présenté dans la Table 5.11 est **bourse de lycée** avec 36 occurrences. Il s'agit d'une prestation versée par le ministère de l'Éducation aux lycéens. Ce dispositif ne dispose pas de sigle ou d'acronyme.

— **bourse de lycée** (36 occurrences sur `service-public.fr`)

Le deuxième terme relevé est **bourse sur critères sociaux**, avec 45 occurrences. Il s'agit d'une bourse versée par le Crous. Elle dispose d'un sigle : **BCS**, dont on ne relève que 5 occurrences. Ce sigle n'apparaît que sur le site `service-public.fr`, et non sur le site du Crous, qui est l'organisme responsable de ce dispositif. Ce dispositif est également doté d'un variant : **bourse d'enseignement supérieur sur critères sociaux**. Ce variant est en réalité le terme juridique présent dans les textes de loi. Le terme le plus utilisé, **bourse sur critères sociaux**, est donc un variant modifié et raccourci pour simplifier la communication.

— **bourse sur critères sociaux** (45 occurrences, 33 occurrences sur `service-public.fr` et 12 occurrences sur `crous-paris.fr`)
— **BCS** (5 occurrences sur `service-public.fr`)
— **bourse d'enseignement supérieur sur critères sociaux** (2 occurrences, 1 sur `service-public.fr` et 1 sur `crous-paris.fr`)

Le troisième terme relevé est **bourse de/des collèges** avec 31 occurrences. Ce terme semble avoir deux variants : **bourse de collège** ou **bourse des collèges**. Ces deux variants sont utilisés de la même façon et à la même fréquence. Le dispositif est similaire à la **bourse de lycée** et ne présente pas de sigle ou d'acronyme.

— **bourse de collège** (17 occurrences sur `service-public.fr`)
— **bourse des collèges** (15 occurrences sur `service-public.fr`)

Le quatrième terme relevé, **bourse au mérite**, compte 16 occurrences et n'a pas de sigle ou d'acronyme. Il s'agit également d'une aide financière réservée à des lycéens et basée sur certains critères de réussite. Les termes **aide au mérite** et **bourse au mérite**

sont similaires mais n'ont pas le même public cible. L'**aide** est destinée aux étudiants du supérieur tandis que la **bourse** est destinée aux lycéens. Cette différence de noyau ne semble pas répondre à des critères sémantiques, mais plutôt à un besoin de distinguer les deux aides.

— **bourse au mérite** (16 occurrences sur `service-public.fr`)

Le terme **bourse d'étude**, qui compte 3 occurrences, n'est pas un dispositif spécifique mais plutôt un hypéronyme désignant différents types de bourses : **bourse au mérite**, **bourse sur critères sociaux**, etc.

— **bourse d'étude** (3 occurrences sur `service-public.fr`)

Le dernier terme relevé, **bourse d'été** qui compte 3 occurrences, est un volet complémentaire de la **bourse sur critères sociaux**.

— **bourse d'été** (3 occurrences sur `service-public.fr`)

Enfin, la Table 5.12 met en avant deux autres termes. Le premier est **bourse Talents** avec 22 occurrences. Il s'agit d'une bourse spécifique aux concours de la fonction publique. Il n'existe pas d'acronyme ou de sigle correspondant.

— **bourse Talents** (22 occurrences sur `service-public.fr`)

Le deuxième terme issu de la Table 5.12 est **bourse Erasmus** avec 3 occurrences. Il s'agit d'une bourse spécifique aux étudiants réalisant un échange Erasmus.

— **bourse Erasmus** (3 occurrences sur `service-public.fr`)

Il ne semble pas y avoir de critères spécifiques pour la formation des dénominations des dispositifs à partir de **bourse**. Néanmoins, nous pouvons noter la spécialisation de ce noyau pour les aides destinées aux élèves et étudiants.

5.1.1.1.8 Synthèse des résultats de l'exploration du corpus AIDESOC-FR

En conclusion, le nom noyau **allocation** est utilisé presque uniquement pour former des dénominations de prestations sociales. Le nom noyau **aide** est également largement utilisé pour les dénominations de prestations sociales, mais il reste un élément polysémique du discours de l'accès aux droits. Le nom noyau **complément**, comme il l'indique, désigne parfois des dispositifs versés en complément d'autres dispositifs. Toutefois, ce n'est pas toujours le cas.

Le nom noyau **prestation** est en même temps un hyperonyme qui désigne l'ensemble

des dispositifs dans un domaine, par exemple dans **prestations familiales**, mais aussi un nom noyau utilisé pour former les noms de ces dispositifs, comme dans **prestation d'accueil du jeune enfant**.

Le nom noyau **prime** peut indiquer une somme d'argent versée en une fois, comme la **prime de déménagement** ou la **prime à la naissance**. Toutefois, la **prime d'activité** semble faire figure d'exception, car il s'agit ici d'une aide versée tous les mois et réévaluée tous les trois mois.

Enfin, le nom noyau **bourse** est réservé aux dispositifs qui concernent les études. Les dispositifs relevés ne semblent pas correspondre aux mêmes caractéristiques que les autres dispositifs et sont réservés aux étudiants. Le nom **bourse** n'est donc pas un nom noyau comme les autres, même s'il est utilisé pour former des dénominations de dispositifs d'aide.

5.1.1.2 Terminologie dans le corpus anglais sur le thème des aides sociales

Les explorations présentées dans cette section ont été faites sur le corpus BEN-AN qui contient l'ensemble des pages dédiées aux aides sociales sur les sites choisis. Une partie de ces explorations a donné lieu à une étude pilote dans laquelle nous avons identifié et analysé tous les noms noyaux représentant les dispositifs d'aide sociale à partir du noyau clé *benefits* (Bouchet, 2020). Les noyaux relevés sont les suivants :

— benefit,
— credit,
— allowance,
— payment.

Notre étude pilote a permis d'identifier que les dispositifs ne présentent pas toujours des structures nominales complexes et qu'il est donc nécessaire de faire une recherche à partir des noms noyaux et non d'une structure lexicale. L'expression CQL[2] utilisée est la suivante :

```
[tag="NP.*"] ([lemma="credit"] | [lemma="allowance"] | [lemma="payment"]
| [lemma="benefit"])
```

Elle correspond au patron syntaxique suivant : [nom][nom noyau].

Rappelons que les explorations sont initialement réalisées à partir d'étiquettes syntaxiques attribuées automatiquement par le logiciel Sketch Engine. Une fois les exemples relevés, nous les explorons manuellement en corpus. C'est pourquoi le nombre d'occurrences relevé dans les tableaux initiaux n'est pas toujours le même que celui relevé ensuite. L'étiquetage automatique n'est pas toujours précis et l'utilisation des majuscules peut fausser le nombre d'occurrences ressorti par le logiciel.

2. Corpus Query Language

5.1.1.2.1 Analyse des patrons lexico-syntaxiques avec le noyau « *credit* »

Nous avons commencé par une exploration des noms de dispositifs formés à partir de *credit*. Ils sont présentés dans la Table 5.13 ci-dessous.

word	Frequency	Freq. per million	% of concordance
Universal Credit	2133	4298,66123	70,30323
Pension Credit	358	721,48182	11,7996
Tax Credit	309	622,73151	10,18457
tax credit	70	141,07186	2,30719
Guarantee Credit	34	68,52062	1,12063
pension credit	23	46,35218	0,75808
Carer's Credit	23	46,35218	0,75808
Savings Credit	18	36,27562	0,59328
universal credit	15	30,22969	0,4944
temporary credit	11	22,16844	0,36256

TABLE 5.13 – Résultats les plus fréquents pour le patron ([nom] ou [adjectif]) [*credit*]

Le premier terme et le plus fréquent est *Universal Credit*, avec 2 148 occurrences. Il s'agit d'un nouveau dispositif qui remplace plusieurs anciens dispositifs. Il présente ainsi une fréquence particulièrement élevée, la plus haute sur l'ensemble des termes analysés, anglais ou français. Il dispose d'un sigle *UC*, qui est utilisé sur deux sites seulement. Il semble donc que l'utilisation du sigle n'est pas généralisée. Cela peut venir de la nouveauté du dispositif, ou encore du fait qu'il ne se compose que de deux éléments.

— *Universal Credit* (2 148 occurrences, tous sites)
— *UC* (32 occurrences, dont 1 sur `gov.uk` et 31 sur `citizensadvice.org.uk`)

Le deuxième terme relevé est *Tax Credit*, qui est très présent dans le corpus avec 1 108 occurrences. Néanmoins, il ne s'agit pas vraiment d'un dispositif d'aide, mais d'une somme qui peut être déduite des impôts d'un usager pour diverses raisons (enfants, dons, etc.). Il s'agit donc d'une acception différente du terme *credit*.

— *Tax Credit* (1 108 occurrences, tous sites)

Le troisième terme relevé est *Pension Credit*. Il s'agit du système de retraite britannique. Il n'existe pas de sigle correspondant.

— *Pension Credit* (379 occurrences, dont 101 occurrences sur `gov.uk`, 21 sur `bristol.gov.uk` 4 sur `understanduniversalcredit.gov.uk` et 253 sur `citizensadvice.org.uk`)

Enfin, le terme *Carer's Credit* propose une nouvelle utilisation du terme *credit*, puisqu'il s'agit ici de cotiser à la *National Health Insurance*. Il s'agit donc d'une aide financière pour laquelle on verse des cotisations.

— *Carer's Credit* (25 occurrences dont 24 occurrences sur `gov.uk` et 1 sur `citizensadvice.org.uk`)

Le nom noyau *credit* prend différents sens dans les termes relevés en corpus. Il peut s'agit d'un concurrent de *benefit* dans certains cas, mais il peut également garder son sens initial, par exemple dans le terme *Tax Credit*. Ce nom noyau présente une polysémie.

5.1.1.2.2 Analyse des patrons lexico-syntaxiques avec le noyau « *allowance* »

Nous allons maintenant étudier le nom noyau *allowance*. La Table 5.14 ci-dessous présente les occurrences les plus fréquentes des groupes nominaux formés à partir de *allowance*.

word	Frequency	Freq. per million	% of concordance
Jobseeker's Allowance	390	794,03306	23,55051
Support Allowance	282	568,31808	16,85595
Carer's Allowance	277	558,24152	16,55708
Attendance Allowance	264	532,04246	15,78004
Disablement Allowance	64	128,97999	3,82546
Maternity Allowance	58	116,88812	3,46683
Housing Allowance	33	66,50531	1,9725
Seekers Allowance	33	66,50531	1,9725
Parent's Allowance	30	60,45937	1,79319
Jobseekers Allowance	22	44,33687	1,315
Maintenance Allowance	20	40,30625	1,19546
Guardian's Allowance	19	38,29093	1,13568
Work Allowance	17	34,26031	1,01614
Jobseeker's allowance	16	32,245	0,95637
standard allowance	15	30,22969	0,89659
Earnings Allowance	11	22,16844	0,6575

TABLE 5.14 – Résultats les plus fréquents pour le patron ([nom] ou [adjectif]) [*allowance*]

Le premier terme présenté est le plus fréquent, avec 390 occurrences. Il s'agit de *Jobseeker's Allowance*, qui fait l'objet d'une étude spécifique dans la section 5.1.3.1. Il s'agit d'un terme désignant une aide financière pour les personnes au chômage.

— *Jobseeker's Allowance* (390 occurrences, 263 occurrences sur `gov.uk`, 22 occurrences sur `bristol.gov.uk`, 105 occurrences sur `citizensadvice.`

org.uk)

Le deuxième terme, *Support Allowance*, correspond en réalité à *Employment and Support Allowance* avec 286 occurrences. Ce terme est également utilisé par son sigle *ESA*.

— *Employment and Support Allowance* (390 occurrences, 263 occurrences sur gov.uk, 22 occurrences sur bristol.gov.uk, 105 occurrences sur citizensadvice.org.uk)
— *ESA* (889 occurrences, tous sites)

Le troisième terme *Carer's Allowance* montre que certains compléments choisis pour former des termes complexes (ici *carer*) sont réutilisés avec plusieurs noms de dispositifs. Ce terme correspond bien à un dispositif d'aide financière, contrairement à *Carer's Credit*.

— *Carer's Allowance* (287 occurrences, 109 occurrences sur gov.uk, 4 sur bristol.gov.uk, 167 occurrences sur citizensadvice.org.uk et 4 occurrences sur understandinguniversalcredit.gov.uk)

Le terme suivant est *Attendance Allowance*. Il s'agit d'une aide financière pour les personnes en situation de handicap, afin de financer des aides à domicile. Le terme *attendance* juxtaposé au terme noyau fait donc référence à la personne engagée.

— *Attendance Allowance* (271 occurrences, 118 occurrences sur gov.uk, 7 sur bristol.gov.uk, 143 occurrences sur citizensadvice.org.uk et 3 occurrences sur understandinguniversalcredit.gov.uk)

Le groupe nominal suivant est *Disablement Allowance* qui correspond en réalité à *Severe Disablement Allowance* avec 65 occurrences. Il s'agit d'un ancien dispositif maintenant remplacé par *Employment and Support Allowance*.

— *Severe Disablement Allowance* (65 occurrences, 52 occurrences sur gov.uk, 4 sur bristol.gov.uk et 9 occurrences sur citizensadvice.org.uk)

Le terme suivant, *Maternity Allowance*, compte 58 occurrences et possède un équivalent en corpus qui est *Maternity Pay*. Les deux dispositif versent une aide financière à une jeune mère. Le dispositif est différent suivant le statut de la mère bénéficiaire (salariée, auto-entrepreneuse, etc.) Dans ce cas, on retrouve à nouveau la réutilisation du terme juxtaposé *Maternity* pour nommer plusieurs dispositifs qui concernent les mêmes usagers.

— *Maternity Allowance* (59 occurrences, 26 occurrences sur `gov.uk`, 4 sur `bristol.gov.uk` et 25 occurrences sur `citizensadvice.org.uk`)
— *Maternity Pay* (187 occurrences, 27 occurrences sur `gov.uk`, 4 sur `bristol.gov.uk`, 1 occurrence sur `understanduniversalcredit.gov.uk`) et 154 occurrences sur `citizensadvice.org.uk`)

Le terme suivant, *Local Housing Allowance*, compte 38 occurrences. Il est versé par les municipalités et se trouve donc en majorité sur les sites des municipalités. Néanmoins, il est intéressant de noter que les sites d'accès aux droits gouvernementaux mentionnent également ce dispositif.

— *Local Housing Allowance* (38 occurrences, 3 occurrences sur `gov.uk`, 15 occurrences sur `bcpcouncil.gov.uk`, 9 sur `bristol.gov.uk`, 5 occurrences sur `understanduniversalcredit.gov.uk`) et 5 occurrences sur `citizensadvice.org.uk`)

Nous retrouvons ensuite 40 occurrences de *Widowed Parent's Allowance*. Le nom du dispositif décrit le statut de l'usager.

— *Widowed Parent's Allowance* (29 occurrences, 19 occurrences sur `gov.uk`, 2 sur `bristol.gov.uk`, 7 sur `citizensadvice.org.uk`)

Le terme suivant *Education Maintenance Allowance* compte 20 occurrences et a un sigle correspondant : *EMA*, qui compte 54 occurrences.

— *Education Maintenance Allowance* (28 occurrences sur `falkirk.gov.uk`)
— *EMA* (54 occurrences sur `falkirk.gov.uk`)

Le terme *Guardian's Allowance* compte 21 occurrences et est destiné aux personnes ayant un enfant à charge dont ils ne sont pas les parents, mais les tuteurs. Encore une fois, le terme juxtaposé *Guardian* décrit le statut de l'usager.

— *Guardian's Allowance* (21 occurrences, 12 sur `gov.uk`, 1 sur `falkirk.gov.uk`, 3 sur `citizensadvice.org.uk` et 4 sur `bristol.gov.uk`)

Les deux termes *work allowance* (23 occurrences) et *standard allowance* (18 occurrences) sont des composantes du dispositif *Universal Credit*. Ils constituent chacun un volet de l'aide financière.

— *work allowance* (25 occurrences, 3 occurrences sur `gov.uk`, 22 sur `understanduniversalcredit.gov.uk`)
— *standard allowance* (18 occurrences, 7 occurrences sur `gov.uk`, 1 sur

citizensadvice.org.uk, 2 sur falkirk.gov.uk et 7 sur understandinguniversalcredit.gov.uk)

Enfin, le terme *Reduced Earnings Allowance* compte 11 occurrences et a un sigle correspondant *REA* qui compte 5 occurrences.

— *Reduced Earnings Allowance* (11 occurrences dont 10 sur gov.uk et 1 sur citizensadvice.org.uk)
— *REA* (5 occurrences sur gov.uk)

Il existe donc un grand nombre de dispositifs utilisant le nom noyau *allowance*. D'après nos relevés, ce nom ne semble pas polysémique puisqu'on le retrouve spécifiquement dans la formation des termes nominaux désignant les divers dispositifs d'aide. Nous avons également observé que certains termes choisis pour être juxtaposés au nom noyau étaient réutilisés avec plusieurs noms noyaux (*Carer*,*Maternity*,...), comme nous allons le voir dans les sections suivantes.

5.1.1.2.3 Analyse des patrons lexico-syntaxiques avec le noyau « *benefit* »

Dans cette section, nous allons recenser les termes formés à partir du nom noyau *benefit*. Nous présentons dans la Table 5.15 les termes relevés à partir du patron présenté au début de cette section.

word	Frequency	Freq. per million	% of concordance
Housing Benefit	675	1360,33583	45,06008
Child Benefit	342	689,23682	22,83044
Disablement Benefit	71	143,08718	4,73965
Incapacity Benefit	67	135,02593	4,47263
housing benefit	56	112,85749	3,73832
child benefit	27	54,41343	1,8024
more Benefit	25	50,38281	1,66889
DWP benefit	20	40,30625	1,33511
other benefit	14	28,21437	0,93458
Tax Benefit	11	22,16844	0,73431
disability benefit	10	20,15312	0,66756
calculators Benefit	9	18,13781	0,6008

TABLE 5.15 – Résultats les plus fréquents pour le patron ([nom] ou [adjectif]) [*benefit*]

Le premier terme, *Housing Benefit*, compte 771 occurrences. Il n'existe pas de sigle correspondant. Il s'agit d'une aide financière pour le logement. Elle est progressivement remplacée par l'aide *Universal Credit*. Il existe également des occurrences de ce terme

au pluriel. Dans ce cas, il est présent dans les titres et est utilisé en tant qu'hyperonyme pour désigner les aides financières en lien avec le logement.

— *Housing Benefit* (771 occurrences, tous sites)
— *housing benefits* (39 occurrences, 38 sur `gov.uk` et 1 sur `bcpcouncil.gov.uk`)

Le deuxième terme relevé, *Child Benefit*, compte 375 occurrences. Il s'agit d'une aide financière réservée aux personnes ayant des enfants. Nous avons également trouvé 4 occurrences du pluriel (*children benefits*) qui se trouve dans des titres et qui a une fonction d'hyperonyme.

— *Child Benefit* (375 occurrences, dont 122 sur `gov.uk`, 1 sur `bcpcouncil.gov.uk`, 18 occurrences sur `bristol.gov.uk`, 165 sur `citizensadvice.org.uk` et 69 sur `falkirk.gov.uk`)
— *children benefits* (4 occurrences sur `gov.uk`)

Le terme relevé suivant, *Industrial Injuries Disablement Benefit*, compte 72 occurrences. La dénomination choisie pour ce terme est descriptive.

— *Industrial Injuries Disablement Benefit* (72 occurrences, 52 occurrences `gov.uk`, 3 sur `citizensadvice.org.uk`, 12 sur `falkirk.gov.uk`, 1 sur `understandinguniversalcredit.gov.uk` et 2 sur `bristol.gov.uk`)

Le terme relevé suivant, *Incapacity Benefit*, est une aide financière destinée aux personnes en situation de handicap.

— *Incapacity Benefit* (68 occurrences, 55 sur `gov.uk`, 4 sur `bristol.gov.uk`, 5 sur `citizensadvice.org.uk` et 1 occurrence sur `understandinguniversalcredit.gov.uk`)

Elle est souvent utilisée en lien avec *Severe Disablement Allowance*, comme dans l'exemple ci-dessous :

(4) If you get Incapacity Benefit or Severe Disablement Allowance

Cet exemple montre que différents noms noyaux peuvent être utilisés pour former des dispositifs d'aide similaires.

Les exemples suivants (*DWP benefit*[3], *other benefit* et *more benefit*) montrent que *benefit* peut être utilisé comme hyperonyme pour désigner une catégorie de dispositifs, comme **prestation** dans le corpus français.

3. Department of Work and Pensions, donc les aides délivrés par le DWP

— *DWP benefit(s)* (26 occurrences)
— *other benefit(s)* (226 occurrences)
— *more benefit(s)* (8 occurrences)

Le terme relevé suivant est le *Council Tax benefit* avec 11 occurrences. C'est un dispositif qui a été remplacé par le *Council Tax Reduction Scheme*. On remarque que dans ce cas, le nom noyau a été remplacé par un nom plus descriptif et plus clair, tandis que le terme juxtaposé *Council Tax* a été conservé.

— *Council Tax benefit*
— *Council Tax Reduction Scheme*(37 occurrences, 36 sur `bristol.gov.uk` et 2 sur `citizensadvice.org.uk`)

Le dernier groupe nominal analysé, *Carer's and disability benefit*, fait état de plusieurs aides financières. Il s'agit d'une catégorie d'aides dont fait partie le dispositif *Carer's allowance*. Le terme *benefit* serait alors un hyperonyme.

— *Carer's and disability benefit* (2 occurrences, `gov.uk`)

Les dispositifs formés à partir de *benefit* ne donne pas lieu à des sigles. Cela pourrait s'expliquer par la matrice lexicale choisie pour nommer les dispositifs d'aide en anglais avec ce noyau, étant donné qu'il s'agit majoritairement des noms composés court : N N. Le terme *benefit* est parfois un nom noyau, mais peut aussi être un nom plus général, ou un hyperonyme qui rassemble plusieurs dispositifs d'aide, comme dans *DWP benefits*.

5.1.1.2.4 Analyse des patrons lexico-syntaxiques avec le noyau « *payment* »

Pour conclure, nous allons étudier les dispositifs formés à partir de *payment*. La Table 5.16 ci-dessous présente les résultats de l'exploration du patron syntaxique sur Sketch Engine.

word	Frequency	Freq. per million	% of concordance
Independence Payment	243	489,7209	20,38591
Credit payment	138	278,1131	11,57718
first payment	81	163,2403	6,7953
Fuel Payment	70	141,07186	5,87248
Support Payment	55	110,84218	4,61409
advance payment	52	104,79624	4,36242
hardship payment	46	92,70437	3,85906
Weather Payment	40	80,61249	3,3557
Housing Payment	24	48,3675	2,01342
monthly payment	20	40,30625	1,67785
emergency payment	19	38,29093	1,59396
Alternative Payment	16	32,245	1,34228

TABLE 5.16 – Résultats les plus fréquents pour le patron ([nom] ou [adjectif]) [*payment*]

Le premier terme relevé par Sketch Engine est *Personal Independance Payment* qui compte 254 occurrences et qui est désigné par l'acronyme *PIP*, avec 676 occurrences. L'acronyme est donc beaucoup plus utilisé que le terme déballé.

— *Personal Independance Payment* (218 occurrences, 116 occurrences sur `gov.uk`, 1 sur `bcpcouncil.gov.uk`, 9 sur `bristol.gov.uk`, 63 occurrences sur `citizensadvice.org.uk`, 29 sur `falkirk.gov.uk`)
— *PIP* (676 occurrences, tous sites)

Le deuxième groupe nominal relevé est *Universal Credit payment*, avec 211 occurrences. Ici, *payment* fait référence au versement de l'aide financière comme dans l'exemple ci-dessous :

(5) You'll usually get your Universal Credit payment 7 days after each monthly assessment period ends.

Cette utilisation peut paraître surprenante, car le terme *payment* est polysémique. Parfois, il s'agit d'un nom noyau, tandis qu'il peut également faire référence au versement d'une aide, comme l'illustre le groupe nominal suivant *Universal Credit payment* ou encore le groupe nominal suivant dans la Table 5.16 : *first payment*.

Le groupe nominal suivant, *Support payment* qui compte 66 occurrences, est utilisé pour former différents dispositifs, comme le montre la Table 5.17 ci-dessous.

word	Frequency	Freq. per million	% of concordance
Funeral Support Payment	21	42,32156	38,18182
Bereavement Support Payment	16	32,245	29,09091
Trace Support Payment	12	24,18375	21,81818
Pandemic Support Payment	5	10,07656	9,09091
Carer Support Payment	1	2,01531	1,81818

TABLE 5.17 – Résultats les plus fréquents du patron [mot][*Support Payment*]

Nous retrouvons une méthode de construction des dispositifs similaires. Le dispositif *Carer Support Payment* est l'équivalent de *Carer's Allowance* en Écosse. Les termes *support payment* et *allowance* semblent donc être concurrents dans ce cas.

— *Carer Support Payment* (1 occurrence sur `gov.uk`)

Le groupe nominal *advance payment* n'est pas un dispositif. Il s'agit d'une avance qui peut être versée pour certains dispositifs.

— *advance payment* (55 occurrences, tous sites)

Le terme suivant, *hardship payment*, compte 62 occurrences et est une composante de *Universal Credit*. Il s'agit d'un versement réalisé en cas de circonstances exceptionnelles.

— *hardship payment* (62 occurrences, 8 occurrences sur `gov.uk`, 8 occurrences sur `bristol.gov.uk`, 38 occurrence sur `citizensadvice.org.uk`, 4 occurrences sur `falkirk.gov.uk` et 4 occurrences sur `understandinguniversalcredit.gov.uk`)

Le terme suivant est *Cold Weather payment* et compte 65 occurrences. Il s'agit d'une aide ponctuelle versée pour contribuer aux factures d'énergie.

— *Cold Weather payment* (38 occurrences sur `gov.uk`, 4 occurrences sur `citizensadvice.org.uk`, 22 occurrences sur `falkirk.gov.uk` et 1 occurrence sur `understandinguniversalcredit.gov.uk`)

Le terme suivant, *housing payment*, peut faire référence à deux types de dispositifs. Tout d'abord il existe le *Discretionary Housing Payment* (50 occurrences) qui est une allocation logement versée par la municipalité. Il existe un sigle pour ce dispositif : *DHP* (34 occurrences). Le deuxième dispositif est une composante de l'*Universal Credit* qui compte 16 occurrences en corpus. Il s'agit du versement dédié au logement. Dans ces deux cas, il s'agit de versements réguliers, mensuels ou hebdomadaires.

— *Discretionary Housing Payment* (50 occurrences, 7 occurrences sur `gov.uk`,

20 occurrences sur `bristol.gov.uk`, 7 occurrences sur `citizensadvice.org.uk`, 10 occurrences sur `falkirk.gov.uk` et 6 occurrences sur `understanduniversalcredit.gov.uk`)

— *DHP* (34 occurrences)

— *housing payment* (20 occurrences, 12 occurrences sur `gov.uk`, 8 occurrences sur `understanduniversalcredit.gov.uk`)

Le terme suivant, *emergency payment*, compte 32 occurrences. Il s'agit d'une aide financière versée par la municipalité en cas d'urgence en une fois.

— *emergency payment* (32 occurrences sur `bristol.gov.uk`)

Le dernier groupe nominal relevé (*Alternative Payment*) est en réalité *Alternative Payment Arrangement*, il ne s'agit donc pas d'un dispositif d'aide financière.

L'analyse des dispositifs formés à partir du nom *payment* montre que ce terme peut faire référence à une aide versée en une fois (*harship payment*), mais aussi à des versements réguliers (*Discretionary Housing Payment*). Ce terme est aussi polysémique, puisqu'il fait également référence au versement d'une aide, comme dans le groupe nominal *Universal Credit payment*.

5.1.1.2.5 Synthèse des résultats de l'exploration du corpus BEN-AN

L'étude de ces différents noms noyaux nous a montré que certains ne sont utilisés que dans les dispositifs d'aide financière comme *allowance*. D'autres, comme *benefit*, sont en même temps l'hyperonyme et l'hyponyme. D'autres encore, comme *credit* et *payment*, peuvent avoir plusieurs significations. Toutefois, nous avons noté l'introduction d'une aide qui centralise plusieurs dispositifs : *Universal Credit*. Ce dispositif, qui avait pour but de simplifier le paysage des aides sociales, n'implique pas un changement complet de stratégie terminologique. En effet, on retrouve les termes suivants dans nos relevés :

— Universal Credit,

— standard allowance,

— work allowance,

— hardship payment,

— housing payment.

Ces termes font référence à chaque composante du dispositif *Universal Credit*. La terminologie de ce dispositif est donc tout aussi complexe et présente les mêmes noms noyaux. Dans nos relevés, nous en retrouvons donc trois pour un seul dispositif : *credit*, *allowance*, *payment*. Nous pouvons cependant noter que chacune des composantes se limite à un groupe nominal adjectif + nom ou nom + nom, contrairement à d'autres dis-

positifs plus longs. De plus, seul le nom du dispositif principal *Universal Credit* dispose d'un sigle *UC*, tandis que les autres restent en minuscule et ne sont pas raccourcis.

En conclusion, malgré une apparente simplification, les dispositifs d'aide financière présentent une terminologie complexe dans le corpus britannique, avec une variation au niveau du noyau et un manque de stabilité dans le choix des termes utilisés.

5.1.1.3 Analyse contrastive des résultats des corpus AIDESOC-FR et BEN-AN

L'analyse des principaux termes décrivant les dispositifs d'aide sociale au Royaume-Uni et en France révèle de nombreuses similitudes. En effet, les analyses ont permis d'identifier des principes de formations similaires.

Les termes sont formés à partir d'un certain nombre de noms noyaux synonymes. En français, les termes noyaux sont les suivants :

— allocation,
— aide,
— revenu,
— complément,
— prime,
— prestation.

En anglais britannique, les termes noyaux sont les suivants :

— benefit,
— credit,
— allowance,
— payment.

Nous avons relevé six noyaux en français et quatre noyaux en anglais britannique. Nous avons choisi d'écarter le nom noyau *bourse* qui ne correspondait plus aux critères après analyse. Le système français comporte un nombre plus important de dispositifs d'aide, ainsi qu'un plus grand nombre d'organismes émetteurs de ces aides, ce qui pourrait expliquer le plus grand nombre de noms noyaux différents.

Dans les deux langues, les dispositifs sont formés par des termes nominaux, des termes composés et des groupes nominaux complexes construits à partir des noyaux identifiés. Dans les deux cas, nous avons pu formaliser les patrons lexico-syntaxiques par une expression CQL qui nous a permis d'identifier la plupart des dispositifs. Toutefois, les mécanismes de formations des termes ne semblent pas répondre aux mêmes logiques. Les termes français sont davantage des termes complexes (**Prestation d'accueil du jeune enfant**, **Allocation personnalisée au logement** ou encore **Allocation de logement à caractère sociale**) avec des sigles et acronymes associés (respectivement **Paje**, **APL** et **ALS**) tentant de rendre la nature nominale de ces termes plus évidente. De

plus, dans le corpus français, les termes sont formés afin de décrire une aide, mais aussi selon une idéologie. Concernant les allocations pour les personnes sans emploi, les deux principales aides disponibles illustrent ce principe :

— Allocation d'aide au retour à l'emploi (ARE)
— Revenu de solidarité active (RSA)

Dans le premier exemple, le but de l'aide est indiqué : soutenir le chômeur en attendant son retour à l'emploi. Le second exemple illustre lui le type d'aide qu'il s'agit : il s'agit du remplacement d'un revenu, pour une personne qui n'en a pas et ce revenu étant solidaire, on indique qu'il s'agit d'une aide sociale, pour laquelle la personne n'a pas cotisé et enfin que le bénéficiaire de l'aide doit être actif, ou doit vouloir redevenir un actif. Dans les deux cas, le nom des dispositifs décrit la situation dans laquelle se trouve l'usager.

Dans le corpus britannique, la formation des termes suit un processus similaire. Toutefois, nous avons noté des différences, car certains termes semblent se complexifier avec le temps (de termes composés ils deviennent des termes complexes) tandis que la terminologie française cherche à raccourcir ses termes (en passant d'un terme complexe à un sigle). Les allocations pour les personnes sans emploi illustrent cette différence (voir section 5.1.1.2). En français, on trouve les termes **Allocation d'aide au retour à l'emploi** ou **Revenu de solidarité active** qui sont réduit pas leur sigle ou acronyme **RSA** ou **ARE**. En anglais, le terme choisi initialement est similaire au terme français (*Jobseeker Allowance*) et il présente également un sigle (*JSA*). Ce terme est en fait un hyperonyme, qui a plusieurs hyponymes, notamment : *income-based Jobseeker's Allowance* et *contribution-based Jobseeker's Allowance*. Le terme initial devient alors complexe. En anglais, le terme fait aussi référence au statut de l'usager, qui est ici en recherche d'emploi. Ces deux hyponymes décrivent le mode de financement des deux dispositifs d'aide : le premier est accordé selon les revenus de l'usager, tandis que le second est accordé selon les contributions cotisées à l'organisme de sécurité sociale britannique. Toutefois, contrairement à la terminologie française, la base du terme est réutilisée au fil du temps afin de former de nouveaux termes, au fil de l'évolution de la législation. C'est ainsi que le terme évolue vers *New style Jobseeker's Allowance*. Dans cette dernière version du terme, la description n'est pas liée à l'usager, mais au dispositif directement, pour indiquer qu'il s'agit d'un nouveau type de *Jobseeker's Allowance*. La méthode de formation que l'on retrouve dans les exemples *income-based Jobseeker's Allowance* et *contribution-based Jobseeker's Allowance* repose sur la prémodification nominale qui est caractéristique de l'anglais de spécialité et notamment du discours scientifique (Mestivier, 2015). Nous allons revenir sur ce phénomène dans la section 5.1.3 portant sur l'analyse des types de variation observés en corpus.

Cette évolution est intéressante car elle montre une conscience linguistique dans la dénomination des dispositifs, ainsi qu'une diversité dans les processus de dénomination des dispositifs. En effet, le terme juxtaposé *New Style* ne nous dit rien des usagers à qui est destinée cette aide. Il nous donne des informations sur la diachronie des dispositifs.

La terminologie britannique fait également l'objet d'une transformation qui est visible dans le corpus étudié. En effet, l'administration britannique est en train de modifier sa législation afin de recentrer ses dispositifs d'aide sur un dispositif principal : *Universal Credit*, qui remplace cinq dispositifs. Le terme *Universal Credit* indique une volonté de simplification de l'administration britannique. En effet, il fait référence au *Universal Basic Income* ou revenu universel en français, qui demande la suppression des différentes aides sociales pour donner à tous un revenu de base, sans condition. La dénomination *Universal Credit* est considérée comme idéologique, parfois même mensongère, car le dispositif n'est pas universel et dispose d'un certain nombre de conditions (Parker, 2023)

Ce dispositif se découpe en plusieurs allocations suivant les besoins des usagers. La *standard allowance* désigne le versement de base, tandis que le *work allowance* remplace les allocations liées à l'emploi. Ce nouveau dispositif et sa terminologie montrent une volonté de simplification, en proposant des groupes nominaux simples nom + adjectif. Toutefois, il est utile de noter que les dénominations sont moins descriptives[4] et ne renseignent pas sur la nature des aides ni sur le profil des usagers.

Enfin, dans les deux langues, nous avons remarqué une volonté de rendre la communication plus efficace en utilisant les sigles et acronymes des dispositifs (*JSA, ESA, UC* en anglais et **ARE, PreParE, APL** en français). L'utilisation de ces sigles et acronymes n'apparaît pas dans les textes juridiques et vient directement d'une pratique discursive du domaine. Elle opacifie le terme et le rend plus spécialisé car il est nécessaire de connaître sa définition. Toutefois, un grand nombre des dénominations choisies pour les dispositifs ne sont plus transparentes et doivent donc faire l'objet d'un apprentissage. Dans la majorité de ces dispositifs, les mots prennent un sens spécifique à notre domaine, ce qui implique une spécialisation du discours et une terminologie qui nécessite un apprentissage. De plus, l'utilisation des sigles et acronymes fluidifie la communication.

En conclusion, les terminologies dans les deux langues sont complexes et font l'objet de choix linguistiques dictés par des enjeux politiques et juridiques. Toutefois, cette terminologie a une place et une évolution spécifique au sein du domaine de l'accès aux droits et c'est sur elle que repose en grande partie la compréhension des dispositifs d'aide et du discours administratifs dans l'ensemble, et finalement l'accès aux droits.

4. à l'exception de *work allowance*, mais qui peut porter à confusion, puisque les usagers percevant cette aide ne travaillent pas forcément.

5.1.1.4 Terminologie dans le corpus français sur le thème du droit au séjour

Dans cette section, nous présentons les explorations faites dans le corpus ETR-FR qui est composé de l'ensemble du contenu dédié au droit au séjour.

Pour explorer le corpus ETR-FR, nous avons repris le schéma formalisé à partir du corpus d'aide sociale pour savoir s'il s'appliquait à notre corpus. Le patron utilisé est le suivant :

[noyau] [préposition] [complément1] [complément2] [complément3] [...]

Nous l'avons traduit par le patron syntaxique et l'expression CQL suivants :

— Patron syntaxique : [Nom 1] [Preposition] [Nom 2]

— CQL : [tag="NC.*"][tag="SP.*"][tag="NC.*"]

Étant donné que nous n'avons pas réalisé d'étude préalable sur ce thème, nous ne disposons pas de noms noyaux qui pourrait nous aider à identifier les termes principaux. Nous allons donc explorer le corpus grâce au patron afin d'identifier les différents dispositifs. Les résultats de l'exploration automatique grâce au patron syntaxique sont présentés dans la Table 5.18 ci-dessous.

word	Frequency	Freq. per million	% of concordance
titre de séjour	1459	2010,97144	4,14996
carte de séjour	1262	1739,44206	3,58961
acte de naissance	478	658,8378	1,35962
travail des métaux	452	623,00143	1,28566
conducteurs de travaux	440	606,46157	1,25153
carte de résident	392	540,30213	1,115
cadres d'étude	380	523,76227	1,08087
développement en informatique	360	496,19583	1,02398
cours de validité	350	482,41261	0,99553
droit de timbre	338	465,87275	0,9614
carte d'identité	326	449,33289	0,92727
autorisation de travail	318	438,30632	0,90451
certificat de résidence	286	394,20002	0,81349
agents de maîtrise	260	358,36366	0,73954
cachets d'entrée	260	358,36366	0,73954
dates de validité	252	347,33708	0,71678
titres de séjour	229	315,63568	0,65136
chefs de projets	220	303,23079	0,62576
demande de carte	216	297,7175	0,61439
industries de process	210	289,44757	0,59732

TABLE 5.18 – Résultats les plus fréquents pour le patron [nom] [préposition] [nom]

Pour plus de flexibilité, nous avons également exploré un patron simplifié :

[nom] [préposition]

Ce patron se traduit par l'expression CQL suivante :

[tag="NC.*"][tag="SP.*"]

Les résultats de l'exploration automatique réalisée avec ce patron sont présentés dans la Table 5.19 ci-dessous.

word	Frequency	Freq. per million	% of concordance
carte de	1683	2319,71551	3,06641
titre de	1659	2286,63579	3,02268
demande de	1074	1480,31757	1,95682
acte de	756	1042,01125	1,37743
moins de	680	937,25879	1,23895
cas de	649	894,53082	1,18247
délai de	577	795,29165	1,05129
droit de	569	784,26508	1,03671
certificat de	554	763,59025	1,00938
visa de	512	705,70074	0,93286
tension pour	481	662,97276	0,87638
conducteurs de	440	606,46157	0,80168
chefs de	424	584,40842	0,77252
pays de	409	563,7336	0,74519
cadres d'	400	551,3287	0,7288
cours de	393	541,68045	0,71604
pays d'	390	537,54548	0,71058
internet de	379	522,38395	0,69053
autorisation de	372	512,73569	0,67778
durée de	364	501,70912	0,6632

TABLE 5.19 – Résultats les plus fréquents pour le patron [nom] [préposition]

Les Tables 5.18 et 5.19 nous ont permis d'identifier les principaux noms noyaux de ce thème, ainsi que les principaux dispositifs mis en place.

Rappelons que les explorations sont initialement réalisées à partir d'étiquettes syntaxiques attribuées automatiquement par le logiciel Sketch Engine. Une fois les exemples relevés, nous les explorons manuellement en corpus. C'est pourquoi le nombre d'occurrences relevé dans les tableaux initiaux n'est pas toujours le même que le nombre d'occurrences relevé ensuite, car l'étiquetage automatique n'est pas toujours précis et l'utilisation des majuscules peut fausser le nombre d'occurrences ressorti par le logiciel.

Nous allons tout d'abord réaliser l'exploration avec le premier terme apparaissant dans la Table 5.18 qui est **titre de séjour**. Pour cela, nous allons étudier les termes formés à partir de **titre de séjour** grâce à un nouveau patron syntaxique. La Table ci-dessous 5.20 représente les résultats pour le patron syntaxique suivant :

[titre de séjour][nom]

Ce patron se traduit par l'expression CQL suivante :

[word="titre"][word="de"][word="séjour"][tag="NC.*"]

word	Frequency	Freq. per million	% of concordance
titre de séjour requis	6	8,27137	30
titre de séjour passeport	6	8,27137	30
titre de séjour étudiant	4	5,51424	20
titre de séjour visiteur	1	2,75712	10
titre de séjour titre	1	1,37856	5
titre de séjour mention	1	1,37856	5

TABLE 5.20 – Résultats les plus fréquents pour le patron [**titre de séjour**][nom]

Pour compléter cette exploration, nous avons ensuite fait la même exploration en changeant légèrement le patron syntaxique en replaçant le nom par un adjectif :

[titre de séjour][adjectif]

Ce patron se traduit par l'expression CQL suivante :

[word="titre"][word="de"][word="séjour"][tag="A.*"]

La Table 5.21 ci-dessous présente ces résultats.

word	Frequency	Freq. per million	% of concordance
titre de séjour précédent	18	24,8141	24
titre de séjour pluriannuel	12	20,67842	20
titre de séjour permanent	9	12,40705	12
titre de séjour autonome	7	9,64993	9,33333
titre de séjour français	4	5,51424	5,33333
titre de séjour valide	4	5,51424	5,33333
titre de séjour Circulaire	2	2,75712	2,66667
titre de séjour valable	2	2,75712	2,66667
titre de séjour actuel	2	2,75712	2,66667
titre de séjour européen	1	1,37856	1,33333

TABLE 5.21 – Résultats les plus fréquents pour le patron [**titre de séjour**][adjectif]

Les Tables 5.20 et 5.21 présentent donc les résultats de ces différents patrons syntaxiques. Elles montrent que le terme **titre de séjour** est bien présent en corpus et qu'il apparaît avec plusieurs compléments.

Le premier groupe nominal de la Table 5.20 ne correspond pas à un dispositif et ne fait donc pas l'objet d'un commentaire.

Le deuxième groupe nominal que nous avons exploré est **titre de séjour Passeport talent**, avec 14 occurrences. Il s'agit d'un type de titre de séjour dont nous aborderons les spécificités dans la section 5.1.5.

— **titre de séjour Passeport talent** (14 occurrences sur `welcometofrance.fr`)

Le troisième terme est le suivant : **titre de séjour étudiant** avec 12 occurrences. Il s'agit encore une fois d'un type de titre de séjour réservé aux étudiants.

— **titre de séjour étudiant** (12 occurrences dont 11 occurrences sur `service-public.fr` et 1 occurrence sur `prefecturedepolice.interieur.gouv.fr`)

Le quatrième terme est **titre de séjour visiteur**, avec 1 occurrence. Il s'agit encore ici d'un type de titre de séjour.

— **titre de séjour visiteur** (1 occurrence sur `service-public.fr`)

Le cinquième groupe nominal, **titre de séjour mention**, compte 2 occurrences. Le complément **mention** est utilisé pour introduire les différents types de titres de séjour. Ce groupe nominal est suivi d'un complément indiquant la raison du séjour comme **titre de séjour mention « Etudiant »** (1 occurrence) ou **titre de séjour mention « Séjour permanent »** (1 occurrence).

— **titre de séjour mention** (2 occurrences sur `welcometofrance.fr`)

Dans la Table 5.21, le deuxième terme **titre de séjour pluriannuel** présente 12 occurrences. Il s'agit d'une catégorie de titre de séjour et plus précisément d'un titre de séjour longue durée. Le troisième terme est **titre de séjour permanent**, avec 9 occurrences. Il s'agit d'une autre catégorie de titre de séjour qui ne nécessite pas de renouvellement et qui est également un titre de séjour longue durée.

— **titre de séjour pluriannuel** (12 occurrences, 5 sur `service-public.fr`, 6 occurrences sur `welcometofrance.fr` et 1 occurrence sur `prefecturedepolice.interieur.gouv.fr`)
— **titre de séjour permanent** (8 occurrences, 7 sur `service-public.fr`, 1 occurrence sur `welcometofrance.fr`)

Les autres termes relevés ne sont pas des dispositifs de droit au séjour, mais d'autres types d'occurrences en corpus.

Le deuxième terme composé de notre corpus est **carte de séjour**. Il s'agit *a priori* d'un équivalent ou d'un terme en lien avec **titre de séjour**. Notre exploration pourra nous aider à déterminer si leur utilisation est similaire ou différente. Nous avons donc récolté dans les Tables 5.22 et 5.23 ci-dessous les occurrences des deux patrons syntaxiques

présentés au début de cette section.

word	Frequency	Freq. per million	% of concordance
carte de séjour vie	45	62,03525	33,08824
carte de séjour passeport	40	55,14245	29,41176
carte de séjour mention	12	16,54273	8,82353
carte de séjour pluriannuellevie	8	11,02849	5,88235
carte de séjour membre	7	9,64993	5,14706
carte de séjour étudiant	6	8,27137	4,41176
carte de séjour étudiant-programme	4	5,51424	2,94118
carte de séjour résident	4	5,51424	2,94118
carte de séjour recherche	4	5,51424	2,94118
carte de séjour carte	1	1,37856	0,73529
carte de séjour pluriannuelleentrepreneur	1	1,37856	0,73529
carte de séjour visiteur	1	1,37856	0,73529
carte de séjour entrepreneur	1	1,37856	0,73529
carte de séjour permanent	1	1,37856	0,73529
carte de séjour temporairevie	1	1,37856	0,73529

TABLE 5.22 – Résultats les plus fréquents pour le patron [**carte de séjour**][nom]

word	Frequency	Freq. per million	% of concordance
carte de séjour temporaire	143	197,13425	45,83333
carte de séjour pluriannuelle	125	172,32015	40,0641
carte de séjour Citoyen	20	39,97827	9,29487
carte de séjour travailleur	3	4,13568	0,96154
carte de séjour jeune	2	2,75712	0,64103
carte de séjour Justificatif	2	2,75712	0,64103
carte de séjour française	2	2,75712	0,64103
carte de séjour chercheur	2	2,75712	0,64103
carte de séjour étudiant	2	2,75712	0,64103
carte de séjour citoyen	1	1,37856	0,32051

TABLE 5.23 – Résultats les plus fréquents pour le patron [**carte de séjour**][adjectif]

Ces deux tableaux présentent plusieurs utilisations communes avec **titre de séjour**. Tout d'abord, l'occurrence de **carte de séjour passeport** est en fait **carte de séjour passeport talent**, avec 42 occurrences. Il semble donc que ce variant soit plus utilisé en corpus que le variant précédent **titre de séjour Passeport Talent** (14 occ.).

— **carte de séjour passeport talent** (42 occurrences sur `service-public.fr`)

Nous avons ensuite identifié la **carte de séjour pluriannuelle**, avec 131 occurrences, qui est également un variant de **titre de séjour pluriannuel** (12 occ.). À nouveau, le variant formé à partir de **carte de séjour** est bien plus présent dans le corpus.

— **carte de séjour pluriannuelle** (132 occurrences, 103 sur `service-public.fr`, 2 sur `prefecturedepolice.interieur.gouv.fr` et 27 sur `welcometofrance.fr`)

Le terme **carte de séjour étudiant** qui compte 26 occurrences est également un variant de **titre de séjour étudiant**. Encore une fois, le nombre d'occurrences du nom **étudiant** modifiant le terme **carte de séjour** est plus important qu'avec le variant **titre de séjour** (on ne trouve que 12 occurrences de **titre de séjour étudiant** dans le corpus).

— **carte de séjour étudiant** (26 occurrences sur `service-public.fr`)

À l'inverse, nous retrouvons une seule occurrence de l'adjectif **permanent** avec le terme **carte de séjour**, qui était plus fréquemment observé en corpus avec le variant **titre de séjour** (on trouve 8 occurrences de **titre de séjour permanent** dans le corpus).

— **carte de séjour permanent** (1 occurrence sur `service-public.fr`)

Le nom **mention**, qui introduit un statut, est également présent avec 8 occurrences en complément de **carte de séjour** et 2 occurrences en complément de **titre de séjour**. Encore une fois, il semble que le variant avec **carte de séjour** est légèrement plus utilisé.

— **carte de séjour mention** (8 occurrences, 7 sur `service-public.fr`, 1 occurrence sur `welcometofrance.fr`)

Certains groupes nominaux relevés avec **carte de séjour** ne sont pas présents avec **titre de séjour**. Le terme de **carte de séjour temporaire** semble effectivement inédit. En explorant le corpus manuellement, nous avons relevé une seule occurrence de **titre de séjour temporaire**, tandis que nos relevés indiquent 125 occurrences de **carte de séjour temporaire**. Le terme est donc formé principalement à partir de **carte de séjour**.

— **titre de séjour temporaire** (1 occurrence sur `welcometofrance.fr`)

— **carte de séjour temporaire** (125 occurrences, 103 sur `service-pubic.fr`, 2 sur `prefecturedepolice.interieur.gouv.fr` et 20 occurrences sur `welcometofrance.fr`)

Nous avons également identifié les relevés suivants comme étant des dispositifs formés uniquement avec **carte de séjour** :

— **carte de séjour jeune au pair** (2 occurrences sur `service-public.fr`)

— **carte de séjour chercheur** (4 occurrences sur `service-public.fr`)

— **carte de séjour Citoyen UE/EEE/Suisse** (22 occurrences sur `service-public.fr`)

— **carte de séjour résident permanent** (3 occurrences sur `service-public.fr`)

Il existe donc certains dispositifs de droit au séjour qui ne sont employés qu'avec **titre de séjour** ou **carte de séjour**. D'autres dispositifs peuvent être formés à partir des deux termes.

Le glossaire de la Direction générale des étrangers en France propose la définition suivante pour **titre de séjour** :

> « Document sécurisé assurant la reconnaissance par l'autorité publique du droit à séjourner sur le territoire national pour un ressortissant étranger majeur. Un titre de séjour se définit par sa nature juridique, son motif d'admission et sa durée de validité. Les cartes de séjour temporaires, les cartes de séjour, les cartes de résident et les certificats de résidence pour Algérien sont des titres de séjour. Voir aussi : carte de séjour temporaire, carte de résident. »

Il n'existe pas de définition pour **carte de séjour** dans ce glossaire. D'après cette définition et les relevés réalisés en corpus, il semble donc que **titre de séjour** soit en même temps un hyperonyme et un équivalent de **carte de séjour**.

Nous proposons ci-dessous une liste récapitulative des types de titres de séjour recensés par notre exploration, qui sera complétée dans la section 5.10 afin de proposer une ontologie du concept de document de séjour :

— résident
— temporaire
— pluriannuel
— permanent

Nous avons ensuite fait des explorations plus poussées sur un certain type de titre de séjour : le **passeport talent**. La liste ci-dessous rassemble l'ensemble des relevés faits à partir du terme **passeport talent**.

Carte de séjour passeport talent possède les compléments suivants :

— **carte de séjour passeport talent (famille)** ou **"famille"** (15)
— **carte de séjour passeport talent-salarié qualifié** (1)
— **carte de séjour passeport talent-carte bleue européenne** (2)
— **carte de séjour passeport talent salarié en mission** (2)
— **carte de séjour passeport talent- chercheur** (2)
— **carte de séjour passeport talent-création d'entreprise** (2)
— **carte de séjour passeport talent-projet économique innovant** (2)
— **carte de séjour passeport talent-investissement économique** (2)

— **carte de séjour passeport talent-mandataire social** (2)

— **carte de séjour passeport talent-profession artistique et culturelle** (2)

Les **passeports talents** présentent des mentions, à la manière des titres de séjour plus classiques. En étudiant ce terme en corpus, nous avons observé qu'il présentait plusieurs stades de néologie dus au transfert du terme de son statut juridique. Tout d'abord, on trouve des occurrences de :

(6) (…) visa de long séjour portant la mention « Passeport Talent ».

Dans ce cas, le passeport talent n'est qu'une mention de carte de séjour. Nous avons également trouvé une version raccourcie de ces occurrences :

(7) (…) un titre de séjour « Passeport Talent ».

Dans cet exemple, le terme est simplifié et on ne retrouve pas le complément « mention » qui a été supprimé probablement pour simplifier et rendre plus efficace la communication. Nous trouvons ensuite des occurrences de type :

(8) (…) carte de séjour pluriannuelle passeport talent.

Dans cet exemple, il n'y a plus de majuscules ni de guillemets, montrant que le terme est utilisé couramment et adopté par les experts du domaine. Enfin, on retrouve des occurrences du terme sans hyperonyme ou catégorie, soit des formes réduites, comme les occurrences suivantes :

(9) Je souhaite renouveler mon passeport talent.

(10) Vous souhaitez solliciter un passeport talent.

Dans ces deux derniers exemples, le terme est totalement intégré dans le discours, et on ne retrouve plus de trace d'explicitation du terme. Le terme n'est plus considéré comme un transfert du discours juridique, mais bien comme un terme à part entière du discours de l'accès aux droits ne nécessitant pas d'explicitation. Il est important de noter que l'ensemble de ces occurrences se côtoient dans un corpus en synchronie, et que nous ne savons pas s'il s'agit réellement d'une évolution diachronique ou simplement d'une manifestation des différents statuts et degrés de spécialisation des rédacteurs. Pour notre recherche, l'ensemble de ces utilisations nous permet de mieux comprendre le terme, et de le situer dans le discours. Toutefois, cela est possible grâce à notre approche par corpus et impossible pour l'usager qui ne consulte qu'une ou deux pages à la fois. Il semble donc que pour lui, une utilisation cohérente soit plus utile.

Enfin, nous souhaitons analyser l'utilisation du terme **certificat de résidence** qui est un type de titre de séjour réservé aux personnes de nationalité algérienne. Nous avons relevé deux types de **certificat de résidence**, un an et dix ans :

— **certificat de résidence** (268 occurrences, tous sites)
— **certificat de résidence d'1 an** (11 occurrences sur `service-public.fr`)
— **certificat de résidence d'un an** (5 occurrences sur `service-public.fr`)
— **certificat de résidence de 10 ans** (50 occurrences, 40 occurrences sur `service-public.fr` et 10 occurrences sur `welcometofrance.fr`)

Ici, nous souhaitons mettre en avant un cas où des facteurs juridiques dictés par la politique influencent le discours de l'accès aux droits. Un certificat de résidence est un titre de séjour, les personnes détentrices de ces titres ont des cartes de séjour et des récépissés identiques aux autres ressortissants étrangers hors UE. Ici, le terme décrit la situation spécifique des personnes de nationalité algérienne, qui ne répondent pas aux mêmes critères d'immigration pour des raisons politiques, comme le décrit le GISTI dans son cahier juridique Les droits des Algériennes et des Algériens en France :

> « L'accord franco-algérien régit les conditions dans lesquelles les ressortissants algériens peuvent être admis à séjourner en France et y exercer une activité professionnelle, ainsi que les règles concernant la nature et la durée de validité des titres de séjour qui peuvent leur être délivrés. Toutefois, dès lors qu'elles ne sont pas incompatibles avec l'accord franco-algérien, les dispositions de droit commun applicables aux autres personnes étrangères concernent également les Algérien·ne·s. »

Ce terme, issu d'un accord, traduit donc une différence en matière de politique de l'immigration, mais peut porter des mentions identiques à une carte de séjour comme « vie privée et familiale » ou « étudiant ». Nous observons donc ici une terminologie dictée et complexifiée par les accords politiques.

Pour aller plus loin, nous avons récolté l'hyperonyme de l'ensemble des dispositifs permettant l'accès au séjour en France, grâce au contexte suivant :

(11) document de séjour (visa, VLS-TS ou titre). (service-public.fr)

À partir de ce relevé, nous avons décidé de faire des explorations supplémentaires autour du concept de **visa**. Pour explorer la formation des termes par rapport au concept de **visa**, nous avons formalisé l'expression CQL suivante :

```
[lemma="visa"][]{0,2}[tag="N.*"]
```

Cette expression correspond au patron syntaxique suivant :

[**visa**] ([mot]) ([mot]) [nom]

Les deux mots qui suivent **visa** sont facultatifs.

word	Frequency	Freq. per million	% of concordance
visa de long séjour	241	469,29158	25,15658
visas) ou attestation	107	208,35767	11,1691
visa de régularisation	73	142,15056	7,62004
visa de court séjour	63	122,67788	6,5762
visa, VLS-TS	52	101,25794	5,42797
visas (ambassade	31	60,36531	3,23591
visa long séjour	28	54,5235	2,92276
visa Schengen	23	44,78716	2,40084
visa par timbres	16	31,15629	1,67015

TABLE 5.24 – Résultats les plus fréquents pour le patron [**visa**] ([mot]) ([mot]) [nom]

La Table 5.24 permet d'identifier deux types de visas qui semblent être en lien : les visas **court séjour** et les visas **long séjour**. On remarque qu'il existe deux variants : **visa de long séjour** et **visa long séjour**. Cette variation existe également pour l'autre l'hyponyme désignant l'autre type de visa (**visa de court séjour** et **visa long séjour**). Le **visa de régularisation** n'est pas un type de visa permettant de rester en France, mais un droit de visa financier. Le **VLS-TS** est un autre type de visa, il s'agit ici du sigle qui correspond au terme **visa long séjour valant titre de séjour**.

- **visa de court séjour** (65 occurrences, 32 sur `service-public.fr` et 33 sur `welcometofrance.fr`)
- **visa court séjour** (6 occurrences, 1 sur `prefecturedepolice.interieur.gouv.fr` et 5 sur `welcometofrance.fr`)
- **visa de long séjour** (321 occurrences, 228 occurrences sur `service-public.fr`, 93 occurrences sur `welcometofrance.fr`)
- **visa long séjour** (32 occurrences, 5 occurrences sur `service-public.fr`, 1 occurrence sur `prefecturedepolice.interieur.gouv.fr` et 26 occurrences sur `welcometofrance.fr`)
- **visa long séjour valant titre de séjour** (21 occurrences, 3 occurrences sur `service-public.fr`, 18 occurrences sur `welcometofrance.fr`)
- **VLS-TS** (189 occurrences, 124 occurrences sur `service-public.fr`, 1 occurrence sur `prefecturedepolice.interieur.gouv.fr`, 64 occurrences sur `welcometofrance.fr`)

Les explorations réalisées avec ces termes montrent que chaque site a une préférence dans l'utilisation d'un variant ou d'un autre. Le site `service-public.fr` semble utiliser **visa de long séjour** et **visa de court séjour**, qui est la version issue du discours juridique. Sur le site `welcometofrance.fr`, on retrouve une utilisation des deux variants, ce qui indique que la terminologie est moins stable, ou plus diverse, sur ce site. On remarque également sur ce même site édité par Business France que **visa** est

parfois utilisé à la place de **titre de séjour** dans l'expression **visa « étudiant »** par exemple. Il s'agit d'une utilisation spécifique à ce site web qui n'est pas rédigé par le même organisme et les mêmes rédacteurs que les autres sites. De plus, il est destiné aux personnes souhaitant s'installer en France qui initialement demandent un **visa** avant de s'installer définitivement et de demander un **titre de séjour**.

La terminologie autour des visas peut être complexe, car elle rassemble deux types de législation. D'abord celle des titres de séjour, car lorsqu'on entre en France, il est nécessaire de demander un **visa** qui permettra de rentrer sur le territoire français, avant de demander la **carte de séjour** une fois arrivé sur le territoire. Ces deux documents demandés (**visa** puis **carte de séjour**) doivent porter la même mention (**étudiant** par exemple). Un **visa étudiant** peut alors devenir une **carte de séjour étudiant**. Les visas portent également d'autres dénominations lorsqu'ils ne sont pas obtenus dans l'objectif de résider sur le territoire de façon permanente (**vacances-travail** ou **visa de long séjour temporaire**). On retrouve alors deux types de terminologie issus de deux domaines juridiques différents.

La terminologie autour des différents types de documents de séjour est complexe et parfois instable. Les différentes façons de désigner le droit au séjour sont combinées avec des **mentions** de droit au séjour qui précisent les raisons pour lesquelles le visa a été accordé. Ces mentions ont un objectif descriptif qui peut faciliter leur compréhension. Toutefois, l'utilisation des différents termes comme **titre de séjour**, **carte de séjour**, **certificat de résidence** et **passeport talent** ne semble pas guidée par une logique descriptive. Il peut donc être difficile de comprendre ou encore d'utiliser l'ensemble de ces termes.

Afin de guider la compréhension des usagers, nous avons choisi de formaliser la terminologie autour du concept de **document de séjour** dans une ontologie dans la section 5.1.5.

5.1.1.5 Terminologie dans le corpus anglais sur le thème du droit au séjour

Pour explorer le corpus IMM-AN, nous avons identifié les relevés suivants à partir d'un patron syntaxique :

[nom] [nom] [visa]

L'expression CQL correspondante est la suivante :

[tag=" N.*"] [tag=" N.*"] [word=" visa"]

La Table ci-dessous présente les résultats de cette exploration automatique.

word	Frequency	Freq. per million	% of concordance
Global Talent visa	50	357,20409	22,52252
Standard Visitor visa	37	264,33103	16,66667
Child Student visa	19	135,73756	8,55856
International Sportsperson visa	18	128,59347	8,10811
Intra-company Transfer visa	9	64,29674	4,05405
Graduate Trainee visa	8	57,15265	3,6036
replacement BRP visa	8	57,15265	3,6036
UK Ancestry visa	7	50,00857	3,15315
International Agreement visa	7	50,00857	3,15315
Creative Worker visa	6	42,86449	2,7027
Seasonal Worker visa	5	35,72041	2,25225
Mobility Scheme visa	5	35,72041	2,25225
Overseas Business visa	5	35,72041	2,25225
Marriage Visitor visa	3	21,43225	1,35135
Charity Worker visa	3	21,43225	1,35135
Standard Visitor visas	3	21,43225	1,35135
Care Worker visa	3	21,43225	1,35135
Domestic Worker visa	3	21,43225	1,35135
Contact UK Visas	2	14,28816	0,9009
family reunion visa	2	14,28816	0,9009
visa Entrepreneur visa	1	7,14408	0,45045
visa Family visa	1	7,14408	0,45045
parent's work visa	1	7,14408	0,45045
General work visa	1	7,14408	0,45045
sponsor UK Visas	1	7,14408	0,45045
guide Eligible visas	1	7,14408	0,45045
Child Student visas	1	7,14408	0,45045
months Student visa	1	7,14408	0,45045
dentist student visa	1	7,14408	0,45045
Agreement Worker visa	1	7,14408	0,45045
partner's work visa	1	7,14408	0,45045
employment work visa	1	7,14408	0,45045
requirements Intra-company visas	1	7,14408	0,45045
Child student visa	1	7,14408	0,45045
visa Sportsperson visa	1	7,14408	0,45045
visitor visa visa	1	7,14408	0,45045
appointment UK Visas	1	7,14408	0,45045
Graduate Entrepreneur visa	1	7,14408	0,45045
Healthcare Visitor visa	1	7,14408	0,45045

TABLE 5.25 – Relevés réalisés à partir du patron [nom] [nom] [visa]

À partir de ces relevés, nous avons réalisé une exploration manuelle. Nous avons ensuite pu lister et classer l'ensemble des types de visas présents en corpus sur les sites d'accès aux droits. Ils sont présentés dans la liste ci-dessous. Ce corpus n'est composé que d'un seul site web : `gov.uk`, et l'ensemble des exemples et termes présentés

viennent donc de ce site.

— *Student visa* (81 occurrences)
— *Global Talent visa* (51 occurrences)
— *Family visa* (48 occurrences)
— *Standard visitor visa* (40 occurrences)
— *Child student visa* (21 occurrences)
— *Seasonal worker visa* (20 occurrences)
— *International sportsperson visa* (18 occurrences)
— *Creative worker visa* (13 occurrences)
— *Intra-company transfer visa* (9 occurrences)
— *Graduate trainee visa* (8 occurrences)
— *UK Ancestry visa* (7 occurrences)
— *International Agreement visa* (7 occurrences)
— *Domestic worker visa* (7 occurrences)
— *Mobility scheme visa* (5 occurrences)
— *Overseas Business visa* (5 occurrences)
— *Marriage visitor visa* (3 occurrences)
— *Charity worker visa* (3 occurrences)
— *Care worker visa* (3 occurrences)
— *Family reunion visa* (2 occurrences)
— *Entrepreneur visa* (1 occurrence)
— *Graduate entrepreneur visa* (1 occurrence)
— *General work visa* (1 occurrence)
— *International Agreement worker visa* (1 occurrence)
— *Healthcare Visitor visa* (1 occurrence)

Ces termes désignent des concepts distincts, car chacun commence par une majuscule, ce qui indique qu'il s'agit de noms propres représentant des catégories de visas. Il semble exister un grand nombre de catégories différentes qui proposent différents degrés de précision. Dans le cas d'un *worker visa*, il existe un grand nombre d'adjectifs antéposés précisant la catégorie professionnelle du ressortissant (par exemple *Domestic worker visa* ou *Charity worker visa*). L'ensemble des catégories professionnelles sont descriptives, mais elles ne sont pas toujours compréhensibles par une personne non initiée, à l'image de la dénomination *International Agreement worker visa*. Néanmoins, ceci reste un cas isolé. Au Royaume-Uni, il n'existe que différents statuts de visa, qui sont délivrés selon certains critères. Pour les résidents permanents, ils bénéficient d'un *indefinite leave to remain*. Nous n'avons pas identifié de sigle en corpus pour ce dispositif.

— *indefinite leave to remain* (91 occurrences)

Toutefois, il est intéressant de noter que d'autres sites qui ne font pas partie de notre corpus utilisent un sigle (*ILR*). Nous avons trouvé des occurrences du sigle dans des guides publiés par des cabinets d'avocats (Davidson Morris et Sable International par exemple)[5]

En conclusion, les dispositifs de droit au séjour ne proposent pas un système terminologique complexe, mais plutôt une description de statut selon le visa obtenu. Les dénominations des visas sont dictées par la législation.

5.1.1.6 Analyse contrastive des résultats des corpus ETR-FR et IMM-AN

Nous pouvons observer que la manière de former les termes liés au droit au séjour, notamment en ce qui concerne les visas et les titres de séjour, diffère entre la France et le Royaume-Uni. Les relevés révèlent un grand nombre de dénominations pour les visas britanniques, qui témoignent d'une pluralité de statuts et de termes. En français, les mentions de titre de séjour sont plus restreintes et font l'objet d'un système conceptuel plus compliqué avec différents types de visas : **titre de séjour**, **VLS-TS**, **visa long séjour**, etc. Dans les deux pays, les dénominations choisies décrivent le profil des ressortissants, et plus précisément la raison qui leur permet de rester sur le territoire. En France, le **titre de séjour « vie privée et familiale »** indique clairement que la raison de présence d'un individu n'est pas liée à son activité professionnelle. Au Royaume-Uni, le *employment work visa* indique clairement qu'à l'inverse, le ressortissant est présent pour des raisons professionnelles, et plus précisément qu'il est salarié.

Dans le domaine de l'accès aux droits au séjour, il est intéressant de noter que le processus de spécialisation semble moins marqué en anglais qu'en français. En effet, si la terminologie française semble avoir instauré un sens spécifique aux différents types de visas et de titres de séjour afin de communiquer de façon claire et sans ambiguïté pour une communauté de discours experte, la terminologie britannique ne répond pas toujours à ces critères. Les dénominations britanniques concernant les visas sont plutôt descriptives, ne donnant que rarement un sens spécifique aux termes utilisés. De plus, elles se retrouvent peu utilisées en corpus, c'est-à-dire par les locuteurs de la communauté de discours. Il semble que ces dénominations constituent davantage des catégories juridiques que des termes à part entière. À l'inverse, en français, on remarque clairement l'évolution et l'utilisation des différents types de titres de séjour. Le terme **passeport talent** par exemple est utilisé pour remplacer son hyperonyme **titre de séjour** :

(12) Les conditions pour obtenir un passeport talent « carte bleue européenne »

5. `sableinternational.com/indefinite-leave-to-remain` et `davidsonmorris.com/indefinite-leave-to-remain/`, consultés le 17 juillet 2024.

Dans l'exemple 12 ci-dessus, le rédacteur omet de préciser qu'il s'agit d'un « titre de séjour mention passeport talent », comme cela est indiqué dans le discours juridique et parle d'un « passeport talent ». Cela montre une évolution de la terminologie au sein du discours de l'accès aux droits, qui devient un discours spécialisé à part entière avec une utilisation des termes propres à sa communauté de discours. Par ailleurs, l'extension « carte bleue européenne » illustre le cas d'une dénomination qui repose sur la précision du profil du ressortissant éligible pour ce type de titre, c'est-à-dire être hautement qualifié (trois années d'études supérieures ou cinq années d'expérience professionnelle) ou détenteur d'une **carte bleue européenne (CBE)** accordée par un autre État membre de l'UE[6].

Le terme **VLS-TS** semble faire l'objet d'un processus similaire. Il s'agit initialement du sigle du terme **visa long séjour valant titre de séjour** qui semble avoir fait l'objet d'un processus de transformation selon le principe de l'économie du langage. Afin de simplifier la communication et de la rendre plus efficace, les rédacteurs réduisent les termes à leurs acronymes ou sigles qui sont une forme plus facile à utiliser en texte. Toutefois, sur les 189 occurrences du terme **VLS-TS** en corpus, seules 55 sont accompagnées du terme initial complet (**visa long séjour valant titre de séjour**). De plus, certains contextes montrent que le terme a maintenant une place à part entière :

(13) Vous devez déposer votre demande de carte en ligne, au plus tôt 4 mois et au plus tard 2 mois avant la fin de validité de votre document de séjour (visa, VLS-TS ou titre). (service-public.fr)

(14) Vous devez déposer votre demande de carte à la préfecture (ou sous-préfecture) de votre domicile, 2 mois avant la fin de validité de votre VLS-TS ou de votre titre de séjour. (service-public.fr)

Dans ces deux exemples similaires, le sigle est placé au même niveau que *titre (de séjour)* et *visa*. De plus, dans le premier contexte (13), il s'agit d'une phrase répétée qui sert à l'usager à se repérer sur le site web, car elle précède un lien (voir section 5.4.1).

Nous avons également relevé que les systèmes conceptuels étaient différents dans les deux pays. Au Royaume-Uni, on retrouve une différence entre le détenteur de visa et le résident permanent. Il n'existe donc que deux catégories de ressortissants étrangers. En France, le système conceptuel est plus complexe puisqu'il est composé de plusieurs catégories et sous-catégories, les deux grandes catégories étant les **visas** et les **titres de séjour**, mais chacune de ces catégories ayant plusieurs sous-catégories. L'ensemble des systèmes conceptuels sont représentés à la section 5.1.5.

En conclusion, nos analyses terminologiques montrent que dans le corpus anglais

6. voir sur `https://www.welcometofrance.com/fiche/passeport-talent-cbe`

IMM-AN, le discours est moins spécialisé que dans le corpus comparable français (ETR-FR). La spécialisation retrouvée dans le corpus britannique émane de son origine juridique et administrative et ne constitue pas un type de spécialisation spécifique au domaine de l'accès aux droits au séjour. Ces conclusions peuvent être liées aux différents types de législation au Royaume-Uni et en France.

5.1.1.7 Terminologie dans le corpus français sur le thème de la citoyenneté

Pour explorer le corpus CIT-FR, nous avons utilisé des schémas similaires à ceux utilisés pour les thèmes précédents. Le premier patron utilisé est le suivant :

[nom][préposition][nom]

Il est traduit par l'expression CQL suivante :

[tag="NC.*"][tag="SP.*"][tag="NC.*"]

Les résultats relevés par le logiciel avec l'expression CQL ci-dessus sont présentés dans la Table 5.26.

word	Frequency	Freq. per million	% of concordance
acte de naissance	312	510,45534	1,13781
renouvellement de prestations	292	477,73384	1,06488
lieu de résidence	252	412,29085	0,919
demande en ligne	218	356,66431	0,79501
Guide d'utilisation	193	315,76244	0,70384
utilisation du certificat	192	314,12636	0,70019
chef d'établissement	180	294,49346	0,65643
parents d'élèves	174	284,67701	0,63455
lundi au vendredi	169	276,49664	0,61632
juge aux affaires	153	250,31944	0,55797
pension de réversion	146	238,86692	0,53244
livret de famille	132	215,96187	0,48138
pièce d'identité	127	207,7815	0,46315
service en ligne	114	186,51253	0,41574
droit de visite	113	184,87645	0,41209

TABLE 5.26 – Résultats les plus fréquents pour le patron [nom][préposition][nom]

Les groupes nominaux relevés dans ce tableau sont variés et désignent différents types de dispositifs, objets, etc. Certains termes sont relatifs à l'identité comme : **acte de naissance**, **livret de famille** et **pièce d'identité**.

— **acte de naissance** (383 occurrences, 42 occurrences sur `paris.fr`, 272 occurrences sur `service-public.fr`, 30 occurrences sur `marseille.fr`, 39 occurrences sur `rennes.fr`)

— **livret de famille** (172 occurrences, 52 occurrences sur `paris.fr`, 60 occurrences sur `service-public.fr`, 10 occurrences sur `marseille.fr`, 50 occurrences sur `rennes.fr`)
— **pièce d'identité** (162 occurrences, 38 occurrences sur `paris.fr`, 75 occurrences sur `service-public.fr`, 8 occurrences sur `marseille.fr`, 41 occurrences sur `rennes.fr`)

D'autres font référence au caractère numérique du domaine comme **demande en ligne** ou encore **service en ligne**.

— **demande en ligne** (224 occurrences, 3 occurrences sur `paris.fr`, 216 occurrences sur `service-public.fr` et 5 occurrences sur `rennes.fr`)
— **service en ligne** (115 occurrences, 7 occurrences sur `paris.fr`, 106 occurrences sur `service-public.fr`, 2 occurrences sur `marseille.fr`)

Les groupes nominaux **guide d'utilisation** et **renouvellement de prestations** sont issues de phrases répétées liées au menu et à la navigation et ne sont donc pas à prendre en compte.

— **guide d'utilisation** (199 occurrences, 7 occurrences sur `paris.fr`, 192 occurrences sur `service-public.fr`)
— **renouvellement de prestations** (292 occurrences, sur `service-public.fr`)

Le terme **pension de réversion** appartient plutôt à la thématique aide sociale, traité dans la section 5.1.1.1.

À partir de nos connaissances et de ces relevés, nous pouvons orienter nos recherches vers les termes qui relèvent de la terminologie **acte de** en utilisant le patron suivant :

[acte de][préposition][mot]

[lemma="acte"][tag="SP.*"][]

La Table 5.27 ci-dessous recense les résultats obtenus.

word	Frequency	Freq. per million	% of concordance
acte de naissance	312	510,45534	37,7724
acte d'état	53	86,71196	6,41646
acte de mariage	40	65,44299	4,84262
actes de naissance	40	65,44299	4,84262
acte de notoriété	38	62,17084	4,60048
acte de décès	32	52,35439	3,87409
acte de procédure	28	45,81009	3,38983
actes de la	23	37,62972	2,7845
acte sous signature	18	29,44935	2,17918
acte d'huissier	18	29,44935	2,17918
actes d'état	16	26,1772	1,93705
actes de disposition	16	26,1772	1,93705
acte de donation	11	17,99682	1,33172
actes d'administration	10	16,36075	1,21065

TABLE 5.27 – Résultats les plus fréquents pour le patron [**acte de**][préposition][mot]

Le second résultat, **acte d'état**, correspond à **acte d'état civil**, l'hyperonyme des termes **acte de naissance**, **acte de mariage** et **acte de décès**.

— **acte d'état civil** (73 occurrences, 2 occurrences sur `paris.fr`, 32 occurrences sur `service-public.fr`, 28 occurrences sur `marseille.fr` et 10 occurrences sur `rennes.fr`)
— **acte de naissance** (383 occurrences, 42 occurrences sur `paris.fr`, 272 occurrences sur `service-public.fr`, 30 occurrences sur `marseille.fr` et 39 occurrences sur `rennes.fr`)
— **acte de mariage** (44 occurrences, 7 occurrences sur `paris.fr`, 24 occurrences sur `service-public.fr`, 4 occurrences sur `marseille.fr` et 9 occurrences sur `rennes.fr`)
— **acte de décès** (40 occurrences, 8 occurrences sur `paris.fr`, 21 occurrences sur `service-public.fr`, 6 occurrences sur `marseille.fr` et 5 occurrences sur `rennes.fr`)

L'étude de l'ensemble de ces termes permet de mettre en avant une utilisation similaire et proportionnelle d'**acte de mariage** et **acte de décès** qui ont généralement un nombre d'occurrences similaire sur les sites (par exemple une vingtaine d'occurrences sur `service-public.fr`). De plus, la grande proportion du terme **acte d'état civil** sur `service-public.fr` tend à montrer qu'il s'agit d'un site plus généraliste.

Cette terminologie administrative est déjà bien documentée puisqu'on la retrouve dans le lexique administratif publié par le ministère de la Fonction publique et de la réforme de l'État en 2004. Une partie de ces termes sont également des termes juridiques, dont on trouve la définition dans des glossaires juridiques ou encore des dictionnaires de

langue générale, comme **acte de notoriété**.

— **acte de notoriété** (52 occurrences, 51 occurrences sur `service-public.fr` et 1 occurrence sur `rennes.fr`)

Certains de ces termes peuvent être considérés comme purement administratifs comme **lieu de résidence**, qui a une signification particulière dans le discours de l'accès aux droits, car il conditionne certains droits et certaines démarches.

— **lieu de résidence** (257 occurrences, 3 occurrences `paris.fr`, 252 occurrences sur `service-public.fr`, 2 occurrences sur `rennes.fr`)

Toutefois, les termes issus du corpus CIT-FR ne correspondent pas aux mêmes schémas lexicaux que ceux des thèmes précédents (AIDESOC-FR et ETR-FR) et il est donc nécessaire d'élaborer le nouveau patron suivant :

[nom][adjectif]

Il correspond à l'expression CQL suivante :

[tag="N.*"][tag="A.*"]

Cette recherche permet de relever de nouveaux termes que nous présentons dans la Table ci-dessous 5.28.

word	Frequency	Freq. per million	% of concordance
certificat médical	933	1526,45778	4,12357
pièces justificatives	568	929,29048	2,51039
état civil	448	732,96151	1,98002
autorité parentale	340	556,26543	1,5027
pension alimentaire	259	423,74337	1,1447
Certificat médical	220	359,93645	0,97233
Tribunal judiciaire	207	338,66748	0,91488
maison départementale	200	327,21496	0,88394
bilan auditif	192	314,12636	0,84858
affaires sociales	192	314,12636	0,84858
bilan ophtalmologique	192	314,12636	0,84858
établissement scolaire	165	269,95234	0,72925
Caisse nationale	153	250,31944	0,67621
tribunal judiciaire	105	171,78785	0,46407
enseignement supérieur	102	166,87963	0,45081
affaires familiales	97	158,69926	0,42871

TABLE 5.28 – Résultats les plus fréquents pour le patron [nom][adjectif]

Dans l'ensemble, les termes relevés dans la Table 5.28 ne correspondent pas à des groupes nominaux complexes. De plus, contrairement aux deux thèmes précédents, il

semble que la terminologie de ce domaine est bien établie.

À partir de nos connaissances et des résultats précédents, nous pouvons envisager une variation autour du complément **justificatif**. Nous avons donc fait une recherche grâce au patron suivant :

[nom][justificatif]

Ce patron correspond à l'expression CQL suivante :

[tag="N.*"][lemma="justificatif]

Nous n'avons trouvé que deux variants autour de cet adjectif : **document justificatif** et **pièce justificative**.

— **document justificatif**
— **pièce justificative**
— **justificatif**

Ces deux variants ont un sens identique et sont donc synonymes. Toutefois, nous pouvons également considérer que **justificatif** est un terme en soi, un terme adjectival, qui par glissement est devenu un nom, **un justificatif de : domicile, d'identité, etc.** Ce glissement, appelé dérivation impropre ou conversion, consiste donc en un changement de catégorie grammaticale sans modification de terme sur le plan formel ou morphologique. Cette hypothèse serait à vérifier dans une étude diachronique du discours administratif.

Tout comme **acte**, on retrouve un certain nombre de termes formés à partir de **justificatif** comme relevé dans le tableau 5.29 ci-dessous.

word	Frequency	Freq. per million	% of concordance
justificatif de domicile	34	55,62654	21,11801
justificatif d'identité	27	44,17402	16,77019
Justificatif de domicile	24	39,2658	14,90683
justificatifs de domicile	15	24,54112	9,31677
Justificatif d'identité	14	22,90505	8,69565
justificatif de taxe	6	9,81645	3,72671
justificatives par lettre	3	4,90822	1,86335
justificatives au consulat	2	3,27215	1,24224
justificatifs au directeur	2	3,27215	1,24224
justificatif de tarif	2	3,27215	1,24224
justificatifs d'absences	2	3,27215	1,24224
justificatif de moins	2	3,27215	1,24224
Justificatifs des prestations	2	3,27215	1,24224
justificatif de paiement	2	3,27215	1,24224
justificatif de vaccination	1	1,63607	0,62112
justificatives au directeur	1	1,63607	0,62112
justificatif des liens	1	1,63607	0,62112
Justificatifs des changements	1	1,63607	0,62112
Justificatif de concubinage	1	1,63607	0,62112
justificatifs de ressources	1	1,63607	0,62112

TABLE 5.29 – Résultats les plus fréquents pour le patron [**justificatif**][préposition][mot]

Ici, contrairement aux groupes nominaux complexes formés par les dispositifs d'aide sociale ou de droit au séjour, les termes ne prennent pas de sens particulier dans le domaine.

Toutefois, on retrouve un phénomène de spécialisation présent dans l'ensemble des corpus, l'utilisation d'acronymes. Le terme **carte d'identité** ou son variant **carte nationale d'identité** ont un sigle correspondant : **CNI**.

— **carte d'identité** (76 occurrences, 12 sur `paris.fr`, 24 sur `service-public.fr`, 21 sur `marseille.fr` et 16 sur `rennes.fr`)
— **carte nationale d'identité** (66 occurrences, 14 sur `paris.fr`, 18 sur `service-public.fr`, 17 sur `marseille.fr` et 18 sur `rennes.fr`)
— **CNI** (21 occurrences, 2 sur `paris.fr` 10 sur `marseille.fr` et 9 sur `rennes.fr`)

Ces analyses montrent que l'ensemble de ces variants sont utilisés de façon concomitante en corpus. Toutefois, on remarque que le sigle **CNI** n'est pas utilisé sur le site `service-public.fr`, alors même que pour le thème aide sociale ce site montrait une préférence pour l'utilisation des sigles (voir section 5.1.1.1).

D'autres acronymes et sigles présents dans le texte relèvent du thème aide sociale, comme illustré dans le tableau 5.30 ci-dessous.

word	Frequency	Freq. per million	% of concordance
MDPH	599	980,0088	13,73223
HT	314	513,72749	7,19853
AAH	299	489,18636	6,85465
PCH	297	485,91421	6,8088
CMI	293	479,36991	6,7171
CNSA	100	163,60748	2,29253
MSA	81	132,52206	1,85695
PARIS	66	107,98094	1,51307
CPJ	59	96,52841	1,35259
CAF	49	80,16766	1,12334
CASVP	45	73,62337	1,03164
TTC	45	73,62337	1,03164
CCAS	43	70,35122	0,98579
CIO	41	67,07907	0,93994

TABLE 5.30 – Acronymes et sigles relevés en corpus CIT-FR

Un exemple de sigle provenant du thème aide sociale mais présent dans le corpus CIT-FR est **AAH** pour **Allocations Adultes Handicapé**, ou encore **CMI** pour **Carte Mobilité Inclusion**.

En conclusion, la terminologie et notamment les groupes nominaux complexes du volet citoyen de notre corpus semblent bien établis. L'ensemble de ces termes sont stables et sont apparus avant la transition vers le numérique. Elle nécessite donc moins d'expertise spécifique au domaine et est bien documentée dans les glossaires et dictionnaires spécialisés.

5.1.1.8 Terminologie dans le corpus corpus anglais sur le thème de la citoyenneté

Pour étudier la terminologie du corpus CIT-AN, nous avons d'abord relevé les patrons syntaxiques les plus communs à l'aide de l'outil N-gram de Sketch Engine. Nous n'avons pas utilisé les patrons précédents, car nous n'étions pas sûre que ces patrons étaient applicables. Les relevés réalisés par l'outil sont représentés dans la capture d'écran (Figure 5.1) ci-dessous.

3–4-grams, tag (items: 19,630 , total frequency: 484,810)

	N-gram	Frequency		N-gram	Frequency
1	NP NP NP	7,080	14	NN IN DT NN	1,581
2	IN DT NN	4,571	15	VV DT NN	1,534
3	NP NP NP NP	3,623	16	IN PPZ NN	1,503
4	NN IN DT	2,954	17	NN NN IN	1,461
5	DT NN IN	2,893	18	NP NP NN	1,459
6	DT JJ NN	2,706	19	JJ NN NN	1,410
7	PP MD VV	2,104	20	IN DT NP	1,271
8	NN NN NN	1,878	21	NNS IN DT	1,226
9	NN IN NN	1,753	22	NN TO VV	1,222
10	IN DT JJ	1,701	23	IN DT NN IN	1,203
11	NN CC NN	1,697	24	IN DT JJ NN	1,154
12	JJ NN IN	1,622	25	TO VV DT	1,145
13	DT NN NN	1,609	26	IN NP NP	1,050

FIGURE 5.1 – Patrons syntaxiques les plus fréquents dans le corpus CIT-AN

À partir de ces résultats, nous avons sélectionné les patrons suivants :

- [adjectif][nom][nom]
- [nom][préposition][nom]
- [nom][nom][nom]

Ces patrons correspondent aux termes trouvés dans les corpus sur le thème de la citoyenneté et du droit au séjour. Ci-dessous, nous présentons les trois tableaux 5.31, 5.32 et 5.33 correspondant aux trois patrons.

word	Frequency	Freq. per million	% of concordance
30pm Councillor Surgery	79	232,48012	5,60284
civil partnership ceremony	31	91,22638	2,19858
civil partnership certificate	16	47,08458	1,13475
civil partnership dissolution	13	38,25622	0,92199
Annual Procurement Report	12	35,31344	0,85106
civil partnership Marriage	12	35,31344	0,85106
online application form	11	32,37065	0,78014
free school transport	11	32,37065	0,78014
new application form	11	32,37065	0,78014
entire guide Eligibility	9	26,48508	0,6383
entire guide Overview	8	23,54229	0,56738
local register office	8	23,54229	0,56738
civil marriage ceremony	7	20,5995	0,49645

TABLE 5.31 – Résultats les plus fréquents pour le patron [adjectif][nom][nom]

word	Frequency	Freq. per million	% of concordance
Name of Candidate	21	61,79851	1,19795
notice of marriage	21	61,79851	1,19795
advice in person	19	55,91294	1,08386
date of birth	18	52,97015	1,02681
Curriculum for Excellence	18	52,97015	1,02681
quality of life	16	47,08458	0,91272
Cost of service	15	44,1418	0,85568
Description of service	15	44,1418	0,85568
partnership into marriage	15	44,1418	0,85568
spread of coronavirus	13	38,25622	0,74159
Book of Remembrance	12	35,31344	0,68454
half of term	12	35,31344	0,68454
place of safety	9	26,48508	0,51341
home to school	9	26,48508	0,51341
book of remembrance	9	26,48508	0,51341

TABLE 5.32 – Résultats les plus fréquents pour le patron [nom][préposition][nom]

word	Frequency	Freq. per million	% of concordance
council tax bill	32	94,16916	1,70394
map PDF file	19	55,91294	1,01171
Map PDF file	18	52,97015	0,95847
@School _Services Comment	16	47,08458	0,85197
council tax advice	16	47,08458	0,85197
boundary map Councillor	16	47,08458	0,85197
garlic bread slice	15	44,1418	0,79872
council tax reduction	14	41,19901	0,74547
Scrutiny Panel Scrutiny	13	38,25622	0,69223
Profile PDF file	13	38,25622	0,69223
form PDF file	13	38,25622	0,69223
council tax discount	13	38,25622	0,69223
customer reference number	12	35,31344	0,63898
fruit bag tuna	11	32,37065	0,58573
profile PDF file	11	32,37065	0,58573

TABLE 5.33 – Résultats les plus fréquents pour le patron [nom][nom][nom]

Dans ces relevés, nous pouvons voir qu'il n'y a pas ou peu de phénomènes similaires à ceux observés dans les deux corpus précédents sur le thème du droit au séjour et de l'aide sociale. La Table 5.31 présente néanmoins des résultats similaires à ceux observés en français avec une variation au niveau de *civil partnership*, comme on peut le voir dans les premières lignes :

— *civil partnership* (321 occurrences, 2 sur `bristol.gov.uk`, 174 sur `york.gov.uk`, 67 occurrences sur `gov.uk`, 78 sur `falkirk.gov.uk`)

— *civil partnership ceremony* (41 occurrences sur `york.gov.uk`)
— *civil partnership certificate* (20 occurrences, 7 occurrences sur `gov.uk`, 3 occurrences sur `falkirk.gov.uk` et 10 sur `york.gov.uk`)
— *civil partnership dissolution* (13 occurrences, 1 occurrence sur `gov.uk`, 12 occurrences sur `york.gov.uk`)

Pour approfondir l'analyse de ces termes, nous avons étudié les compléments de *civil partnership*, présentés dans la Table 5.34.

word	Frequency	Freq. per million	% of concordance
civil partnership ceremony	31	91,22638	28,44037
civil partnership certificate	16	47,08458	14,6789
civil partnership dissolution	13	38,25622	11,92661
civil partnership Marriage	12	35,31344	11,00917
civil partnership registration	7	20,5995	6,42202
civil partnership record	6	17,65672	5,50459
civil partnership notice	4	11,77115	3,66972

TABLE 5.34 – Noms formés à partir de *civil partnership*

Dans les groupes nominaux relevés, contrairement aux autres thématiques, la modification est faite au niveau du complément, il n'y a pas de variation liée à la prémodification.

Cette terminologie est initialement utilisée pour le concept de mariage et a été transférée pour le concept de Pacs (*Civil Partnership*) comme le montrent les résultats ci-dessous (5.35).

word	Frequency	Freq. per million	% of concordance
marriage schedule	49	144,19653	23,44498
marriage venue	37	108,88309	17,70335
marriage certificate	27	79,45523	12,91866
marriage ceremony	18	52,97015	8,61244
marriage registration	12	35,31344	5,74163
marriage register	11	32,37065	5,26316
Marriage venue	8	23,54229	3,82775
marriage notice	8	23,54229	3,82775

TABLE 5.35 – Noms formés à partir de *marriage*

Une partie des termes sont présents dans les deux Tables (5.34 et 5.35) ce qui suggère un certain degré de concurrence ou de similitude entre les deux termes, *civil partnership* et *marriage*. Toutefois, certains termes employés en fonction de complément ne sont pas communs, tels que *schedule* puisqu'on le retrouve uniquement dans *marriage schedule*. Cela est dû à la différence de procédure. Dans les deux cas, les termes *certificate,*

ceremony, dissolution, registration, record, notice sont issus du discours administratif. Toutefois l'apposition de *marriage* ou *civil partnership* apporte des précisions sur le contexte précis d'utilisation et d'application de ces termes.

Nous avons également identifié des termes qui décrivent le domaine et sa transformation numérique comme *online application form* ou *new application form*. Nous avons retrouvé le même type de termes en français dans la section précédente.

— *online application form* (11 occurrences, 10 occurrences sur `falkirk.gov.uk` et 1 occurrence sur `york.gov.uk`)
— *new application form* (11 occurrences sur `falkirk.gov.uk`)

Dans le même domaine, nous pouvons discerner le terme *entire guide* issu du schéma phraséologique *Print entire guide*.

— *entire guide* (194 occurrences sur `gov.uk`)
— *Print entire guide* (194 occurrences sur `gov.uk`)

Les relevés montrent que le terme *entire guide* est présent uniquement dans ce schéma phraséologique et uniquement sur le site `gov.uk`. Cette terminologie et phraséologie répétée (voir 5.4.1) est issue du numérique puisqu'elle propose une solution alternative à l'utilisation des pages web en proposant une version non fragmentée des informations.

La Table 5.32 montrant les termes avec le motif [nom][préposition][nom] présente des caractéristiques assez similaires à la Table 5.31 montrant les termes avec le motif [adjectif][nom][nom] puisqu'on n'y observe pas de terminologie autour d'un concept, ou d'un nom noyau qu'on pourrait identifier comme caractéristique du discours sur le thème de la citoyenneté. La majorité des relevés ne sont pas des termes (*home to school*, *half of term*, *Description of service*, etc.). Certains des relevés sont, par contre, des termes bien attestés du discours sur le thème de la citoyenneté, tels que *date of birth*.

La Table 5.33, montrant les termes formés avec le motif [nom][nom][nom] présente différentes caractéristiques puisqu'elles montrent un grand nombre de termes et noms caractéristiques du discours numérique. Tout d'abord, on retrouve cinq occurrences du terme *PDF*, qui montre clairement le caractère numérique du discours. Le seul terme qui relève du discours administratif de l'accès aux droits que l'on peut observer dans ce tableau est *council tax*, qui revient dans trois occurrences (*council tax advice*, *council tax bill* et *council tax discount*) et dans des sources qui sont majoritairement des sites de municipalités.

— *PDF* (912 occurrences, tous sites)
— *council tax* (79 occurrences sur `bristol.gov.uk`, 262 occurrences sur `falkirk.gov.uk`, 1 occurrence sur `service-public.fr` et 323 sur `york.gov.uk`)

Pour compléter ces observations, nous avons analysé les acronymes et sigles les plus utilisés. Pour cela, nous avons fait une requête afin de trouver les suites de trois à quatre lettres en majuscules les plus utilisées. Les résultats de cette requête sont présentés dans la Table 5.36 ci-dessous.

word	Frequency	Freq. per million	% of concordance
PDF	774	2277,71663	35,31022
SNP	270	794,55231	12,31752
BSL	118	347,24879	5,38321
NHS	100	294,27863	4,56204
EPIA	48	141,25374	2,18978
EMA	44	129,4826	2,0073
CLD	38	111,82588	1,73358
VAT	36	105,94031	1,64234
HMO	28	82,39802	1,27737
HMRC	22	64,7413	1,00365
SLC	21	61,79851	0,95803
CCTV	20	58,85573	0,91241
CASC	19	55,91294	0,86679
CTB	15	44,1418	0,68431

TABLE 5.36 – Résultats pour l'expression régulière [A-Z]3,4 sur le corpus CIT-AN

L'analyse de ces résultats nous a permis d'identifier qu'il s'agit pour la plupart d'acronymes issus d'autres domaines. Le premier résultat, *PDF (Portable Document Format)*, illustre ce phénomène et complète les observations réalisées sur les résultats précédents (Table 5.33). Certains termes peuvent être assimilés à la thématique aide sociale, comme *EMA (Education Maintenance Allowance)* ou *HMO (House in Multiple Occupation)*, qui fait référence à un dispositif d'aide. D'autres concernent les différents organismes publics comme *NHS (National Health Service) HMRC (HM Revenue & Customs)* ou encore des partis politiques comme le *SNP (Scottish National Party)*. Ces acronymes et sigles sont en lien avec l'administration, mais ne sont pas spécifiques à l'accès aux droits. D'autres sont issus de domaines connexes, comme le commerce avec *VAT (Value Added Tax)*. Le *CASC (Community Amateur Sports Club)* et *CTB (Council tax base)* peuvent être considérés comme faisant partie de la thématique, mais chacun de ces sigles sont utilisés une fois. Les acronymes et sigles ne sont donc pas une spécificité dans la thématique citoyenneté.

5.1.1.9 Analyse contrastive des résultats des corpus CIT-FR et CIT-AN

La thématique citoyen semble faire figure d'exception dans notre corpus car sa terminologie ne présente pas de nombreux groupes nominaux complexes. En français, nous

retrouvons certains termes autour du noyau *acte*, mais dans une mesure restreinte. Certains acronymes ou sigles sont également présents comme *CNI, PACS*. Toutefois, il ne s'agit pas d'un système complexe comme celui présent dans les thématiques droit au séjour et aide sociale. Dans le corpus britannique, les termes comme *civil partnership* et *marriage* sont présents et se combinent à d'autres termes issus de l'administration comme *registration* ou *notice* pour former des termes composés. Comme dans le corpus français, il s'agit ici de préciser l'objet désigné et nullement de l'instauration d'un système terminologique complexe comme dans le corpus consacré aux aides sociales.

La thématique citoyen semble donc présenter des caractéristiques différentes par rapport aux deux autres thématiques. Les groupes nominaux complexes ne forment pas un système conceptuel complexe, et nous n'avons pas relevé d'instabilité liée à la variation terminologique.

Plusieurs facteurs peuvent expliquer ce phénomène. Tout d'abord, le sous-domaine existe depuis plus longtemps et il est géré par des organismes plus stables, directement liés aux gouvernements et aux municipalités. La terminologie n'a pas ou peu été modifiée par des nouvelles législations, contrairement à la terminologie des aides sociales et du droit au séjour qui fait régulièrement l'objet de nouvelles lois.

5.1.2 Synthèse de l'analyse terminologique anglais-français des trois thématiques

Nous avons identifié un certain nombre de caractéristiques terminologiques dans la formation des dénominations de dispositifs des différentes thématiques de notre corpus.

Tout d'abord, l'analyse de la thématique aide sociale nous a permis d'identifier des similarités entre les principes de formation des termes en anglais et en français. Nous avons conclu à la présence d'un discours spécialisé des aides sociales dans les deux langues, formant un système terminologique et phraséologique complexe.

Nous avons ensuite étudié la thématique du droit au séjour, dans le sous-corpus réservé aux droits des étrangers vivant en France. Dans ce corpus, nous avons exploré les dispositifs relatifs à l'immigration et au droit au séjour. Nous avons conclu à l'existence d'un discours spécialisé en français, qui repose sur une terminologie complexe autour des **documents de séjour**. Toutefois, nous n'avons pas identifié de discours spécialisé propre au domaine de l'accès aux droits pour le corpus britannique. En effet, il semble plutôt que la terminologie et la phraséologie employées soient issues du discours juridique.

Enfin, nous avons réalisé une étude de la thématique de la citoyenneté. Pour cette thématique, nous avons identifié certains éléments terminologiques et phraséologiques.

Il existe bien un discours spécialisé de l'accès aux droits des citoyens. Toutefois, ce discours semble stable et bien documenté à l'aide de glossaires par exemple.

5.1.3 Variation terminologique

Nos observations de la terminologie du domaine en corpus nous ont permis d'arriver à la conclusion que les groupes nominaux complexes sont sujets à la variation terminologique. Nous avons pu identifier plusieurs types de variations terminologiques présentes dans notre discours, que nous allons étudier dans cette section. D'après Sinclair, un manque de clarté dans la terminologie peut avoir des conséquences sur la compréhension du message (Sinclair, 2005b). Il nous a donc paru essentiel d'explorer les différents types de variations présents en corpus. La liste ci-dessous montre les différentes variations repérées dans les deux langues :

— variation au niveau du nom non stabilisé
— variation au niveau du noyau
— variation acronyme/nom
— variation graphique

Nous allons dans un premier temps nous intéresser à la variation au niveau du nom non stabilisé illustré par le concept d'allocation chômage dans les deux langues. Cette variation peut être attribuée à l'évolution de la législation et aux phénomènes de disparition ou raréfaction de termes qui sont remplacés par d'autres et qui, à la manière de ce qui se passe dans le discours scientifique sous l'influence du progrès, sont sur le point de devenir obsolètes. Afin de confirmer ce phénomène, il faudrait, comme le suggèrent Dury et Drouin (2010), regarder l'évolution du terme sur plusieurs années. Toutefois, nous pouvons noter que ces opérations peuvent être difficiles, car la mise à jour constante des pages ne nous permet pas de trouver un corpus figé dans le temps. Le corpus récolté pour cette thèse pourrait donc servir de corpus initial, afin d'observer l'évolution de la terminologie. Cette instabilité du discours administratif n'est pas nouvelle et ne peut être attribuée uniquement à la composante numérique du discours. En effet, Langrod, dans une étude de la terminologie administrative de 1953, fait les conclusions suivantes :

> « on tolère des confusions infinies, des malentendus sémantiques et souvent un véritable chaos terminologique » (Langrod, 1953, p. 1)

Il indique également que ce chaos est « sans prétention à créer un cadre doctrinal propre » (Langrod, 1953, p. 1). Ces deux citations montrent que l'instabilité des termes ou le « chaos terminologique » est un phénomène attesté et documenté dans l'histoire de la terminologie de l'administration. Nous allons analyser la terminologie actuelle pour voir si cela est toujours le cas et quelles pourraient être les raisons de ce « chaos ». Les

termes non stabilisés sont en particulier liés à des aides répandues, qui sont accessibles à un grand nombre de personnes. Ces termes font l'objet de discussion dans les médias, et plus généralement, leur concept est présent dans la langue générale.

5.1.3.1 Analyse contrastive du concept d'allocation chômage/*Jobseeker's allowance* en corpus

Dans cette section, nous allons démontrer nos propos quant à la variation liée à l'instabilité à travers une analyse contrastive du concept d'allocation chômage dans nos deux corpus.

5.1.3.1.1 Analyse en corpus français

Nous allons nous concentrer sur la variation terminologique autour du concept d'allocations chômage afin de procéder à une analyse fine des divers variants. La Table 5.37 ci-dessous présente l'ensemble des différents termes faisant référence à ce concept, ainsi que le nombre d'occurrences dans le corpus, tous thèmes confondus.

Terme	Occurrences
allocation chômage	571
allocation d'aide au retour à l'emploi	43
ARE	474
assurance chômage	85
allocation de retour à l'emploi	19
allocation chômage d'aide au retour à l'emploi	4

TABLE 5.37 – Récapitulatif des termes relevés autour du concept d'allocation chômage

La Table 5.37 propose un récapitulatif des termes relatifs à l'allocation chômage, c'est-à-dire à l'allocation reçue sous certaines conditions lorsque l'on perd son travail. Nous retrouvons quatre variants du terme. Le terme principal **allocation chômage** est la dénomination la moins spécialisée. C'est celle qui est connue du grand public et qui est présente dans d'autres types de discours, notamment journalistique. Pour autant, on retrouve également ce terme sur les sites spécialisés qui composent notre corpus. Dans notre corpus, il fait référence au dispositif spécifique de l'allocation chômage, comme on le voit dans ces contextes.

(15) Le salaire journalier de référence (SJR) sert à calculer vos allocations chômage.

(16) Pour toucher des allocations chômage, vous devez avoir travaillé une période minimale de travail dite durée d'affiliation.

(17) Cependant, le montant de l'allocation chômage sera calculé sur vos seuls salaires perçus en France.

Dans ces trois contextes, **allocation chômage** fait référence au dispositif mis en place par Pôle emploi. Les explications font en effet référence à un dispositif spécifique, en énonçant les conditions de perception de l'allocation. Dans le premier contexte, on retrouve une référence à la base de calcul de l'allocation, le « salaire journalier de référence (SJR) ». Dans le deuxième contexte, on retrouve la condition de la « période minimale de travail ». Enfin, dans le dernier contexte, on retrouve une autre condition, celle de percevoir ses salaires en France afin de valider la période minimale de travail.

Toutefois, d'autres utilisations du terme sont plus générales. Dans certains cas, **allocation chômage** ne fait pas référence à un dispositif particulier, mais plutôt à la possibilité de recevoir une allocation lorsque l'on perd son travail.

(18) Le droit aux allocations chômage diffère selon que vous reveniez d'un pays de l'Espace économique européen (EEE), de Suisse ou d'un autre pays. (caf.fr)

(19) Vous devez prioritairement faire valoir vos droits à l'ensemble des autres prestations sociales (allocation chômage, retraite...) auxquelles vous pouvez prétendre. (service-public.fr)

Dans ces deux contextes, l'utilisation des termes fait plutôt référence au concept général de l'allocation chômage (qui existe dans plusieurs pays) par rapport à d'autres prestations.

Enfin, on retrouve le terme **allocation chômage** dans un grand nombre de titres et de phrases répétées (en-tête, pied de page, liens, etc.)

(20) Prolongation des droits à l'allocation chômage pour les demandeurs d'emploi en fin de droits (paris.fr)

(21) Prime exceptionnelle et allocation chômage : quelle articulation ? (paris.fr)

(22) Estimer le montant de ses allocation chômage Pôle emploi (service-public.fr)

Les contextes ci-dessus montrent l'utilisation d'**allocation chômage** dans les titres, en-tête et pieds de page des sites. Sur les 571 occurrences d'**allocation chômage**, 493 sont des titres ou des phrases multipages (que l'on retrouve sur plusieurs pages différentes).

Le concept d'**allocation chômage** a également une désignation plus juridique et plus spécialisée. Celle-ci provient directement du texte de loi : **allocation d'aide au retour à l'emploi**. La première version de cette allocation est mise en place en 2000 (arr, 2000), puis elle a été modifiée pour la dernière fois en 2019 (dec, 2019).

Ce terme fait référence au dispositif spécifique, comme le montrent les contextes suivants :

(23) L'allocation d'aide au retour à l'emploi (ARE) est un revenu de remplacement versé par Pôle emploi, sous certaines conditions, aux personnes inscrites comme demandeurs d'emploi et involontairement privées d'emploi.

Il existe des définitions du dispositif qui se situent généralement dans le chapô de la page web. Le chapô est un texte court situé entre le titre et le corps du texte d'un article (Gué, 2023). Dans les pages étudiées, il contient généralement une définition générale des dispositifs abordés. Il est particulièrement important en rédaction web car l'inclusion de mots-clés et de termes dans le titre et le chapô permettent de faire remonter l'article dans les moteurs de recherche (Deseilligny, 2022).

On retrouve également une utilisation d'**allocation d'aide au retour à l'emploi** similaire à celle d'**allocation chômage**, comme le montre le contexte suivant.

(24) La durée de versement de l'allocation d'aide au retour à l'emploi (ARE) dépend de votre durée d'affiliation lors des 36 derniers mois.

Cette phrase montre une utilisation du terme pour faire référence au dispositif spécifique et à ses conditions. On trouve donc des acceptions qui sont des quasi-synonymes de **allocation chômage**. Dans d'autres cas, les deux termes sont utilisés pour faire référence à des dispositifs différents.

(25) Vous touchez une allocation chômage inférieure ou égale à l'allocation d'aide au retour à l'emploi (ARE) (soit 31,59 € par jour) (service-public.fr)

Dans cette définition, le rédacteur différencie l'**allocation chômage** et l'**allocation d'aide au retour à l'emploi**, qui serait un type d'allocation chômage. Ces différentes utilisations des termes se trouvent toutes les deux sur le site officiel de l'administration française `service-public.fr`. Toutefois, l'**allocation chômage** est parfois assimilée à l'**allocation d'aide au retour à l'emploi**, comme on le voit dans le contexte suivant.

(26) (...) si vous touchez l'allocation chômage d'aide au retour à l'emploi (ARE) ou l'allocation de solidarité spécifique (ASS) (Service-public.fr)

Dans ce contexte, les deux termes sont combinés pour créer un nouveau variant **allocation chômage d'aide au retour à l'emploi**. L'acronyme utilisé pour faire référence aux **allocations d'aide au retour à l'emploi** est largement utilisé dans notre corpus. Le variant **ARE** (474 occurrences) rassemble le plus d'occurrences par rapport à tous les variants d'**allocation chômage**. L'ensemble des occurrences proviennent des sites `service-public.fr` et `paris.fr`. Il semble donc que son utilisation ne soit pas commune sur tous les sites de notre corpus. Toutefois, sur le site `service-public.fr`, il s'agit du variant le plus utilisé.

Son utilisation est similaire à celle des deux variants précédents (**allocations chômage** et **allocations d'aide au retour à l'emploi**), car il fait référence au dispositif spécifique et à ses conditions comme le montrent les citations suivantes.

> Remplir les conditions d'attribution de l'**ARE** (autres que la condition de privation involontaire d'emploi) (service-public.fr)

> Oui, pour toucher l'**ARE**, vous devez être inscrit comme demandeur d'emploi ou accomplir une formation inscrite dans votre projet personnalisé d'accès à l'emploi (PPAE). (service-public.fr)

Ces deux contextes mentionnent la privation involontaire d'emploi qui est l'une des principales conditions d'obtention de l'aide. L'utilisation est similaire à **allocation chômage**. Dans certains cas, leur utilisation peut également être interchangeable.

> Peu importe le montant de l'**allocation chômage** perçue (supérieure ou non à l'**ARE** minimale). (service-public.fr)

Dans le contexte ci-dessus, **ARE** et **allocation chômage** semblent être synonymes. **ARE** est utilisé de façon interchangeable avec **allocation chômage**.

Un autre variant présent dans notre corpus est **assurance chômage**. Il s'agit d'un terme plus ancien, qui fait également référence au système de prélèvement sur les salaires qui permet ensuite d'obtenir l'allocation. Dans notre corpus, il s'agit également d'un variant d'**allocation chômage** et son utilisation est similaire. Il s'agit d'un terme devenu rare, car il date de lois publiées en 1958. Toutefois, il reste utilisé comme variant. Il est parfois utilisé à la place d'**allocation d'aide au retour à l'emploi**, comme le montre le contexte ci-dessous.

> La rémunération mensuelle est équivalente au montant de l'allocation d'**assurance chômage** (ARE). (service-public.fr)

L'acronyme utilisé entre parenthèses est identique à celui de l'**allocation d'aide au retour à l'emploi**. Dans cet exemple, les deux variants **assurance chômage** et **allocation d'aide au retour à l'emploi** sont donc assimilés au même terme.

Plus généralement, **assurance chômage** est utilisé majoritairement dans les titres des pages et des sections. Sur les 83 occurrences dénombrées dans le corpus, 39 sont en fait des titres, répétés pour la plupart sur plusieurs pages.

(27) Quelle que soit votre situation, vérifiez vos droits à l'assurance chômage. (pole-emploi.fr)

(28) L'essentiel à savoir sur l'assurance chômage (pole-emploi.fr)

Ces deux titres figurent sur le site de Pôle emploi, ce qui indique qu'ils sont utilisés comme variants du concept.

Certaines utilisations montrent également une combinaison d'**allocation** avec **assurance chômage**, présentée dans le contexte suivant :

(29) Si vous êtes en fin de droit, vous devez informer l'organisme assureur de la cessation du versement des allocations d'assurance chômage.

Dans ce contexte, les rédacteurs associent les deux termes **allocation chômage** et **assurance chômage**, ce qui montre l'instabilité ou possiblement une réintroduction du concept principal qui est **allocation**.

Nous pouvons observer que le maintien de ce type de variation permet de créer un fil reliant les dispositifs de longue date aux dispositifs actuels, auquel cas la variation a pour fonction d'indiquer qu'il s'agit de mêmes types de dispositifs. Néanmoins elle demeure un obstacle dans la compréhension des types et sous-types de dispositifs, si bien que les indications explicites pour assurer ce lien chronologique seraient nécessaires, de type :

> *Si vous êtes en fin de droit, vous devez informer l'organisme assureur de la cessation du versement des **allocations d'aide au retour à l'emploi** (autrefois appelées **allocations chômage** et **assurance chômage**, et parfois encore communément appelées **allocations chômage** ou **ARE**).

Enfin, nous retrouvons des variants que nous qualifions de « dérivés ». Ces variants sont utilisés sur plusieurs sites et sont formés à partir d'une dérivation, c'est-à-dire d'une modification du terme initial. Ils sont minoritaires dans le corpus et ne proviennent pas de la législation. Le premier variant dérivé est **allocation de retour à l'emploi**. Il s'agit d'une simplification du terme juridique initial **allocation d'aide au retour à l'emploi**. Toutefois, son utilisation ne semble pas venir d'un effort concerté de simplifier le terme juridique. Il s'agit probablement d'une décision, ou d'une erreur singulière, qui s'est répercutée dans le corpus. Ce variant est présent sur deux sites : le site officiel de l'administration et le site de la ville de Paris. Il y a 12 occurrences de ce variant dans notre corpus, dont 9 sont suivis de l'acronyme. Sur le site `service-public.fr`, ce type de variants se trouvent parfois dans les pages dont le sujet principal est l'allocation chômage. Dans les deux pages suivantes, le variant dérivé est utilisé dans le chapô de la page :

1. Un salarié peut-il toucher l'allocation chômage en cas de démission ?
2. Un agent public peut-il percevoir l'allocation chômage en cas de démission ?

Sur la page 1, la phrase présentant le variant est la suivante :

(30) En principe, la démission ne vous permet pas de bénéficier de l'allocation de retour à l'emploi (ARE). (service-public.fr)

Sur la page 2, la phrase présentant le variant est la suivante :

(31) En principe, vous n'avez pas droit à l'allocation chômage (allocation de retour à l'emploi— ARE) lorsque vous démissionnez. (service-public.fr)

Ces phrases sont toutes deux issues de la présentation du sujet de la page. Il s'agit d'extraits qui semblent contenir des ressemblances dans les principes de rédaction, mais qui contiennent également des différences. Ces extraits seront analysés plus en détail et dans leur environnement numérique dans la section 5.1.3.2.

Les autres occurrences des variants se trouvent sur des pages qui ne sont pas toujours directement liées aux allocations chômage. Les pages sont les suivantes :

— Peut-on cumuler la pension d'invalidité avec d'autres revenus ?
— Contrat de sécurisation professionnelle (CSP)
— Dépenses d'énergie : les aides de la Ville de Paris

Dans le cas de ces trois pages, l'utilisation de ce variant pourrait venir d'un rédacteur non spécialisé dans les allocations chômage, qui a pu modifier accidentellement le terme juridique initial. La structure et la nature des sites web a pour conséquence la répercussion de cette utilisation sur plusieurs autres pages.

Il existe un deuxième variant modifié qui est **allocation chômage d'aide au retour à l'emploi** (ARE) et qui constitue donc la combinaison des deux variants **allocation chômage** et **allocation d'aide au retour à l'emploi**. Ce variant est uniquement présent sur le site officiel de l'administration française. Nous avons relevé 4 occurrences de ce variant en corpus. Ces occurrences semblent être présentes dans les titres des pages. Il semble donc possible que l'association de ces deux variants ait été faite pour garantir la compréhension ou l'apparition de la page dans les moteurs de recherche. En effet, il est possible d'accéder à la page du site directement depuis un moteur de rechercheur. L'apparition des deux variants dans le titre permet à l'usager non seulement de trouver cette page, mais également de comprendre son titre, peu importe le variant qui lui est familier. Dans ce cas, nous pouvons émettre l'hypothèse qu'il ne s'agit pas d'un variant dérivé, mais d'une modification volontaire dans un but de clarification.

La Table 5.38 propose un récapitulatif des différents variants et de leurs contextes d'utilisation.

Utilisation	Terme
Général	allocation chômage, assurance chômage, ARE
Dispositif spécifique	allocation chômage, assurance chômage, ARE, allocation d'aide au retour à l'emploi
Termes apparaissant dans la législation	assurance chômage, allocation d'aide au retour à l'emploi
Titre	Allocation chômage, assurance chômage, allocation chômage d'aide au retour à l'emploi
Termes modifiés	allocation chômage d'aide au retour à l'emploi, allocation de retour à l'emploi

TABLE 5.38 – Récapitulatif des modes d'utilisation des variants d'allocation chômage

La Table 5.39 récapitule l'ensemble des variants autour du concept d'**allocation chômage** avec le nombre d'occurrences en corpus. Nous avons ensuite proposé un commentaire pour chaque variant qui décrit sa place dans le discours. Pour consulter des exemples d'utilisation, voir l'annexe .2.

Terme	Occ.	Commentaire
allocation chômage	571	Dénomination « générale », mais présente sur les sites spécialisés. Terme très présent dans les titres (sur les 571 occurrences, 493 sont issues de titre).
allocation d'aide au retour à l'emploi	43	Terme qui figure dans la législation.
ARE	474	Terme uniquement utilisé dans la communication grand public ou par les experts entre eux.
Allocation de retour à l'emploi	19	Utilisation dérivée par rapport au terme juridique qui est « allocation d'aide au retour à l'emploi ». Mais il est utilisé sur Légifrance 1 fois (Communiqué de presse du conseil des ministres du 25 novembre 2020).
Assurance chômage	85	Terme introduit en 1958, qui n'est plus le terme désignant le dispositif actuel d'allocation chômage mais qui apparaît néanmoins dans le corpus.

TABLE 5.39 – Récapitulatif des termes et de leur utilisation en corpus autour du concept de l'allocation chômage

5.1.3.1.2 Analyse en corpus britannique

Dans le corpus britannique, le concept d'allocation chômage fait également l'objet de variations. Nous avons exploré le corpus afin de faire un récapitulatif de l'ensemble des variants autour de ce concept. Le premier terme que nous avons exploré en corpus

est *unemployment benefits* (15 occurrences), qui est rarement utilisé. On le retrouve dans certaines descriptions ou présentations de pages :

(32) Ideally, a Jobcentre gives out advice on the available jobs and also facilitates unemployment benefits to British citizens."

Toutefois, l'utilisation de ce variant n'est pas répandue et le concept est plus souvent exprimé par une paraphrase, notamment lorsqu'il apparaît dans un menu comme dans le contexte ci-dessous :

(33) Benefits and financial support if you're temporarily unable to work.

Il s'agit d'un emploi général, qui ne vient pas du domaine juridique et qui est peu présent en corpus.

Le deuxième terme étudié est celui de *work allowance* (23 occurrences), qui est la composante dédiée au travail dans le dispositif d'aide *Universal Credit*. Nous n'avons trouvé que deux cooccurrences entre *Universal Credit* et *work allowance*, qui montrent le lien entre les deux termes, comme dans le contexte ci-dessous.

(34) Read more about Universal Credit Work Allowances"

La majorité des occurrences en corpus indiquent son fonctionnement et ses conditions d'accès :

(35) The Work Allowance only applies to you if : you have responsibility for one or more children (or qualifying young persons)

Nous n'avons pas trouvé de sigle correspondant en corpus. Ce dispositif remplace l'ensemble des aides financières versées aux personnes sans emploi et le peu d'occurrences peut paraître surprenant, notamment par rapport aux nombreuses occurrences de *Universal Credit* (environ 2 000 occurrences) ou *Jobseeker's Allowance* (326 occurrences).

Le troisième terme étudié est donc *Jobseeker's Allowance* qui était, jusqu'à la transition vers *Universal Credit*, le dispositif principal et donc le terme principal. Il y a 326 occurrences de ce terme en corpus. Il s'agit d'un nombre d'occurrences ajusté par rapport aux autres termes formés à partir de *Jobseeker's Allowance*. Il s'agit du terme juridique, retrouvé dans les différentes législations concernées (job, 2022). Ce terme est utilisé dans plusieurs situations. On le retrouve tout d'abord avec d'autres dispositifs issus du même concept :

(36) Universal Credit, Jobseeker's Allowance or Income Support (gov.uk)

La citation ci-dessus fait référence aux différents types d'aide disponible.

Nous trouvons également des définitions du dispositif, comme dans le contexte ci-dessous :

(37) Jobseeker's allowance is the benefit paid to unemployed people who are actively looking for work.

Ce terme est également présent dans de nombreuses phrases répétées, titres et menus.

(38) (...) You can apply for Jobseeker's Allowance (JSA) to help you when you look for work. (23 occurrences)

(39) (...) Jobseeker's Allowance and low income benefits (138 occurrences)

Il semble que ce terme soit donc un repère pour les usagers sur les sites. Le premier exemple nous permet également d'identifier le sigle *JSA* que nous allons maintenant analyser.

Le sigle *JSA* compte 491 occurrences, sans compter les occurrences des autres termes formés à partir de ce sigle. Le sigle est donc plus fréquemment utilisé que la forme déballée du terme. Cela montre une certaine spécialisation du discours, puisque des termes spécifiques au discours de l'accès aux droits sont apparus. En effet, *JSA* n'apparaît pas dans le discours juridique correspondant.

Il semble que le sigle soit utilisé pour donner des précisions, des détails et explications sur le dispositif, contrairement à la forme complète qui est utilisée plus souvent dans les menus et hyperliens.

(40) You do not need to go to JSA interviews and appointments at the moment because of coronavirus (COVID-19).

(41) If you don't keep to your agreement your JSA could be stopped or reduced - this is called a sanction.

Ces deux contextes montrent que le sigle est utilisé dans des contextes spécialisés dans le but de fournir des précisions sur ce dispositif, et non pour présenter le concept général comme le terme complet présenté précédemment (*Jobseeker's Allowance* dans le contexte 37).

Par ailleurs, nous avons identifié deux types de *Jobseeker's Allowance* qui sont présents en corpus : le *contribution-based Jobseeker's Allowance* (6) et le *income-based Jobseeker's Allowance* (77). Dans ces deux cas, il s'agit d'une précision du mode de calcul de l'aide financière qui est juxtaposée au terme. Nous pouvons observer avec ces deux exemples le recours à une matrice typique du discours scientifique, la prémodification (Mestivier, 2015), que l'on retrouve par exemple dans des termes tels que *corpus-based studies* ou *Artemisinin-based combination therapy*. Ces exemples pourraient montrer une tendance globale de la terminologie anglaise à rendre les concepts spécialisés

transparents en opérant une dénomination selon ce type de matrice. Elle consiste à recourir au processus d'emballage (ou *packing*, cf. Halliday et Webster (2006)) pour condenser les informations qui sont habituellement exprimées dans les propositions, et ainsi d'éviter l'utilisation des prépositions (ici « *on* ») et des termes spécifiques (ici « *calculated* » qui est remplacé par « *-based* ») :

JSA (which is) calculated/based on income.

Nous remarquons une grande disparité dans l'utilisation des deux aides en corpus. L'aide *income-based Jobseeker's Allowance* est plus fréquemment accordée, ce qui peut expliquer cette disparité. Chacun de ces termes a un variant utilisant le sigle : *contribution-based JSA* (24 occurrences) et *income-based JSA* (151 occurrences). Les occurrences intégrant le sigle sont plus fréquentes que les termes complets, ce qui suggère encore une fois une spécialisation du discours, comme le montrent les deux contextes ci-dessous :

(42) If your income-based JSA claim is ending because you're making a new claim for Universal Credit, you'll automatically continue to get the amount of JSA you currently receive, as long as you're still eligible.

(43) The benefits that can be used are : Universal Credit income-based Jobseeker's Allowance income-related Employment and Support Allowance Income Support Pension Credit

Ces deux contextes montrent une utilisation différente des deux variants (sigle et terme complet). Le sigle est utilisé lors de précisions et d'explications concernant le dispositif. Le terme complet est utilisé dans des listes ou fragments de texte servant au repérage par l'utilisateur. Le sigle est utilisé pour donner des précisions à l'usager qui perçoit déjà l'aide. On remarque à nouveau une disparité entre spécialisation et non-spécialisation du discours.

Il est également intéressant de noter que nous avons trouvé 3 occurrences d'une dénomination « dérivée », voire fautive, du dispositif *income-related JSA*, présentée dans les contextes suivants :

(44) If you're eligible for income-related JSA you can get up to :

(45) (...) if you get income-related JSA as a couple

(46) You can get both contribution-based JSA and income-related JSA at the same time.

Ces occurrences se trouvent sur le site `citizensadvice.org.uk`, où se trouvent également 37 occurrences de *income-based JSA*. Nous pouvons donc supposer qu'il

s'agit d'une dérivation non intentionnelle du terme, qui n'est recensée nulle part ailleurs.

Un autre terme a été créé à partir des termes ci-dessus, il s'agit du *New style Jobseeker's Allowance* (4 occurrences). Avec ce terme, nous observons la formation de nouveaux termes en langue de spécialité par apposition ou juxtaposition, mais aussi en recourant à des termes très peu spécialisés, voire pas spécialisés, puisque *New Style* fait partie de la langue général (Mestivier, 2015). Ce terme a plusieurs variants comme le *'new style' JSA* avec 112 occurrences, puis le *new style JSA* avec 33 occurrences. Le premier variant est plus fréquent et il est présent dans des menus et des listes comme dans le contexte ci-dessous :

(47) Make a claim for 'new style' JSA or contribution or income-based JSA

(48) Apply for 'new style' JSA

L'utilisation des guillemets et le choix du terme juxtaposé montrent une conscience linguistique des dénominations du dispositif. En effet, dans ce cas, on ne cherche plus à décrire le statut de l'usager ou la nature de l'aide, mais on la différencie des précédentes.

Enfin, nous nous intéressons pour clore notre étude de la variation du concept de *unemployment benefit* aux termes suivants : *Employment and Support Allowance* et *New Style Employment and Support Allowance* (256 occurrences). Il s'agit d'autres types d'aides pour les personnes sans emploi. Le premier dispositif (*Employment and Support Allowance*) présente également un variant avec un sigle, *ESA*, qui compte 889 occurrences. Le deuxième dispositif montre une utilisation d'un terme juxtaposé avec *New Style Employment and Support Allowance* (6 occurrences) et un variant, *New Style ESA* (40 occurrences). L'utilisation de *New Style* montre à nouveau qu'il y a un effort de simplification ou de clarification du système complexe d'aides. Nous pouvons noter qu'il y a une reprise de la même séquence juxtaposée dans plusieurs termes, comme nous l'avions déjà observée dans la section 5.1.1.2. Cette stabilité linguistique permet à l'usager de comprendre en trouvant une certaine cohérence dans le discours et dans la construction des informations. En effet, la répétition des mêmes items lexicaux joue un rôle cohésif.

Le concept d'*unemployment benefit* présente donc de nombreux termes avec des niveaux de spécialisation différents. Certains sont équivalents tandis que d'autres font référence à différents types de calcul ou de bénéficiaire. Le concept présente certaines caractéristiques de stabilité comme la reprise d'éléments de juxtaposition, mais d'autres éléments montrent davantage une instabilité, comme l'utilisation de guillemets et de sigles, ou encore la multiplication des différentes aides.

La Table ci-dessous présente l'ensemble des variations autour du concept de *unemployment benefits*. Pour consulter des exemples d'utilisation, voir l'annexe .1.

Termes	Occ.	Commentaire
New Style Jobseeker's Allowance (JSA)	4	Terme complet avec sigle.
'new style' JSA	112	Terme principalement utilisé dans les titres de pages ou dans les sous-titres.
new style JSA	33	Terme présentant peu d'occurrences sur gov.uk (3). Il est plutôt utilisé sur les autres sites.
income-based Jobseeker's Allowance (JSA)	77	Terme désignant une aide pour laquelle il n'est plus possible d'ouvrir des droits. Toutefois, certains usagers la perçoivent encore. Le terme est presque systématiquement utilisé dans des listes contenant d'autres aides sociales.
income-based JSA	151	Terme désignant une contribution qui fait référence aux cotisations réalisées dans le cadre d'une activité salariée, afin de percevoir le chômage. Collocation : "contribution or income-based JSA".
contribution-based JSA	24	Terme désignant une contribution qui fait référence aux cotisations réalisées en dehors d'une activité salariée afin de percevoir le chômage.
Jobseeker's Allowance	326	Terme complet tel qu'il apparaît dans la législation.
Unemployment benefits	15	Terme général, qui n'apparaît presque pas.
Income Support	439	Terme désignant une aide pour laquelle il n'est plus possible d'ouvrir des droits. Toutefois, certains usagers la perçoivent encore. Le terme est presque systématiquement utilisé dans des listes contenant d'autres aides sociales.
Employment and Support Allowance	256	Terme complet.
new style Employment and Support Allowance	4	Terme complet avec juxtaposition.
Employment Support allowance	33	Utilisation « dérivée ».
Work allowance	23	Il s'agit d'une composante du nouveau dispositif *Universal Credit* dédiée aux versements destinés aux personnes en recherche d'emploi
JSA (sans parenthèses ni compléments)	371	Terme uniquement présent dans le discours de l'accès aux droits.
income-related JSA	3	Utilisation « dérivée ».

TABLE 5.40 – Récapitulatif des termes et de leur utilisation autour du concept de *unemployment benefits*

5.1.3.2 Analyse contrastive du concept d'allocation chômage / *Jobseeker's allowance* dans son environnement textuel

Nous allons maintenant comparer l'utilisation des termes autour du concept d'allocation logement dans leur environnement délinéarisé.

5.1.3.2.1 Analyse en texte français

Nous réaliserons tout d'abord une étude du concept sur le site français. Pour illustrer nos analyses, nous avons sélectionné deux chapôs de pages, qui présentent la description générale du sujet de la page. Ces deux pages concernent les conditions de perception du chômage en cas de démission. Les textes ci-dessous sont présentés dans leur environnement, dans une capture d'écran du texte. La Figure 5.2 est le chapô de la page « Un salarié peut-il toucher l'allocation chômage en cas de démission ? ».

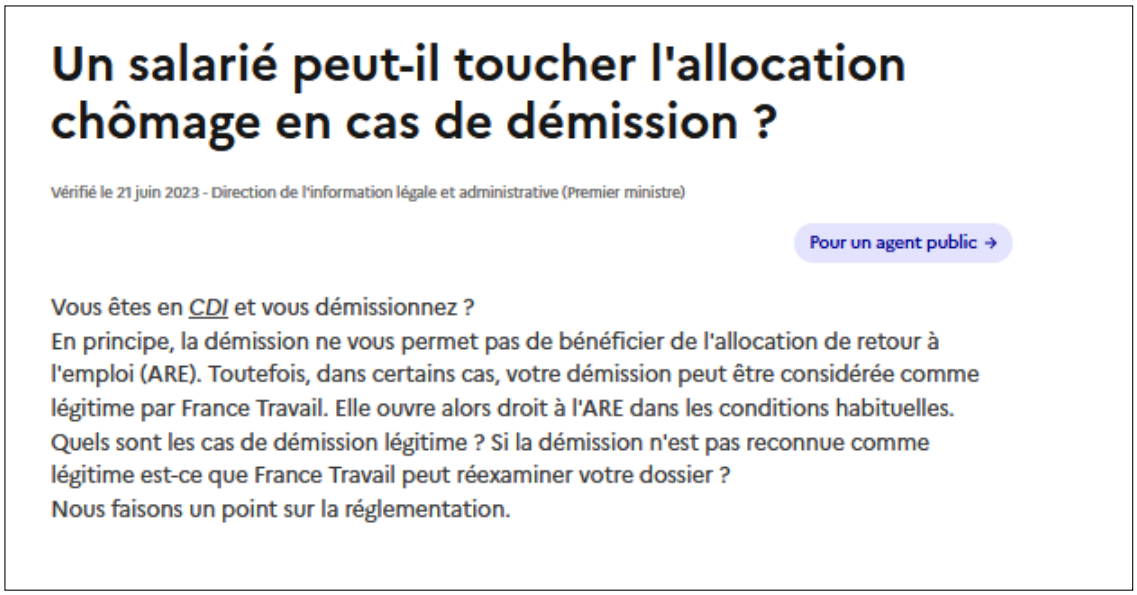


FIGURE 5.2 – Capture d'écran du chapô d'une page du site service-public.fr concernant les droits au chômage des salariés

source :/www.service-public.fr/particuliers/vosdroits/F89, consulté le 1er février 2024

Cette page a été vérié pour la dernière fois le 21 juin 2023, et consulté et récupéré le 1er février 2024 pour notre étude. Nous pouvons observer un paragraphe de présentation de la page, avec plusieurs questions rhétoriques qui indiquent à l'utilisateur les sujets qui seront abordés, par exemple :

> Quels sont les cas de démission légitime ?

Nous avons visité cette page à nouveau en juillet 2024, et nous nous sommes rendu compte que le texte avait été modifié. Le nouveau chapô daté du 4 juillet 2024 est pré-

senté ci-dessous dans la Figure 5.3. À noter que la date de vérification indiquée sur la page (le 21 juin 2023) n'a pas été mise à jour.



FIGURE 5.3 – Capture d'écran du chapô d'une page du site service-public.fr concernant les droits au chômage des salariés

source :/www.service-public.fr/particuliers/vosdroits/F89, consulté le 4 juillet 2024

Les Figures 5.2 et 5.3 ci-dessus, ajoutées à titre d'exemple, nous permettent d'illustrer le caractère dynamique voire instable du discours numérique, qui est constamment mis à jour. Dans cet exemple, la mise à jour consiste en une suppression de deux phrases, les deux avant-dernières questions :

> « Quels sont les cas… peut réexaminer votre dossier ? »

.

On note que la mise à jour n'a pas été datée, car la même date apparaît sur les deux captures, alors même qu'il y a eu modification.

La Figure 5.3 ci-dessus est extraite de la page du site concernant l'allocation chômage des salariés en cas de démission[7]. Le texte montré est uniquement le texte de présentation de la page, c'est-à-dire que d'autres éléments textuels apparaissent sur la page.

En cliquant sur le bouton « Pour un agent public » intégrant une flèche pointant à droite, nous arrivons au texte suivant présenté dans les deux Figures 5.4 et 5.5 ci-dessous. Les deux Figures 5.4 et 5.5 sont des captures d'écran d'une seule page à deux dates différentes (le 1er février 2024 et le 4 juillet 2024). Entre ces deux dates, le texte de cette page a été modifié. La première Figure (5.4) datant du 1er février 2024 est présenté ci-dessous.

7. www.service-public.fr/particuliers/vosdroits/F89



Un agent public peut-il percevoir l'allocation chômage en cas de démission ?

Vérifié le 15 novembre 2022 - Direction de l'information légale et administrative (Premier ministre)

En principe, vous n'avez **pas droit à l'allocation chômage** (allocation de retour à l'emploi - ARE) lorsque vous **démissionnez**.

Toutefois, dans les **3 cas suivants**, vous pouvez **bénéficier d'allocations chômage** :

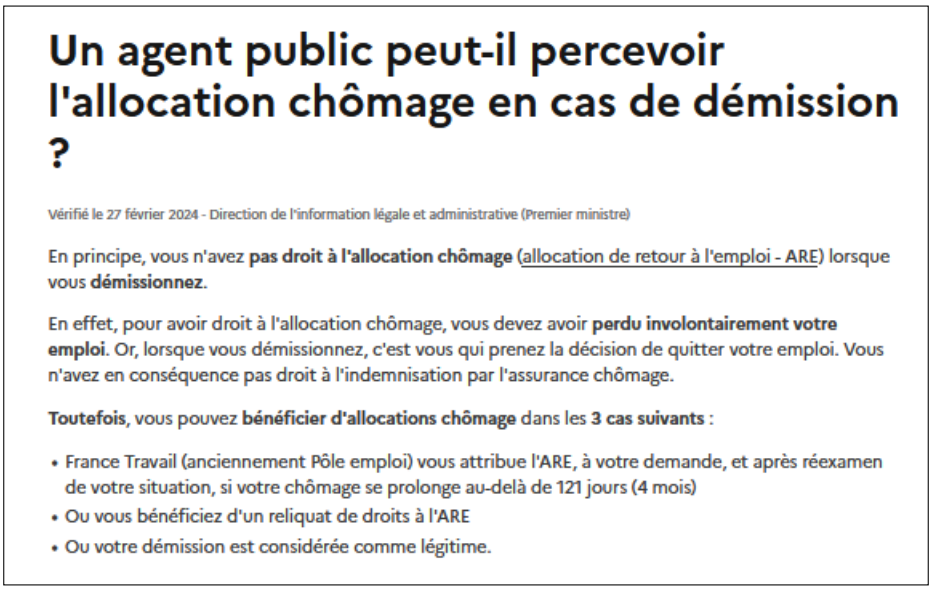
- France Travail (anciennement Pôle emploi) vous attribue l'ARE après le réexamen de votre situation
- Vous bénéficiez d'un reliquat de droits à l'ARE
- Votre démission est considérée comme légitime

FIGURE 5.4 – Capture d'écran du chapô d'une page du site service-public.fr concernant les droits au chômage des agents publics

source :`/www.service-public.fr/particuliers/vosdroits/F35382`, consulté le 1er février 2024

Sur cette page, la date de vérification indiquée est le 15 novembre 2022.

La Figure 5.5 présente le chapô de la page après modification.



Un agent public peut-il percevoir l'allocation chômage en cas de démission ?

Vérifié le 27 février 2024 - Direction de l'information légale et administrative (Premier ministre)

En principe, vous n'avez **pas droit à l'allocation chômage** (allocation de retour à l'emploi - ARE) lorsque vous **démissionnez**.

En effet, pour avoir droit à l'allocation chômage, vous devez avoir **perdu involontairement votre emploi**. Or, lorsque vous démissionnez, c'est vous qui prenez la décision de quitter votre emploi. Vous n'avez en conséquence pas droit à l'indemnisation par l'assurance chômage.

Toutefois, vous pouvez **bénéficier d'allocations chômage** dans les **3 cas suivants** :

- France Travail (anciennement Pôle emploi) vous attribue l'ARE, à votre demande, et après réexamen de votre situation, si votre chômage se prolonge au-delà de 121 jours (4 mois)
- Ou vous bénéficiez d'un reliquat de droits à l'ARE
- Ou votre démission est considérée comme légitime.

FIGURE 5.5 – Capture d'écran du chapô d'une page du site service-public.fr concernant les droits au chômage des agents publics

source :`/www.service-public.fr/particuliers/vosdroits/F35382`, consulté le 4 juillet 2024

Sur cette page, la date de modification est indiquée, il s'agit du 27 février 2024.

En comparant ces deux Figures, nous avons observé l'ajout d'un nouveau paragraphe permettant de préciser les raisons des trois conditions énumérées ouvrant droit à l'ARE

dans la Figure 5.5 :

> En effet, pour avoir droit à l'allocation chômage, vous devez avoir perdu involontairement votre emploi. Or, lorsque vous démissionnez, c'est vous qui prenez la décision de quitter votre emploi. Vous n'avez en conséquence pas droit à l'indemnisation par l'assurance chômage.

La Figure 5.5 ci-dessus est extraite de la page qui concerne les droits au chômage des agents publics en cas de démission. On peut accéder à cette page depuis la page « Un salarié peut-il percevoir l'allocation chômage en cas de démission ? » en cliquant sur le bouton « Pour un agent public » intégrant une flèche pointant à droite. Le texte montré est uniquement le texte de présentation de la page, c'est-à-dire que d'autres éléments textuels apparaissent sur la page mais ne sont pas illustrés dans cette partie. Néanmoins, d'autres éléments textuels typiques du discours numérique seront pris en compte dans l'analyse dans les sections suivantes (cf. 6.2.2, 6.2.3).

Les textes présents dans les Figures 5.4 et 5.2 sont présentés dans la Table suivante et les termes faisant référence aux **allocations chômage** sont soulignés.

page	Page 1	Page 2
Titre de la page	**Un salarié peut-il toucher l'allocation chômage en cas de démission ?**	**Un agent public peut-il percevoir l'allocation chômage en cas de démission ?**
Texte de la page	Vous êtes en CDI et vous démissionnez ? En principe, la démission ne vous permet pas de bénéficier de l'allocation de retour à l'emploi (ARE). Toutefois, dans certains cas, votre démission peut être considérée comme légitime par France Travail. Elle ouvre alors droit à l'ARE dans les conditions habituelles. Quels sont les cas de démission légitime ? Si la démission n'est pas reconnue comme légitime est-ce que France Travail peut réexaminer votre dossier ? Nous faisons un point sur la réglementation.	En principe, vous n'avez **pas droit à l'allocation chômage** (allocation de retour à l'emploi - ARE) lorsque vous **démissionnez**. **Toutefois**, dans les **3 cas suivants**, vous pouvez **bénéficier d'allocations chômage** : — France Travail (anciennement Pôle emploi) vous attribue l'AREaprès le réexamen de votre situation — Vous bénéficiez d'un reliquat de droits à l'ARE — Votre démission est considérée comme légitime

TABLE 5.41 – Textes et titres des pages étudiées

Nous observons dans ces deux textes l'utilisation des variants en contexte. Les variants présents dans ces deux textes sont les mêmes :

1. **Allocation chômage**
2. **Allocation de retour à l'emploi**

3. **ARE**

Les variants 1 et 3 sont parmi les plus fréquemment utilisés dans notre corpus. Leur utilisation sur cette page n'est donc pas surprenante. Toutefois, le variant 2, *allocation de retour à l'emploi* est un variant dit dérivé (voir 5.1.3.1). Cela signifie qu'il s'agit d'une modification opérée par les sites d'accès aux droits, qui n'apparaît pas dans la législation. Leur utilisation sur cette page semble significative, car il n'existe que 12 occurrences de ce variant en corpus, dont 9 sur le site `service-public.fr`. Sur les 9 occurrences du site, 7 concernent des pages qui ne traitent pas directement de l'allocation chômage. Il est donc possible que cette modification soit due à un manque de connaissance du domaine.

L'utilisation de ce variant, ainsi que l'hyperlien présent dans le chapô de la page 1 menant directement à la page 2, indiquent que ces deux pages traitent de sujets similaires. Toutefois, elles semblent répondre à des normes de rédaction très différentes. Tout d'abord, on note une différence de collocation verbale dans les titres. La page 1 utilise la collocation *toucher une allocation*, tandis que la page 2 présente la collocation *percevoir une allocation*. La charte rédactionnelle de la Direction de l'Information Administrative et Légale indique une préférence claire pour le verbe *toucher* par rapport à *percevoir*, qu'il ne faudrait pas ou peu utiliser. Il semble donc que la page 2 ne respecte pas cette charte, ou qu'elle ait été rédigée avant sa mise en œuvre.

La page 1 utilise des questions rhétoriques, qui sont complétées par une explication. La page 2 utilise des marques visuelles avec une liste et des syntagmes en gras. Les deux pages utilisent le *vous* et s'adressent directement à l'usager. Toutefois, la page 1 mélange le *vous* et certaines structures impersonnelles (« Si la démission n'est pas reconnue comme légitime » « la démission ne vous permet pas »). On retrouve donc un mélange entre le vous et l'impersonnel.

L'utilisation des variants est également différente dans les deux pages. Dans la page 1, on retrouve une variation entre le titre (allocation chômage) et le texte (allocation de retour à l'emploi, puis ARE une fois l'acronyme introduit). Cette variation n'apparaît pas sous la même forme dans la page 2. En effet, la page 2 propose une juxtaposition des trois variants utilisés comme suit :

(49) vous n'avez pas droit à l'allocation chômage (allocation de retour à l'emploi — ARE).

Cette juxtaposition permet de désambiguïser l'utilisation des variants, qui sont indiqués comme étant concurrents. Dans la suite du texte, on retrouve l'utilisation des différents variants :

(50) (...) vous pouvez bénéficier d'allocations chômage.

(51) France Travail (anciennement Pôle emploi) vous attribue l'ARE après le réexamen de votre situation

Les stratégies employées dans les deux chapôs sont donc différentes, tant au niveau de la variation que du style. Il est intéressant de noter que les textes ont été publiés à deux dates différentes. La page 1 a été vérifiée pour la dernière fois le 21 juin 2023, tandis que la page 2 a été vérifiée pour la dernière fois le 15 novembre 2022. La mise à jour effectuée le 27 février 2024 n'a pas porté sur l'harmonisation des stratégies rédactionnelles. Les différentes dates peuvent justifier des différentes stratégies utilisées, néanmoins les mises à jour auraient pu permettre de les harmoniser. Par ailleurs, la page 1 propose dans son chapô un lien vers la page 2 (voir Figure 5.2). Ce lien montre que les deux pages sont reconnues comme abordant des sujets complémentaires ou similaires. La page 2 propose un lien similaire redirigeant vers la page 1, mais celui-ci se trouve en bas de page, dans le *contenu complémentaire*.

Nos analyses portant sur la terminologie française montrent une instabilité et une cohabitation des différents variants sur les pages.

5.1.3.2.2 Analyse en texte britannique

Nous allons réaliser une étude similaire afin d'identifier si cette instabilité est présente sur les pages britanniques. Nous illustrons notre propos avec deux pages concernant deux dispositifs. La première page est présentée dans la Figure 5.6. Il s'agit ici uniquement du chapô de la page.

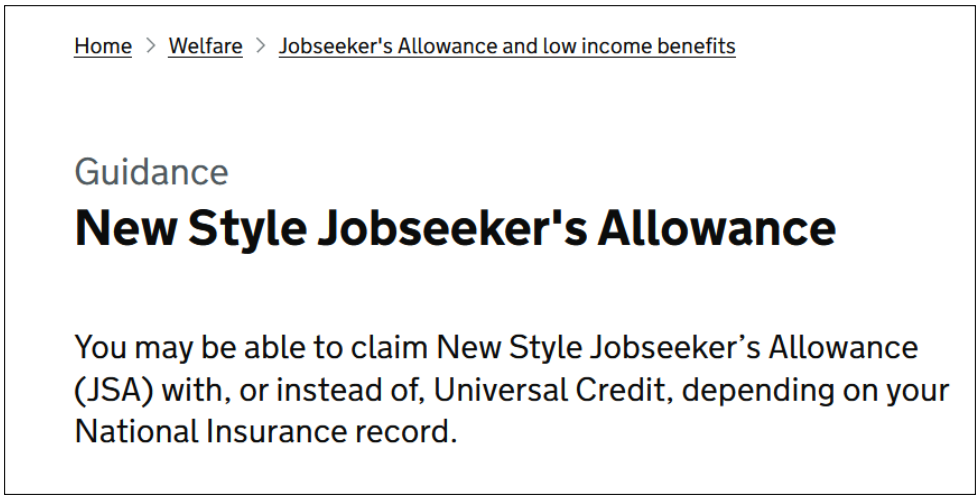


FIGURE 5.6 – Capture d'écran du chapô de la page concernant le *New Style Jobseeker's Allowance*

source : `www.gov.uk/guidance/new-style-jobseekers-allowance`, consulté le 20 mai 2024

Dans la présentation de cette page, nous retrouvons trois termes qui nous ont intéressée lors de l'analyse de ce concept :

— *New Style Jobseeker's Allowance*
— *JSA*
— *Universal Credit*

L'ensemble de la page présente les différences et le fonctionnement des deux dispositifs différents[8]. Plus loin dans le texte de la page, deux termes faisant référence aux dispositifs viennent s'ajouter. Tout d'abord *New Style JSA*, qui devient l'unique terme pour faire référence au dispositif après avoir été introduit dans le premier paragraphe du texte.

La Figure 5.7 présente une autre section de la page.

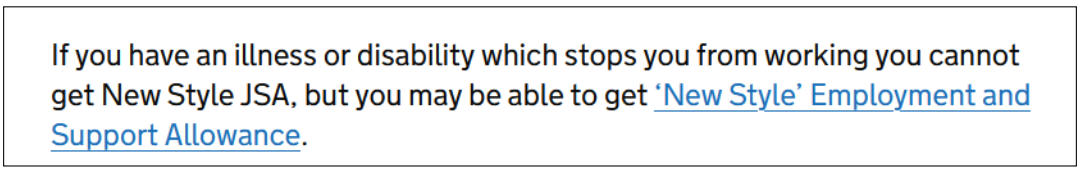


FIGURE 5.7 – Capture d'écran d'un paragraphe de la page concernant le *New Style Jobseeker's Allowance*

source : `www.gov.uk/guidance/new-style-jobseekers-allowance`, consulté le 20 mai 2024

Dans cette section, le dispositif *New Style JSA* est comparé à un autre dispositif similaire *'New Style' Employment and Support Allowance*. Dans ce paragraphe, le rédacteur introduit un nouveau terme, qui est doté d'un hyperlien afin que l'usager puisse consulter la page correspondante. L'usage d'un hyperlien permet de simplifier l'introduction d'un nouveau terme. Toutefois, nous pouvons noter une instabilité graphique, car dans ce dispositif, les guillemets sont utilisés dans *'New Style'*, tandis qu'ils n'apparaissent pas dans *New Style JSA*.

Pour confirmer les observations réalisées précédemment, nous avons fait une analyse similaire sur la page concernant le *Employment and Support Allowance*, dont le chapô est présenté dans la Figure 5.8 ci-dessous.

8. *Universal Credit* et *New Style Jobseeker's Allowance*

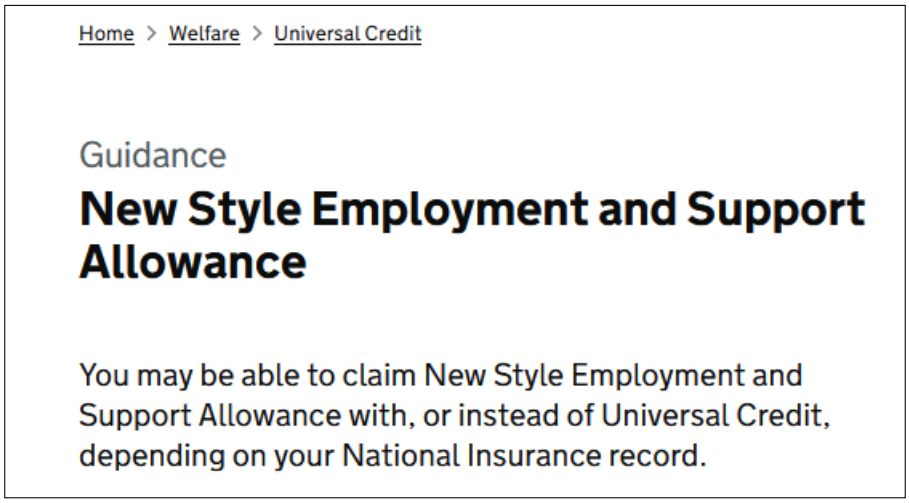


FIGURE 5.8 – Capture d'écran d'un paragraphe de la page concernant le *Employment and Support Allowance*, consulté le 20 mai 2024

`www.gov.uk\/guidance\/new-style-employment-and-support-allowance`

La page propose un texte de présentation presque identique au texte de la page *New Style JSA*. De la même façon, dans la suite du texte (Figure 5.9 ci-dessous), une fois le terme complet introduit ainsi que son sigle, le rédacteur n'utilise plus que la version raccourcie et spécialisée du terme *New Style ESA*. La Figure 5.9 ci-dessous présente une autre section de la page.

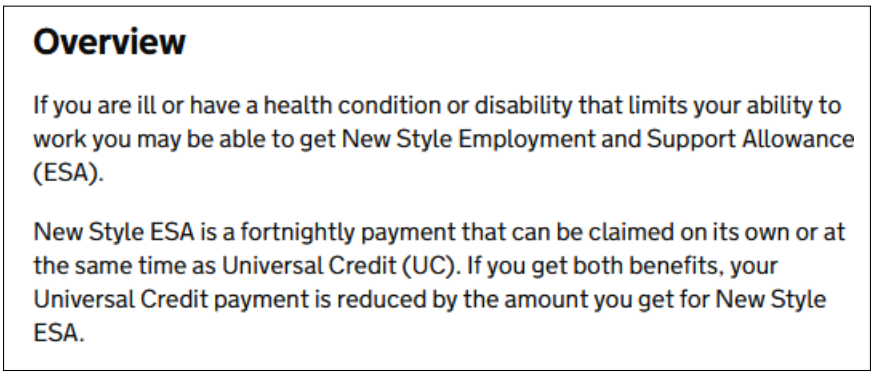


FIGURE 5.9 – Capture d'écran d'un paragraphe de la page concernant le *Employment and Support Allowance*

source : `https://www.gov.uk/employment-support-allowance`, consulté le 25 juin 2024

Le texte ci-dessus (Figure 5.9) montre l'évolution de *New Style Employment and Support Allowance (ESA)* à *New Style ESA*. Nous pouvons également noter une différence avec la page précédente : le dispositif *Universal Credit* est mentionné avec son sigle *UC*, qui est ensuite utilisé dans le reste de la page. Dans la page précédente (Figure

5.7), le sigle n'était pas utilisé. Il est possible de s'interroger sur cette différence entre les deux pages, qui prennent des décisions terminologiques divergentes. Dans les deux cas, la terminologie est profuse, mais explicitée.

L'étude de ces deux pages n'a pas mené à l'identification d'une instabilité similaire aux sites français. La terminologie du domaine, si elle est complexe, n'est pas présentée de façon instable dans l'environnement numérique. En effet, la variation terminologique est mise en valeur de façon explicite pour éviter l'effet d'instabilité. Les résultats présentés dans la section 5.1.1.2 de ce chapitre ont tout de même montré une variation au niveau du noyau des dispositifs, que nous allons maintenant aborder.

5.1.3.3 Variation des noyaux des termes complexes : aide, allocation, complément, prestation, prime, bourse/*benefit*, *credit*, *allowance*, *payment*

Nous allons maintenant nous intéresser à la variation au niveau du noyau des groupes nominaux complexes, c'est-à-dire des termes complexes. Comme nous l'avons présenté précédemment, les termes principaux du domaine désignant des dispositifs d'aide mis en place pour les usagers donnent lieu à des termes complexes. Ces dispositifs peuvent concerner plusieurs sujets : droit au séjour, aide sociale, etc.

Nous avons identifié, dans la section 5.1.1.1, un phénomène de variation du noyau au niveau des dispositifs d'aide sociale en français et en anglais.

Tout d'abord, il est important de noter que les termes sont tous formés de façon similaire. Chaque terme est un groupe nominal complexe, composé d'un noyau qui désigne le concept d'aide (allocation, aide, prestation, etc.) et d'une extension qui se présente généralement sous la forme d'un complément du nom. Ce complément est parfois introduit par une préposition, puis suivi d'un second groupe nominal. Il sert à préciser l'objet de l'aide (soutenir la famille, accueillir un enfant, payer un loyer), si bien qu'il est « descriptif ».

[noyau] [préposition] [complément1] [complément2] [complément3] [...]

Si nous analysons le groupe nominal **allocation de soutien familial**, on note que le noyau est **allocation**. Il nous indique que nous avons affaire à un dispositif d'aide financière. Ce noyau est suivi du complément **soutien familial**, qui apporte une information supplémentaire sur le but de l'aide : il s'agit d'une aide financière pour soutenir les familles. Toutefois, le terme reste opaque, car il n'est pas possible de déduire les conditions d'attribution de l'aide, le complément servant seulement à déduire le domaine d'application. Pour **allocation de soutien familial**, il s'agit de la famille. Nous avons présenté dans la partie 5.1.1.1 certains termes français :

— **allocation de soutien familial (Asf)**

— **prestation d'accueil du jeune enfant (Paje)**
— **prime à la naissance (Pn)**
— **allocation de base (Ab)**
— **aide personnalisée au logement (Apl)**

Ces termes font tous référence à un dispositif d'aide financière. Pour chacun, il s'agit d'une aide versée au bénéficiaire selon certains critères. Chacun de ces termes dispose d'un noyau différent, mais ces noyaux ne font pas toujours état d'une différence conceptuelle. Le noyau **prime** fait parfois état d'un versement unique, dans le cas de **prime à la naissance** par exemple. Dans d'autres cas, il est utilisé pour désigner des aides versées mensuellement, par exemple la **prime d'activité** qui est une aide versée mensuellement par la caisse d'allocations familiales (Caf).

Les noms noyaux identifiés en français sont les suivants :
— aide
— allocation
— complément
— prestation

Il s'agit donc d'un phénomène de variation qu'il convient d'étudier. Cette variation ne semble pas toujours être fonctionnelle, car il est tout à fait possible d'échanger les noyaux des différents termes. C'est le cas pour certaines aides qui sont désignées principalement par leurs acronymes et sigles, et dont la désignation complète n'est donc pas stable. Dans l'exemple suivant, le noyau **aide**, utilisé normalement dans **aide personnalisée au logement** est remplacé par **allocation**.

(52) Locataire d'un studio du parc social, elle a à sa charge un loyer mensuel de 190 €, déduction faite de l'allocation personnalisée au logement (APL) versée directement à son bailleur.

Dans l'exemple suivant, le noyau **aide** est substitué au noyau **allocation**.

(53) La rémunération est imposable comme l'aide au retour à l'emploi (ARE).

Ces changements, même s'ils sont peu fréquents, montrent qu'il s'agit bien d'un même concept, désigné par différents variants. Cette variation apparaît lors de la conception du dispositif. La synonymie pourrait ainsi être démultipliée par des « actions d'aménagement linguistique » (Humbley, 2018, p. 52). Dans le cas de notre étude, les actions sont les multiples dispositifs mis en place par la législation et leurs dénominations. L'introduction incrémentale de nouvelles dénominations pour des dispositifs crée de nouveaux variants. En effet, même si juridiquement le terme précédent devient rapidement obsolète et disparaît des textes juridiques, le transfert dans l'usage et dans d'autres types de discours est plus lent, et le nouveau variant devient concurrent des précédents.

La variation au niveau du noyau est aussi présente dans le discours de l'accès aux droits britannique. Nous l'avons notamment étudié dans la section 5.1.1.2. Dans cette partie, nous allons illustrer cette variation en analysant un dispositif en particulier, *Universal Credit*, qui présente ce phénomène de variation au niveau du noyau. Nous avons identifié quatre composantes dans le dispositif *Universal Credit* :

— *standard allowance*
— *work allowance*
— *hardship payment*
— *housing payment*

Nous avons donc identifié trois noms noyaux au sein d'un seul dispositif :

— credit
— allowance
— payment

Ces noms noyaux sont parfois concurrents, comme dans *standard allowance* et *housing payment*, et peuvent faire référence à des versements réguliers. Mais ils prennent parfois un sens différent comme dans *hardship payment*, où il s'agit d'une somme unique.

Le nom noyau *credit* peut également porter à confusion, car il est polysémique. Dans le terme *Tax Credit*, il ne fait pas référence à une aide financière au sens d'un versement, mais simplement à une réduction d'impôt. Dans le cas du dispositif *Universal Credit* il s'agit bien d'une aide financière versée régulièrement à l'usager.

Un quatrième nom noyau est présent dans notre corpus. Il s'agit de *benefit*, qui est en même temps un terme générique et qui est aussi intégré aux dénominations de dispositifs d'aide, comme dans les aides suivantes :

— *Child Benefit*
— *Incapacity Benefit*
— *Housing Benefit*

Il est toutefois intéressant de noter qu'il s'agit d'un nom noyau moins utilisé que les trois autres noms noyaux (*credit*, *allowance*, *payment*), probablement pour éviter la confusion entre l'holonyme *Universal Credit* et ses méronymes (*standard allowance*, *housing payment*, *work allowance*, *harship payment*).

Dans le cas de *Housing Benefit*, les majuscules et le singulier indiquent qu'il s'agit d'une aide spécifique, mais nous retrouvons également des occurrences de *housing benefits* en corpus, qui font référence à des aides liées au logement en général.

En conclusion, le corpus britannique présente une variation au niveau des noms noyaux. Cette variation peut être particulièrement déroutante, car les noms noyaux choisis ont souvent plusieurs sens en corpus qui se croisent même dans les noms de dis-

positifs. L'introduction d'une nouvelle aide supposée « universelle » (*Universal Credit*) présente les mêmes caractéristiques au niveau de la variation des noms noyaux. Il est toutefois intéressant de noter que cette fois, le nom noyau *benefit*, qui peut être particulièrement complexe car il est en même temps un hyperonyme et un hyponyme, n'apparaît pas dans les différentes composantes de ce nouveau dispositif. Toutefois, une variation au niveau du noyau persiste pour les méronymes du terme *Universal Credit* ainsi que d'autres dispositifs d'aide spécifiques n'entrant pas dans *UC*.

Une autre variation au niveau du noyau est la variation entre **carte de séjour** et **titre de séjour** présente dans le sous-corpus du droit au séjour français.

Il existe trois manières de désigner le document qui permet à un citoyen étranger de résider en France sur le long terme :

1. **titre de séjour**
2. **carte de séjour**
3. **certificat de résidence**

En ce qui concerne l'utilisation des deux termes **titre de séjour** et **carte de séjour**, ils peuvent parfois désigner des concepts distincts, mais sont également utilisés comme synonymes dans certains contextes, comme souligné dans la section 5.1.1.4. Les exemples ci-dessous illustrent ce phénomène.

(54) Si vous ne possédez pas de carte de séjour ou si votre carte est périmée, un document provisoire de séjour vous est remis.

(55) Si l'administration n'a pas répondu dans un délai de 4 mois, votre demande de carte est refusée.

(56) Sous réserve d'éligibilité, il se verra remettre un titre de séjour « carte de résident » d'une durée de validité de dix ans, renouvelable.

(57) Les conditions sont différentes en fonction du titre de séjour ou du visa que vous détenez.

La variation observée pourrait venir en opposition au figement du discours juridique (Rey, 2008). En effet, les rédacteurs peuvent chercher à proposer une diversité terminologique plus grande afin de faciliter la compréhension et de faire disparaître ce figement issu du juridique.

Dans le corpus britannique, nous n'avons pas observé de variation autour de la terminologie du droit au séjour. Au contraire, nous observons une stabilité et une fidélité au discours juridique avec l'utilisation du terme *visa* pour l'ensemble des dispositifs et la description du statut de l'individu comme nous l'avons décrit dans la section 5.1.1.5.

5.1.4 Acronymes et sigles : ARE, Paje, PrePaRe ; *JSA*, *UC*, *DLA*, etc.

Dans chacune des thématiques choisies pour notre corpus, nous avons identifié l'utilisation d'acronymes ou de sigles. Rappelons la différence entre un sigle et un acronyme : bien que les deux soient formés de lettres initiales, l'acronyme est prononcé comme un mot ordinaire, alors que le sigle est épelé. Les sigles et acronymes sont utilisés sur l'ensemble des thématiques de notre corpus. Ils constituent une pratique discursive en langue de spécialité (voir 3.3.4) qui peut être lié à un phénomène de néologie. Les groupes nominaux complexes présentent des inconvénients dans la communication en la rendant moins efficace, et il peut être utile de les réduire à l'aide d'acronymes ou de sigles. On observe alors une spécialisation du discours, qui devient plus efficace, mais ambigu pour les usagers non initiés. De plus, le système dénominatif et conceptuel étant complexe dans le cas des aides sociales et du droit au séjour en particulier, les acronymes sont tout autant complexes. Cela peut mener à des écarts et même à un glissement sémantique qui font apparaître des variations dérivées des termes, comme nous l'avons vu avec *income-related JSA* ou **allocation de retour à l'emploi**.

Pour étudier les acronymes, nous avons relevé les occurrences de 3 à 4 mots en majuscules en corpus grâce à l'expression régulière suivante : [A-Z]3,4

Cela signifie que l'on va chercher des suites de 3 à 4 lettres en majuscules. Les relevés pour les acronymes des termes de 3 et 4 mots réalisés sur le corpus français complet (les trois thématiques incluses) sont présentés dans la Table 5.42 ci-dessous.

word	Frequency	Freq. per million	% of concordance
MDPH	1289	599,6152	9,888
MSA	1065	495,4152	8,16968
RSA	844	392,61073	6,47438
AAH	721	335,39376	5,53084
PCH	614	285,61966	4,71003
CMI	590	274,45537	4,52593
ARE	474	220,49465	3,63608
CASVP	303	140,94911	2,32433
OQTF	261	121,41161	2,00215
FAP	260	120,94643	1,99448
CNSA	203	94,43125	1,55723
BTP	198	92,10536	1,51887
CDI	185	86,05804	1,41915
EEE	176	81,87143	1,35011
DCEM	142	66,05536	1,08929
ICT	120	55,82143	0,92053
PDF	112	52,1	0,85916
CMG	108	50,23929	0,82847

TABLE 5.42 – Résultats de l'expression régulière [A-Z]3,4 dans le corpus français

Lors de notre exploration du corpus dans la section 5.1.1, nous avons identifié certains acronymes ou sigles qui ne sont plus en majuscule et ne prennent qu'une majuscule au départ :

— Aah,
— Are,
— Apl,
— Als,
— Alf,
— PreParE

Certains de ces acronymes ou sigles deviennent plus utilisés que leur dénomination légale et sont donc des termes à part entière. C'est le cas notamment de l'**Are**, qui est un acronyme et se prononce donc comme un mot, et de l'**Apl**, qui est largement plus utilisé que le terme complet correspondant. Certains acronymes ne sont pas nés de la simple réduction des groupes nominaux complexes, c'est-à-dire de façon habituelle, mais ont été choisis, pensés et préparés pour être utilisés en tant qu'acronyme. C'est notamment le cas de la **PreParE** (pour **prestation partagée d'éducation de l'enfant**), dont on a sélectionné plusieurs lettres au début de chaque mot pour former un acronyme. En 2004, l'**allocation Pour Jeune Enfant**, dite **APJE** est remplacée par la **prestation d'accueil du jeune enfant** dite **Paje**. Il semble que cette décision terminologique soit également liée à l'utilisation de l'acronyme, car nous assistons simplement à une redistribution des lettres afin de prononcer facilement l'acronyme. Dans ce cas, les noms noyaux **prestation** et **allocation** semblent complètement interchangeables.

Prenons l'exemple de l'acronyme **Apl**, formé à partir du terme **aide personnalisée au logement**. Il s'agit d'un terme qui fait référence à un certain type d'aide au logement, aux côtés de deux autres aides : **allocation de logement à caractère social** et **allocation de logement à caractère familial**, respectivement **Als** et **Alf**. Pour l'ensemble de ces dispositifs, les acronymes permettent de communiquer de façon plus efficace entre experts, mais aussi de faire des phrases moins compliquées et plus visuelles, ce qui peut être un facteur important sur les sites web. Les trois dispositifs d'aide au logement deviennent alors l'**Apl**, l'**Als** et l'**Alf**. Le sens initial est non seulement perdu, mais il est plus difficile de distinguer les différentes aides.

Le recours à des acronymes au détriment des sigles serait donc une tendance observable dans le choix des dénominations pour les dispositifs d'aide en français.

Nous avons ensuite identifié des sigles dans le corpus britannique. Nous n'avons pour l'instant pas identifié d'acronyme. L'expression régulière présentée au début de cette section a été utilisée pour identifier les sigles présents dans le corpus. Ils sont présentés dans la Table 5.43 ci-dessous.

Acronyme	Occurrence	%
DWP,	1194,	18.47725
ESA,	893,	13.81925
PDF,	821,	12.70504
JSA,	806,	12.47292
DLA,	302,	4.67348
SNP,	270,	4.17827
SMI,	228,	3.52832
NHS,	199,	3.07954
BSL,	186,	2.87837
BRP,	132,	2.04271

TABLE 5.43 – Résultats pour l'expression régulière [A-Z]3,4 dans le corpus anglais

Les résultats présentés dans la Table 5.43 ne sont en grande partie pas des dispositifs liés à l'accès aux droits. *DWP* est un département exécutif du gouvernement britannique (*Department for Work and Pension*). *PDF* est le format de document informatique (*Portable Document Format*). *SNP* est un parti politique écossais (*Scottish National Party*). *NHS* est l'équivalent de la Sécurité sociale (*National Health Service*), et enfin le *BSL* désigne le langage des signes (*British sign language*).

Toutefois, on retrouve trois dispositifs issus de la thématique aide sociale formés de la même manière : *ESA*, *JSA* et *DLA* qui correspondent aux trois aides suivantes :

— *Employment and Support Allowance (ESA)*

— *Jobseeker's Allowance (JSA)*

— *Disability Living Allowance (DLA)*

Dans les trois cas, le sigle est plus utilisé que le terme complet, parfois en montrant des fréquences d'utilisation significativement plus importantes, comme *ESA* qui compte 893 occurrences, contre 263 occurrences pour le terme complet. *Jobseeker's Allowance* compte 414 occurrences, tandis que *JSA* compte 806 occurrences. *Disability Living Allowance* compte 189 occurrences et *DLA* compte 302 occurrences. Les sigles sont donc utilisés bien plus fréquemment que les termes complets. Comme nous l'avons vu dans la section 5.1.3.2, le terme complet est généralement utilisé dans le titre de la page, dans sa présentation, et une première fois dans le texte, puis il est remplacé par le sigle. Cela permet une communication plus efficace. Toutefois, il est important d'envisager que le discours numérique de l'accès aux droits n'est pas consulté de manière toujours linéaire et que des liens permettent d'accéder à différentes sections de la page. Un usager ne consultera pas forcément la présentation ou le début du texte. Dans ce cas-là, l'utilisation directe du sigle peut poser problème, d'autant plus que le sigle n'apparaît pas dans les titres.

Un dernier dispositif apparaît dans cette liste, il s'agit du *Support for Mortgage Interest* qui compte 46 occurrences, et dont le sigle est le suivant : *SMI*. Ce terme n'a pas été relevé lors des analyses précédentes car il n'a pas de nom noyau. Il désigne néanmoins un type d'aide particulier. Cet exemple montre la nécessité de multiplier les interrogations de corpus pour aboutir à une analyse terminologique complète du discours administratif.

Nos explorations précédentes, notamment dans la section 5.1.1.2, ont mis en avant l'utilisation de sigles pour certains dispositifs disposant de deux lettres uniquement. Nous avons donc exploré l'ensemble des sigles formés avec deux lettres.

word	Frequency	Freq. per million	% of concordance
UK	3596	3684,46021	69,46108
KB	749	767,42511	14,46784
EU	368	377,05266	7,10836
MB	86	88,11557	1,66119
GP	38	38,93479	0,73402
HR	37	37,91019	0,7147
UC	32	32,78719	0,61812
TV	29	29,71339	0,56017

TABLE 5.44 – Résultats pour l'expression régulière [A-Z]2 dans le corpus anglais

Une grande partie des sigles présentés dans la Table 5.44 ne sont pas spécifiquement issus du discours de l'accès aux droits. L'acronyme le plus utilisé, *UK*, est naturellement présent pour signifier qu'il s'agit de l'accès aux droits britanniques. Le deuxième sigle mis en avant est *KB*, qui correspond à une unité de taille d'un fichier. Comme avec *PDF*, on retrouve des sigles liés au caractère numérique du discours. Le seul sigle formé par deux lettres correspondant à un dispositif est *UC*, pour *Universal Credit*. Il semble alors que l'utilisation des sigles ne soit pas une pratique régulière pour les termes composés, étant donné que les autres dispositifs formés par des termes NN n'ont pas tendance à être réduits en sigle.

En conclusion, les sigles et acronymes sont très présents en corpus. Ils reflètent en même temps une volonté de rendre plus efficaces les échanges, et de les spécialiser. En effet, en perdant le terme initial et en introduisant un nouveau variant, le discours de l'accès aux droits fait émerger une terminologie spécifique au domaine, dans la mesure où les sigles sont les variants opaques de leur variant développé correspondant. Cette nouvelle terminologie n'est pas transparente pour l'usager, qui doit connaître la signification de ces sigles et leurs différents rôles dans le discours. Les sigles et acronymes, même s'ils peuvent sembler être des outils de simplification, sont en réalité un signe de spécialisation du discours de l'accès aux droits numérique. Enfin, l'exemple de **PreParE** (pour **prestation partagée d'éducation de l'enfant**) illustre non seulement la tendance actuelle à rendre les sigles prononçables en les substituant par les acronymes (cf. **Paje**)

mais pourrait être accompagné d'une volonté de les rendre également moins opaques pour les usagers.

5.1.5 Approche ontologique de la terminologie du discours numérique de l'accès aux droits

Dans cette partie, nous proposons d'utiliser une approche ontologique de la terminologie du domaine, en essayant de comprendre et de représenter le système conceptuel de certaines parties du domaine. Chaque domaine possède des dispositifs spécifiques, et chaque dispositif est désigné par un terme, ou par un terme et ses variants. Nous avons donc réalisé une représentation ontologique autour du concept de document de séjour en France et au Royaume-Uni. Nous avons identifié dans la section 5.1.1.4 que ce système conceptuel était complexe. Nous avons également identifié une variation autour des concepts de **titre de séjour** et **carte de séjour**. Dans notre représentation, nous avons identifié **titre de séjour** comme étant un hyperonyme et **carte de séjour** comme étant un nom noyau des différents types de cartes de séjour. Cette décision reflète l'usage des termes en corpus : **carte de séjour** est plus fréquemment utilisé comme nom noyau pour désigner divers types de **titres de séjour**. De plus, cette utilisation est en accord avec la définition de **titre de séjour** donnée dans le glossaire de la Direction des Étrangers en France, disponible sur leur site [9]

Nous proposons donc dans la Figure 5.10 ci-dessous une représentation ontologique de la terminologie du domaine restreinte au concept de document de séjour.

9. immigration.interieur.gouv.fr/Info-ressources/Etudes-et-statistiques/Glossaire/Titre-de-sejour

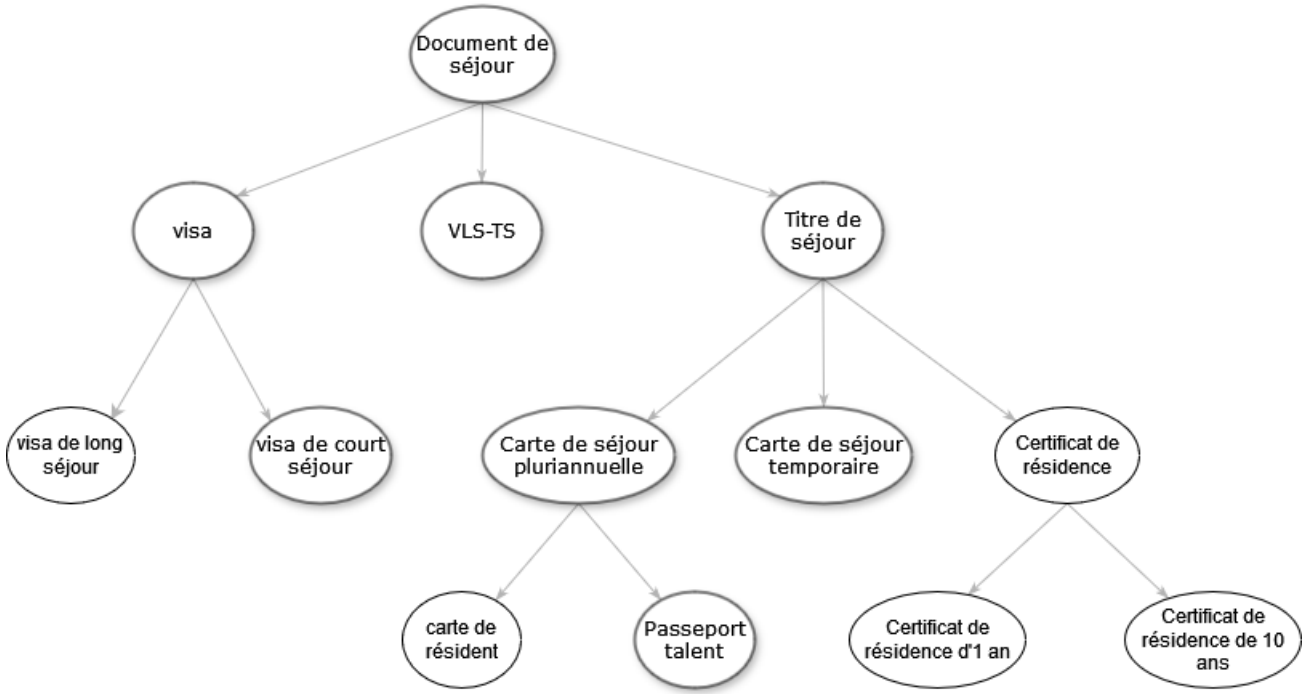


FIGURE 5.10 – Représentation ontologique des termes vedettes autour du concept de document de séjour en France

L'ontologie représentée dans la Figure 5.10 se concentre sur les différentes dénominations des documents de séjour, en excluant les mentions des titres de séjour. Le terme **document de séjour** est le terme principal, englobant l'ensemble des hyponymes. Le contexte riche en connaissance suivant nous a permis de déterminer le début de notre représentation ontologique :

(58) En France Vous devez déposer votre demande sur internet, au plus tôt 4 mois et au plus tard 2 mois avant la fin de validité de votre document de séjour (visa, VLS-TS ou titre). (service-public.fr)

L'extrait souligné montre une organisation des termes que nous avons reportée dans la Figure 5.10. D'après cette phrase, il existe donc trois types de documents : les visas, les titres (de séjour) et les VLS-TS (visa long séjour valant titre de séjour). Nous avons ensuite étudié les différents types de titres de séjour. Les ontologies ont pour but de présenter un terme unique pour chacun des concepts. Notre approche par corpus étant inductive et non prescriptive : lorsque des variants existaient, nous avons choisi les plus fréquents. C'est le cas par exemple pour la **carte de séjour temporaire** et la **carte de séjour pluriannuelle**. Ces termes ont été identifiés grâce aux analyses réalisées sur Sketch Engine et présentées dans la section 5.1.1.4. Nous avons également identifié deux types de **cartes de séjour pluriannuelles** : la **carte de résident** et le **passeport talent**. On remarque ici une instabilité au niveau du nom noyau entre **passeport** et **carte**.

Pour aller plus loin, nous proposons une version enrichie de cette ontologie avec l'ensemble des termes vedettes et des variants présents dans le corpus. Pour les besoins de notre étude et faciliter l'analyse, nous présentons ce travail en différentes Figures qui détaillent chaque partie de l'arborescence. Toutefois, dans une perspective de vulgarisation il est tout à fait possible, et même recommandé, que cette arborescence soit présentée en une seule arborescence sur un grand support.

La première Figure 5.11 ci-dessous rassemble l'ensemble des variants autour de la branche visa.

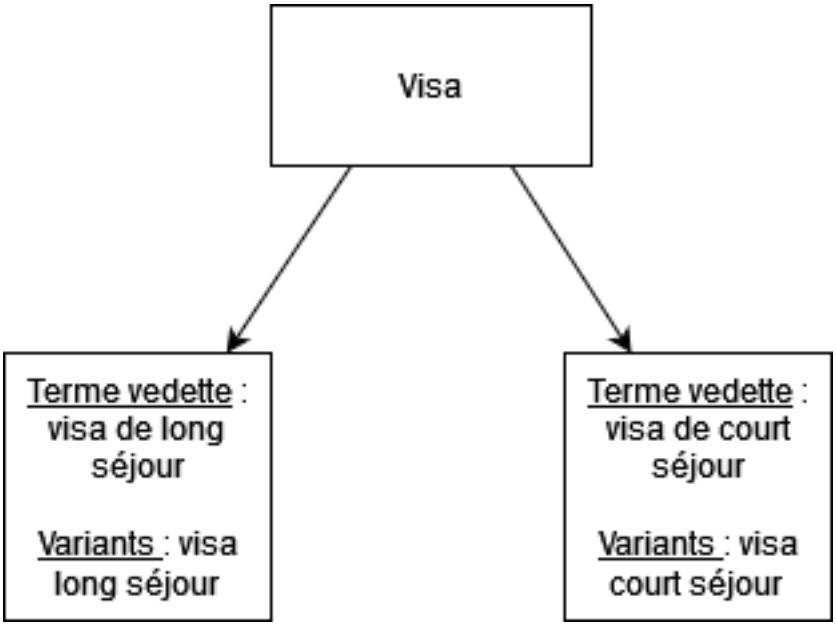


FIGURE 5.11 – Terme vedette et variants dans la branche [visa]

Dans cette branche, il n'existe qu'un variant de chaque terme. Le terme vedette est le terme juridique, tandis que les variants sont des termes dérivés, modifiés et raccourcis.

La Figure 5.12 présente les termes vedettes et variants autour de la branche VLS-TS.



FIGURE 5.12 – Terme vedette et variants dans la branche [VLS-TS]

Dans la Figure 5.12, on peut voir qu'il existe deux variants du terme vedette. Il s'agit en fait des termes complets. On retrouve deux variants, car ce terme est construit à partir du terme **visa de long séjour**. Le premier variant, **visa de long séjour valant titre de séjour**, est le terme juridique. Le second variant est tout simplement formé à partir du variant **visa long séjour**, présenté dans la Figure 5.11.

La Figure suivante présente les termes vedettes et variants autour de la branche titre de séjour.

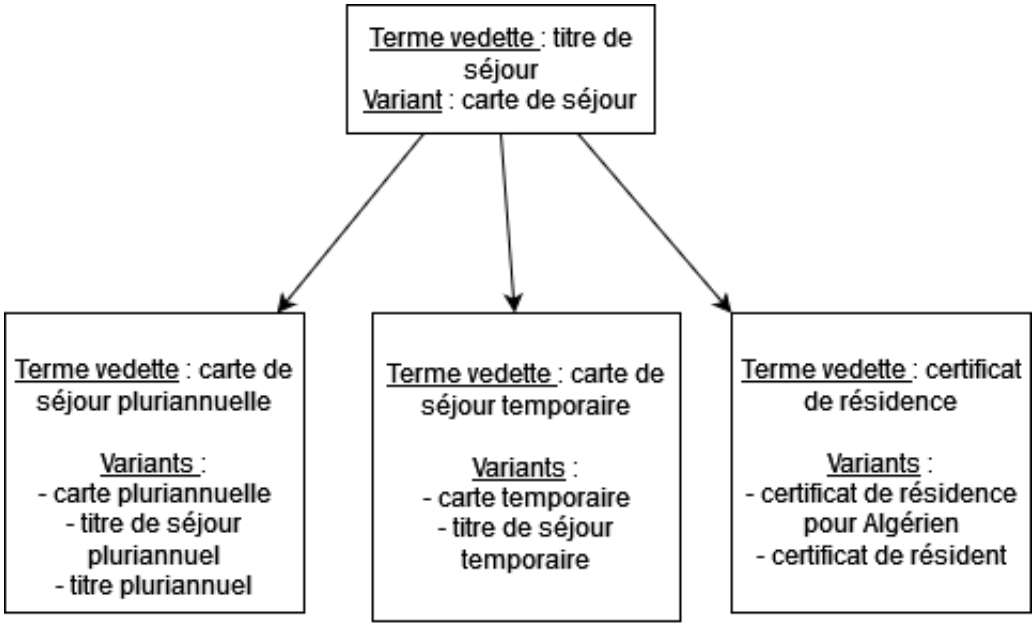


FIGURE 5.13 – Terme vedette et variants dans la branche [titre de séjour]

Dans la Figure 5.13, le terme initial de la branche **titre de séjour** possède un variant qui est **carte de séjour**. Nous avons déjà mis en évidence cette variation dans les sections 5.1.1.5 et 5.1.3. Pour **carte de séjour pluriannuelle** et **carte de séjour temporaire**, il s'agit de variants liés à deux phénomènes. Tout d'abord, l'instabilité au niveau du noyau (entre **titre** et **carte**) est répercutée dans les hyperonymes et on retrouve donc des occurrences de **titre de séjour pluriannuel** ou **titre de séjour temporaire**. Dans un second temps, on observe un phénomène de simplification des termes ou d'économie du langage qui a pour résultat de faire disparaître le complément **de séjour**. Le locuteur ne garde alors que les éléments les plus significatifs, afin de communiquer de façon plus efficace. Il s'agit d'un phénomène commun en langue de spécialité qui, ici, peut être ambivalent. D'une part, cette omission allège le texte et le rend plus facile à lire. D'autre part, il omet une information, et peut empêcher la compréhension du terme. Pour le terme **certificat de résidence**, le premier variant présenté propose au contraire des précisions sur le terme en indiquant les usagers concernés (**pour Algérien**). Le second variant **certificat de résident** est créé par dérivation entre le nom **résidence** et **résident**.

La Figure 5.14 présente les termes vedettes et variants autour de la branche carte de séjour pluriannuelle.

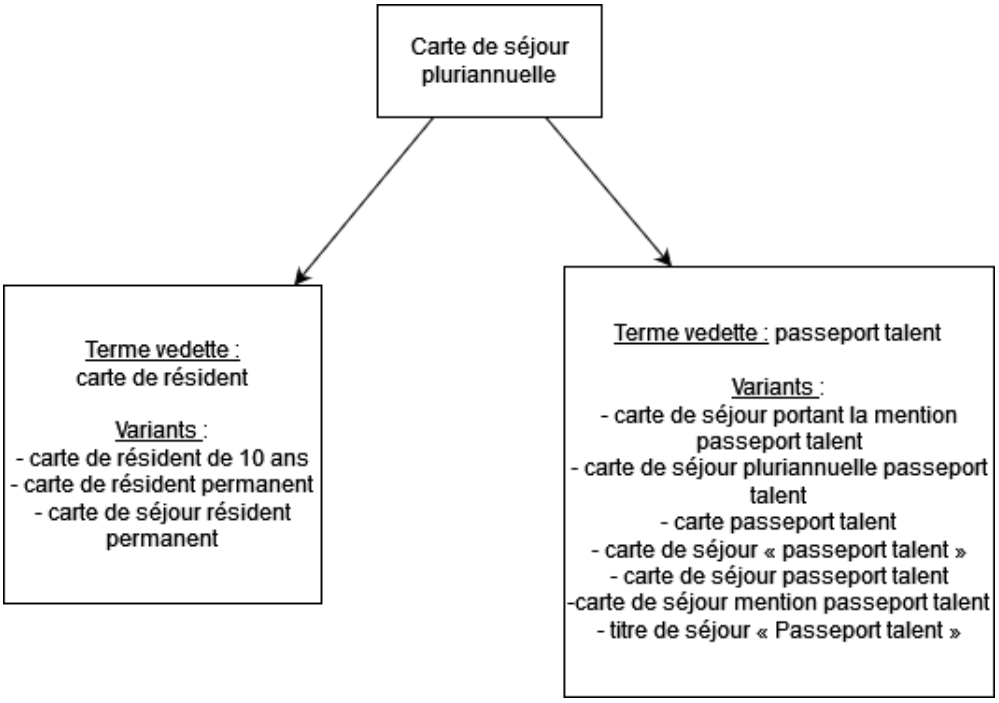


FIGURE 5.14 – Terme vedette et variants dans la branche [carte de séjour pluriannuelle]

Dans la Figure 5.14, nous proposons la liste des variants pour deux termes, **carte de résident** et **passeport talent**, qui sont hyperonymes de **carte de séjour pluriannuelle**. Le premier terme vedette (**carte de résident**) est déjà dans sa forme réduite. Ses variants proposent donc des compléments qui, pour certains comme **carte de résident de 10 ans**, amènent davantage d'informations à l'usager. D'autres reprennent la version complète du terme **carte de séjour de résident permanent**, ou ajoute simplement le complément **permanent** (dans **carte de résident permanent**). Il est intéressant de noter que le terme vedette est le terme juridique.

Le deuxième terme vedette, **passeport talent**, est également le terme dans sa forme réduite. Comme nous l'avons décrit dans la section 5.1.1.4, ce terme présente plusieurs stades de néologie de transfert du terme, de son statut juridique vers son statut de terme à part entière du discours administratif de l'accès aux droits. Le processus est davantage visible et nous formulons deux hypothèses. D'une part, le terme est plus récent. Jusqu'en 2016, il existait une **carte de séjour compétences et talents** qui a été remplacée par le **passepot talent**. D'autre part, la formation du terme est inhabituelle, puisqu'elle ne reprend pas le terme noyau **carte** ou **titre** ou encore **visa**, jusque là systématiquement utilisés. Le terme est formé à partir de **passeport**. Ce changement marque une nouvelle dynamique et une modification dans la méthode de dénomination. De plus, le complément **talent** est vague et imprécis. En réalité, le **passeport talent** ne correspond pas aux dynamiques discursives du domaine car il s'adresse à un public différent et a pour but d'attirer les « talents » contrairement au reste de la terminologie qui n'a pas un but at-

tractif. Cette nouvelle dynamique discursive peut donc être destabilisante pour l'usager comme pour le rédacteur, ce qui explique le grand nombre de variants.

L'ensemble de nos recherches n'abordent pas la présence (ou l'absence) de majuscules dans les termes car celle-ci est variable et ne semble pas correspondre à des règles spécifiques. Le grand nombre de sites et de rédacteurs différents mènent à une instabilité dans l'emploi des majuscules dans les termes. La présence (ou l'absence) de majuscules ne semble pas répondre à des règles ou à des logiques, et la cohérence dans l'emploi des majuscules ou minuscules n'est pas essentielle pour la compréhension du discours. Nous avons donc choisi de ne pas l'intégrer à notre étude.

Nous présentons ensuite l'analyse du domaine équivalent au Royaume-Uni dans la Figure ci-dessous.

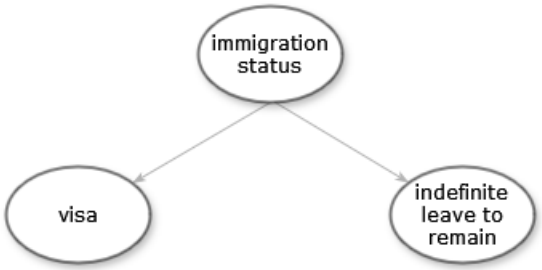


FIGURE 5.15 – Représentation ontologique des termes vedettes autour du concept de [document de séjour] au Royaume-Uni

La Figure 5.15 montre que le paysage terminologique du discours de l'accès au droit au séjour au Royaume-Uni est bien moins complexe. L'absence de terminologie compliquée n'indique pas le degré de conditionnalité des démarches ou la politique migratoire du pays. Toutefois, elle semble indiquer qu'au Royaume-Uni, ce domaine relève davantage du domaine juridique, duquel est issu l'ensemble de la terminologie, et que le droit au séjour ne constitue pas un domaine à part entière, mais un sous-domaine juridique. La terminologie présentée ici reflète donc cette appartenance juridique, avec un figement beaucoup plus marqué. On compte toutefois plusieurs groupes et discussions Reddit sur le sujet, ce qui tend à indiquer que le transfert de ces démarches sur les sites web et le manque d'accès direct aux professionnels du domaine crée une demande pour des experts non professionnels.

La création d'ontologies autour de la terminologie du droit au séjour permet de clarifier la place des différents termes au sein du domaine. De plus, elle permet également de mettre en avant les différences entre ces domaines en France et au Royaume-Uni.

Construire et fournir aux usagers les arborescences des dispositifs par l'analyse ontologique des concepts pourrait être un moyen efficace de leur éviter des difficultés d'inter-

prétation de l'information. Une méthode plus robuste pourrait être d'associer l'approche conceptuelle à l'approche lexicale, en construisant des arborescences qui informent sur le paradigme variationnel pour chaque concept.

5.2 Caractérisation syntaxique

5.2.1 Pronoms

Les pronoms jouent un rôle essentiel dans le discours de l'accès aux droits numérique. Pour déterminer leur place et leur rôle dans le discours étudié, nous avons exploré les pronoms et leur utilisation dans les corpus anglais et français.

5.2.1.1 Pronoms — corpus français

Les pronoms semblent avoir une place importante dans notre corpus. Cette observation est illustrée par la comparaison des pronoms dans le corpus général d'internet FRTenTen, un corpus juridique spécialisé contenant des textes fournis par l'Union européenne, et le corpus spécialisé de l'accès aux droits numérique qui constitue le principal sujet de cette étude.

Dans le corpus FrTenTen, qui représente les sites web dans leur ensemble, nous relevons 4,9 % de pronoms personnels. Pour le corpus EUR-LEX, corpus spécialisé dans le droit européen, nous relevons 0,84 % de pronoms personnels. Enfin, dans le corpus de l'accès aux droits numérique récolté pour cette étude, les pronoms personnels représentent 3,3 % du corpus total. Les pronoms ont donc une place moins importante que dans le corpus général, mais sont bien plus présents que dans le corpus spécialisé juridique.

Nous avons comparé les pronoms les plus utilisés dans le corpus général d'internet (FrTenTen) et dans le corpus spécialisé de l'accès aux droits, comme le montre la Table 5.45 ci-dessous.

	FrTenTen	Corpus d'accès aux droits numérique
1	il (14 %)	Vous (56 %)
2	vous (9 %)	il (9 %)
3	se (7 %)	s' (3 %)
4	nous (5 %)	je (3 %)
5	je (4 %)	se (3 %)
6	ce (4 %)	le (2 %)

TABLE 5.45 – Pronoms personnels les plus fréquents dans le corpus FrTenTen et le corpus d'étude français

La comparaison des deux corpus montre une grande disparité entre l'utilisation des pronoms en corpus numérique général et en corpus spécialisé dans l'accès aux droits. Dans le corpus de discours numérique général, le pronom personnel **il** qui représente 14 % des pronoms personnels dans le corpus est le plus fréquemment employé. Le deuxième pronom personnel le plus utilisé est **vous**, qui représente 9 %. Nous avons ensuite relevé plusieurs pronoms personnels qui représentent 7 % ou moins dans le corpus. Dans le corpus spécialisé de l'accès aux droits, le pronom le plus utilisé est **vous**, avec 56 % des occurrences de pronoms. Le deuxième pronom le plus utilisé est **il**, avec 9 % des occurrences. Le reste des pronoms personnels utilisés représentent 3 % ou moins. Les pronoms sont aussi présents, voire plus présents, dans le discours général numérique, mais leur utilisation est très différente. En effet, en français, d'après nos analyses statistiques, le pronom personnel **vous** est largement surreprésenté dans le corpus. Plus de la moitié des pronoms personnels utilisés (56 %) sont **vous**, ce qui représente une proportion bien plus élevée que dans le discours numérique général. Dans le corpus spécialisé juridique, les pronoms sont bien moins présents. Le pronom **vous** ne représente que 0,33 % des pronoms du corpus. Le discours de l'accès aux droits, s'il est issu du discours juridique, réintroduit l'usager dans le discours, ce qui est largement visible dans l'utilisation des pronoms. Quant à la fréquence très haute des pronoms et notamment du pronom **vous** dans le discours numérique général, elle pourrait provenir du caractère interactif de ce type de discours.

L'utilisation du **vous** dans le discours numérique de l'accès aux droits fait presque uniquement référence à l'utilisateur. Nous avons identifié les schémas les plus fréquents avec ce pronom. Pour ce faire, nous avons détecté les collocations du pronom de façon automatique. La Table 5.46 ci-dessous présente ces résultats.

Collocate	Freq	Coll. freq.	T-score	MI	logDice
si	4328	6693	64,50302	5,67852	12,03326
êtes	3611	4948	59,05197	5,85303	11,84835
Si	3593	6028	58,67192	5,56098	11,79338
pouvez	3320	6478	56,19995	5,34311	11,65994
,	9431	71195	87,85715	3,39118	11,61775
devez	3167	7144	54,67331	5,13386	11,56357
avez	2701	4151	50,9627	5,68751	11,46573
que	2208	5248	45,57924	5,05845	11,12526

TABLE 5.46 – Collocats les plus fréquents de **vous** détectés automatiquement par Sketch Engine

Avec le logiciel Sketch Engine, nous avons pu explorer les collocats du pronom **vous** détectés automatiquement. Cet outil nous présente des collocats candidats que nous présentons dans la Table 5.46. En explorant ces propositions en corpus, nous avons identifié

le schéma collocationnel le plus utilisé qui est le suivant :

si vous êtes [**profil administratif**]

L'étude de ce schéma, ainsi que l'ensemble des schémas conditionnels du corpus est abordée dans la section 5.2.2.

L'étude des collocations proposées par le logiciel Sketch Engine nous a également conduit à nous intéresser aux verbes utilisés avec ce pronom. Les relevés réalisés sont présentés dans la Table 5.47 ci-dessous.

word	Frequency	Freq. per million	% of concordance
Vous devez	4190	1949,0983	11,71045
Vous pouvez	3364	1564,86078	9,4019
vous êtes	3057	1422,05095	8,54388
vous devez	2735	1272,26345	7,64394
vous pouvez	2694	1253,19113	7,52935
vous avez	2100	976,87504	5,8692

TABLE 5.47 – Verbes les plus fréquemment utilisés avec le pronom **vous**

Les résultats présentés dans la Table 5.47 permettent d'identifier une grande utilisation des verbes suivants :

— devoir
— pouvoir
— être
— avoir

Ces résultats indiquent que les verbes les plus utilisés expriment une modalité, à savoir la modalité déontique. Le verbe le plus utilisé exprime une obligation (avec 6 925 occurrences), tandis que le second verbe exprime une possibilité (avec 6 058 occurrences). Cela peut être surprenant, car le domaine de l'accès aux droits indique un domaine de possibilités, de permissions, ou d'accès. Toutefois, la montée en conditionnalité présentée dans la section 1.2.2 du chapitre 2 confirme cette tendance à s'orienter vers la valeur d'obligation de la modalité déontique.

L'ensemble de ces résultats concernant le pronom **vous** montrent la place centrale qu'occupe ce pronom et l'usager. Ils sont placés en tant qu'acteurs principaux de ce domaine, qui bénéficient de possibilités, mais également d'obligations. Ils sont au centre de la conditionnalité, comme le montre le schéma collocationnel relevé. L'interrogation du corpus montre également une place importante occupée par les deux auxiliaires, **être** et **avoir**. Il est intéressant de relever que certains organismes sont présents dans le corpus, mais ne sont pas identifiés par la recherche quantitative réalisée par Sketch Engine avec les pronoms. Il s'agit principalement de la Caf et de Pôle emploi, qui sont mentionnés à plusieurs reprises. Le corpus contient 2 646 occurrences de **caf**, 2 417 occurrences de

Pôle emploi, contre 41 471 occurrences de **vous**. Nous aborderons la phraséologie de ces organismes dans la section 5.3.1.2.

Nous avons ensuite exploré le second pronom personnel présent dans notre corpus : **il**, qui représente 9 % des relevés. Ce pronom peut également faire référence aux usagers à l'aide de tournures impersonnelles, comme nous observons dans les exemples ci-dessous :

(59) Un parent isolé peut bénéficier du RSA s'il remplit plusieurs conditions.

(60) il faut au préalable remplir le dossier social étudiant

Ces deux exemples montrent que même lorsqu'un autre pronom est utilisé, il peut encore faire référence aux usagers, qui restent la figure centrale du discours.

Le pronom **il** fait également référence à un tiers dans les démarches, qui participe aux démarches ou bénéficie des aides. Les exemples suivants montrent cette utilisation en corpus :

(61) Cette obligation s'applique s'il a déjà placé ses salariés en activité partielle au cours des 36 mois précédant la date de dépôt de sa nouvelle demande.

Dans cet exemple, le tiers auquel le texte fait référence est l'employeur.

(62) Il est apprenti, stagiaire ou salarié.

Dans cet exemple, le tiers est un enfant qui perçoit un salaire au sein du foyer fiscal.

(63) Il doit dater d'au maximum 1 an.

(64) Il est établi par le médecin qui suit l'enfant au titre de la maladie, du handicap ou de l'accident.

Ces deux exemples font référence à des pièces justificatives. Ils montrent que ces pièces jouent également un rôle important dans le discours.

Enfin, le dernier type d'utilisation est la tournure impersonnelle simple, comme présentée dans les exemples ci-dessous :

(65) Il n'y a pas d'âge minimum pour demander l'Asi.

(66) Il est inutile de demander le complément familial.

Le troisième pronom personnel utilisé en corpus est le pronom personnel sujet **je**. Il peut sembler surprenant dans le domaine de l'accès aux droits numérique, notamment parce que les principaux rédacteurs de ce discours sont des organismes, qui donc utiliseraient le **nous**, comme pronom. Toutefois, en analysant l'ensemble des occurrences en corpus, le pronom **je** est utilisé uniquement pour faire référence à des usagers. Il s'agit de modulation de la voix du rédacteur, qui se place en tant qu'usager afin de montrer certains exemples. Les occurrences ci-dessous illustrent ce phénomène :

(67) J'ai deux enfants en primaire, mais je n'ai pas touché l'allocation de rentrée scolaire (Ars), pourquoi ?

(68) Je ne travaille plus et je suis en invalidité, le montant de mon allocation aux adultes handicapés va-t-il changer ?

(69) Je suis intérimaire, ai-je droit au Rsa ?

Les trois occurrences présentées ci-dessus proviennent du site `caf.fr`. Nous avons identifié 309 occurrences du pronom personnel **Je** (avec majuscule) sur le site de la Caf. L'ensemble de ces occurrences présentent une question, qu'elle soit directement dans la phrase ou dans la phrase qui suit. L'étude de ces contextes nous permet d'observer que les occurrences de **Je** ont un rôle classificatoire, notamment sur le site `caf.fr` : il existe un autre type d'occurrence du pronom sans majuscule sur le site `caf.fr`, il s'agit d'occurrences issues de menus et de titres dotés d'un hyperlien, faisant référence à l'espace en ligne, comme le montre la citation suivante.

(70) Pour des aides adaptées à ma situation, je déclare immédiatement tout changement dans Mon Compte.

Les sites restants proposent des occurrences avec des fonctions similaires, comme des fonctions classificatoires :

(71) Les différentes allocations auxquelles je peux prétendre en fonction de ma situation. (pole-emploi.fr)

(72) Comment faire si je ne parviens pas à utiliser le téléservice ? (service-public.fr)

En conclusion, les 2 751 occurrences du pronom personnel **je** font référence à l'usager uniquement, et non aux organismes produisant les textes étudiés. Pour une majorité de ces occurrences, il s'agit d'une phrase ayant pour rôle de fournir un exemple pour l'usager, ou de jouer un rôle classificatoire pour celui-ci : se repérer sur le site, ou savoir de quoi relève la page sur laquelle il se trouve.

Par ailleurs, il est intéressant de noter qu'en corpus, le pronom **je** est souvent utilisé pour faire référence à la nature numérique des démarches, comme le montre l'exemple suivant :

> « Je m'inscris aux ateliers organisés par Pôle emploi depuis mon espace personnel »

De la même manière, les termes faisant référence aux espaces en ligne où les usagers font les démarches sont souvent assortis du possessif **mon**. On trouve 127 occurrences de **mon espace** contre 114 occurrences de **votre espace**. Le pronom possessif à la deuxième personne du pluriel étant le plus fréquent dans l'ensemble du corpus, cette

utilisation semble surprenante. Cette caractéristique est encore plus présente dans l'utilisation en corpus du terme **compte**. En effet, on trouve 651 occurrences de **Mon Compte** contre 110 occurrences de **votre compte**. Sur le site `caf.fr`, les références de **votre compte** (30 occurrences) ne concernent pas l'espace en ligne. Nous pouvons donc observer une tendance à utiliser la première personne du singulier afin de faire référence aux outils numériques, tandis que le volet administratif et juridique du site utilise plutôt la deuxième personne du pluriel. Ces disparités montrent à nouveau la double origine de notre discours : numérique et juridico-administrative.

5.2.1.2 Pronoms — corpus anglais

Les conclusions tirées dans le corpus français montrent que les pronoms ont des caractéristiques spécifiques dans le domaine de l'accès aux droits numérique en France. Nous allons maintenant explorer le rôle des pronoms dans le discours de l'accès aux droits numérique au Royaume-Uni.

Nous avons à nouveau comparé trois corpus différents : un corpus d'anglais général numérique, le EnTenTen, un corpus juridique spécialisé, le EUR-LEX spécialisé dans le droit européen, et le corpus spécialisé créé pour notre étude dans le domaine de l'accès aux droits. Le corpus juridique présente 0,85 % de pronoms personnels, avec une large majorité de *it* qui représente 25 % des occurrences relevées. Dans le corpus web général, l'EnTenTen, les pronoms personnels représentent 4,8 % du corpus total, et le pronom *it* est également le plus utilisé, avec 11,5 % de l'ensemble des pronoms. Enfin, dans notre corpus spécialisé, les pronoms représentent 7,8 % de l'ensemble du corpus, et le pronom personnel le plus utilisé est *you*, avec 50 %, le deuxième est *your*, avec 24 %.

Tout d'abord, nous remarquons une grande disparité entre le nombre de pronoms personnels dans le corpus juridique spécialisé, qui est bien inférieur aux deux autres corpus. Les pronoms ne semblent pas avoir un rôle important, ce qui est dû à la nature du discours qui se doit de conserver la neutralité (Rouski, 2018). À l'inverse, le corpus spécialisé dans l'accès aux droits présente un taux de pronom plus élevé que le corpus web général.

Nous allons maintenant comparer les pronoms personnels utilisés dans le discours de l'accès aux droits numérique et dans le discours web général afin d'en identifier les éventuelles différences.

	EnTenTen	Corpus spécialisé de l'accès aux droits
1	it (15,5 %)	you (52 %)
2	I (11 %)	your (26 %)
3	you (10 %)	it (5 %)
4	we (8 %)	they (4 %)
5	their (6,5 %)	their (2,5 %)

TABLE 5.48 – Pronoms personnels les plus fréquents dans le corpus EnTenTen et le corpus d'étude anglais

La Table 5.48 nous permet d'observer une disparité entre l'utilisation des pronoms dans les deux corpus. Dans le corpus de l'accès aux droits, nous retrouvons une majorité de *you*, qui correspond à 52 % des pronoms personnels, puis de *your*, qui représente 26 %. Le lecteur ou l'usager sont donc très présents dans ce corpus. Au contraire, dans le corpus général d'internet, il y a une meilleure répartition des différents pronoms et le pronom le plus utilisé est neutre (*it*). Il est intéressant de noter que le deuxième pronom personnel le plus utilisé est *I*, qui pourrait être issu de certains genres numériques comme le blog. La comparaison de l'utilisation des pronoms en corpus spécialisé de l'accès aux droits montre une différence claire et une place importante à l'usager dans ce discours. La forte présence du pronom *you* peut être liée à la vulgarisation du discours, comme le constate l'étude de Pic et Furmaniak (2012).

Pour approfondir notre étude sur ce pronom, nous allons relever ses collocats détectés automatiquement par l'outil Sketch Engine, afin d'en étudier les caractéristiques. La Table 5.49 ci-dessous propose un récapitulatif des collocats les plus significatifs.

Collocate	Freq	Coll. freq.	T-score	MI	logDice
.	10748	30792	91,62155	3,10481	12,28866
can	6442	8292	76,07029	4,25908	12,1058
if	6308	8320	75,17255	4,22389	12,07463
to	7866	28462	75,66967	2,76796	11,88685
're	5177	5314	68,95475	4,58561	11,88302
If	5146	6380	68,12706	4,31319	11,84052
get	3979	5396	59,60848	4,18381	11,50068
need	3711	4510	57,91411	4,34198	11,42877
have	3806	5935	57,78945	3,98232	11,41937
'll	3641	3986	57,66044	4,49269	11,41854

TABLE 5.49 – Collocats les plus fréquents de *you* détectés automatiquement par Sketch Engine

Les collocats présentés dans la Table 5.49 nous ont permis d'identifier un schéma collocationel fréquent dans le corpus. En effet, les deuxième et troisième collocats proposés par le logiciel sont les suivants : *can* et *if*. En explorant les occurrences en corpus de ces collocats avec *vous*, nous avons trouvé diverses occurrences du schéma suivant :

If you're *[profil administratif]*

Ce schéma conditionnel sera abordé dans la section 5.2.2.

Le second collocat *can* montre également le recours important à la modalité de type déontique dans le discours, liée notamment aux usagers. Nous avons également identifié deux schémas collocationnels qui intègrent les deux collocats :

If you cannot/can't

If you can get

Ces deux schémas montrent que la modalité et la conditionnalité ont une place dans le corpus. Elles permettent de déterminer le profil de certains utilisateurs, mais représentent aussi leur rôle actif dans les démarches et dans le discours en général. Les usagers sont les acteurs principaux de leur droit, notamment depuis la transformation numérique. Ce rôle est traduit par l'abondance de phrases où l'usager se retrouve en position de sujet, et le peu de sujets alternatifs. Pour mieux comprendre la place des usagers, nous avons relevé dans la Table 5.50 les verbes les plus fréquemment utilisés avec le pronom *you*.

word	Frequency	Freq. per million	% of concordance
you 're	4742	4858,65136	23,52416
you have	2128	2180,34797	10,5566
you 've	1468	1504,11223	7,28247
you need	1043	1068,6574	5,17412
you do	1014	1038,94401	5,03026
you get	915	937,50865	4,53914
you are	894	915,99205	4,43496
you apply	625	640,37476	3,10051
you want	390	399,59385	1,93472

TABLE 5.50 – Verbes les plus fréquemment utilisés avec le pronom *you*

La Table 5.50 montre qu'il y a une grande majorité de "you're", probablement issue du schéma :

if you are [*profil*]

Ce schéma a pour but de déterminer le profil des usagers, afin de leur donner les informations adéquates. En effet, chaque dispositif nécessite une situation spécifique, qui est déterminée par un certain nombre de paramètres. Ces situations sont donc importantes aux yeux de l'administration, car elles permettent de catégoriser les usagers et de leur octroyer, ou non, le droit demandé. Cette particularité du discours est visible et explique la présence marquée des deux verbes suivants : *to be* et *to have*.

(73) (...) if you're 21 or under, studying any qualification up to A level or equivalent and do not have parental support.

(74) If you're in a couple but you do not live together, you'll get the amounts for a single person.

(75) If you have a disability or health condition

(76) Financial help if you have children

L'ensemble des exemples ci-dessus montrent que ces collocations ont pour but d'établir des profils. La navigation indépendante de l'usager sur le site suppose une diversité des profils à déterminer. De plus, la nature du site web fait que les usagers doivent déterminer leur profil eux-mêmes et les textes doivent intégrer cette dimension.

Au-delà de la place occupée par le pronom **vous**, et donc par les usagers dans le texte, ceux-ci sont aussi présents dans d'autres pronoms. Nous analysons également la place du pronom *they* dans notre corpus, puisqu'il s'agit du deuxième pronom le plus utilisé. Grâce à notre exploration de corpus, nous avons pu identifier trois utilisations de ce pronom. Tout d'abord, il est utilisé pour faire référence à des usagers, comme dans l'exemple suivant :

(77) If you live with your partner, they will also need to set up an account.

Plus précisément, le pronom *they* peut faire référence à un tiers usager en lien avec l'usager principal ou le lecteur à qui est destinée l'information principale. Le pronom est également utilisé pour faire référence à un type d'usager en particulier : les personnes à charge. Il s'agit d'un tiers qu'un usager soutient fiscalement, qui apparaît donc dans le discours comme dans les phrases suivantes :

(78) (...) the person you care for will stop getting : a severe disability premium paid with their benefits an extra amount for severe disability paid with Pension Credit, if they get one They might also stop getting reduced Council Tax.

À partir de ces relevés, nous avons identifié un schéma collocationnel concernant les personnes à charge dans le corpus. Il s'agit du schéma suivant :

They + *health condition*.

Il est représenté dans les exemples suivants :

(79) (...) they're blind

(80) (...) they're on dialysis

(81) (...) they're terminally ill

(82) They cannot walk

Dans tous ces cas, il s'agit à nouveau d'un usager, même s'il n'est pas directement adressé.

Le pronom personnel *they* est utilisé également pour faire référence à d'autres organismes publics qui font partie du domaine. Toutefois, cette utilisation reste très limitée. Les contextes ci-dessous illustrent les références à d'autres organismes :

(83) HMRC will continue to reclaim overpayments while they review your dispute.

(84) If they do not change it or you still think it's wrong You can ask for the decision to be reconsidered

Les organismes sont donc présents dans le corpus, mais ils ne sont pas aussi présents que les usagers.

Le dernier pronom auquel nous allons nous intéresser est le pronom personnel *we*. Sa présence peut sembler surprenante étant donné que l'on retrouve un effacement de l'administration par rapport aux usagers, qui ne semblent pas avoir d'interlocuteurs. Il n'existe qu'une occurrence de ce pronom sur le site `gov.uk`, mais plusieurs occurrences sur les autres sites. En analysant le contexte de ces occurrences, on observe un ton très différent pour les municipalités, présenté dans les contextes ci-dessous :

(85) We will always ask for your permission before we do this. (Falkirk Council)

(86) When we arrange a place in a home for you, we will assess your income and savings. (Falkirk Council)

(87) If you think that the decision we made about your Housing Benefit is wrong, you can ask us to look at it again. (Bournemouth, Christchurch and Poole Council)

(88) If you rent from a private landlord, the maximum amount of Housing Benefit we can pay is based on the size of your household, not the amount of rent being charged by your landlord. (Bournemouth, Christchurch and Poole Council)

Dans ces exemples, nous observons non seulement une voix indépendante des municipalités, mais également une implication dans les démarches administratives. Il y a ici une identification du sujet, qui est acteur des décisions, ce qui n'est pas le cas dans le reste du corpus. Cela montre que même s'il existe des similarités dans les discours de l'accès aux droits, il reste certaines disparités selon les rédacteurs et les institutions.

En conclusion, ces analyses montrent que l'usager est une figure centrale du discours de l'accès aux droits numérique. Le site web et l'absence d'interlocuteur ont pour conséquence la multiplication des schémas ayant pour but de décrire le profil des usagers. Cette prévalence du pronom *you* peut provenir des caractéristiques liées à la vulgarisation du discours. En effet, le discours juridique est neutre, mais le discours de l'accès aux droits réintroduit les usagers au centre du discours afin qu'ils puissent accéder à leurs droits. C'est ce que Cacchiani (2018, p. 29) observe sur le site `gov.uk` :

> on aspire à engager l'utilisateur en présentant les connaissances comme directement orientées vers ses besoins Cacchiani (2018, p. 29)

5.2.1.3 Analyse contrastive

Nous avons relevé dans les deux corpus une utilisation des pronoms personnels particulière. Dans les deux langues, les corpus d'étude contiennent davantage de pronoms que les corpus spécialisés juridiques, qui en présentent très peu et restent neutres en raison de leur spécialisation. En français, le corpus web général présente plus de pronoms que le corpus de l'accès aux droits, sans que cette différence ne soit forcément significative, puisqu'elle est de moins de 1 %. En anglais, au contraire, le corpus de l'accès aux droits présente un taux plus élevé de pronoms que le corpus web général (3 % de plus). Dans les deux cas, l'utilisation des pronoms semble propre au discours de l'accès aux droits, puisque le pronom personnel sujet **vous** et son équivalent anglais *you* représentent plus de la moitié des pronoms utilisés. Ce phénomène indique une place centrale des usagers dans le discours. D'ailleurs, ils sont presque les seuls sujets présents en discours, tandis que les organismes sont plus absents. Dans le corpus français, on compte environ 6 000 occurrences des organismes d'accès aux droits sociaux, tandis qu'il y a plus de 41 000 occurrences du pronom **vous**. Dans le corpus anglais, nous avons noté une disparité entre les différents sites émetteurs du discours. Les municipalités semblent avoir une position spécifique, en s'impliquant davantage et en devenant le sujet du discours. Ce n'est pas le cas d'autres organismes nationaux, qui ne sont présents que très rarement dans le reste du discours.

En conclusion, les analyses et explorations faites en corpus confirment le rôle central de l'usager dans le discours étudié. Ce rôle est caractéristique de l'accès aux droits, qui réintroduit le sujet au cœur du discours. Les particularités linguistiques peuvent également être attribuées au caractère vulgarisé du discours, dans lequel le lecteur a une place privilégiée (Pic et Furmaniak, 2012).

5.2.2 Conditionnalité : si vous êtes [profil administratif]/if you're [profil administratif]

La conditionnalité est une caractéristique saillante du discours numérique de l'accès aux droits comme décrit dans la section 1.2.2, qui est aussi présente dans les phrases répétées (voir section 5.4.1). Cette conditionnalité est présente dans le mode de communication avec les administrations, par exemple dans la nécessité de posséder un équipement informatique afin d'accéder au discours. Nos analyses préalables ont permis d'identifier certains schémas conditionnels. Cette section est donc consacrée à l'exploration de la conditionnalité du discours au niveau linguistique et discursif.

L'utilisation du pronom *you*, qui est le pronom principal du corpus, est largement liée, dans les deux langues, à des phrases conditionnelles. Dans la section 5.2.1, nous

avons identifié le schéma conditionnel suivant :

if you're *[profil administratif]*

Afin d'éclairer nos connaissances sur ce schéma, nous avons récolté ses occurrences les plus fréquentes en corpus, présentées dans la Table 5.51 ci-dessous.

if you're	eligible
	disabled
	in a couple
	a single parent
	self-employed

TABLE 5.51 – Occurrences les plus fréquentes pour le schéma : if you're [profil]

L'étude des occurrences en corpus nous a permis de déterminer que ce schéma conditionnel est moins fréquent en anglais et qu'il est le plus souvent composé avec le complément *eligible*. Les conditions d'éligibilité ne sont pas toujours exprimées dans ce schéma, mais il indique plutôt que le dispositif est conditionnel. Il existe tout de même quelques occurrences de ce schéma, présentées dans la Table 5.51. Nous avons également identifié un deuxième schéma conditionnel dans le corpus anglais. Il s'agit du schéma :

if you're applying for *[dispositif d'aide sociale]*

Les occurrences les plus fréquentes de ce schéma sont présentées dans la Table 5.52 ci-dessous.

if you're applying for	a visa
	a visitor visa
	an Intra-Company Graduate Trainee visa

TABLE 5.52 – Occurrences les plus fréquentes pour le schéma : if you're applying for [dispositif]

L'analyse des occurrences relevées de ce schéma montre qu'il est majoritairement utilisé dans le domaine des visas. Elle montre également que les schémas conditionnels sont plus ou moins complexes selon les situations et les profils, mais aussi selon les besoins du discours. Le profil de l'usager doit être affiné au cours de ses recherches : d'abord on l'oriente vers une page « visa » puis ensuite vers le visa spécifique.

Nous allons maintenant étudier le schéma équivalent en français :

si vous êtes *[profil administratif]*

Ce schéma conditionnel compte 2 419 occurrences en corpus. Afin d'éclairer nos connaissances sur ces schémas, nous avons récolté ses occurrences les plus fréquentes en corpus, présenté dans la Table 5.53 ci-dessous.

si vous êtes	marié
	dans l'une des situations suivantes :
	allocataire
	étranger
	sous protection juridique
	hébergé gratuitement
	salarié
	enceinte
	étudiant

TABLE 5.53 – Occurrences les plus fréquentes pour le schéma : si vous êtes [profil]

Les résultats présentés montrent que les profils qui apparaissent dans ce schéma peuvent être variés. De plus, il est important de noter que chacun de ces profils peut être accompagné d'un complément, comme dans les exemples ci-dessous :

(89) si vous êtes salarié ou ancien salarié de la construction et la réparation navales,

(90) si vous êtes étranger ayant une carte de séjour portant la mention « retraité »,

(91) si vous êtes étudiant boursier et n'avez pas terminé vos études au 1er juillet,

(92) si vous êtes étudiant isolé ou en couple, sans personne à charge.

Dans les exemples ci-dessus, nous pouvons identifier qu'il s'agit de compléments afin de préciser le statut de l'usager. Ces compléments montrent que les statuts des usagers peuvent être complexes à déterminer. Les phrases conditionnelles peuvent donc être également complexes.

En conclusion, les schémas conditionnels permettent de déterminer le profil des usagers. Leur utilisation dans le discours de l'accès aux droits est probablement antérieure à sa transition numérique. Sur les sites web, ils sont indispensables afin d'aider l'usager à déterminer son profil par rapport à l'aide demandée. Il est important de noter que les usagers n'ont pas un profil unique. Leur « profil administratif » change selon leurs besoins et les dispositifs demandés. De plus, il est déterminé par l'administration, qui fixe les critères et les conditions d'éligibilité. Le schéma conditionnel est donc une caractéristique importante du discours de l'accès aux droits numérique, il sert de point de repérage pour les usagers et permet de les guider sur les pages adaptées à leur besoin.

5.3 Caractérisation lexico-syntaxique

5.3.1 Schémas phraséologiques

Afin de compléter notre étude des schémas phraséologiques, mais aussi des pronoms en corpus, nous allons explorer dans cette section différents schémas phraséologiques caractéristiques de notre corpus. Ces schémas ont été identifiés lors de nos explorations précédentes ou proposés par l'outil de détection automatique *n-gram viewer* de Sketch Engine.

5.3.1.1 Schéma phraséologique autour des verbes modaux

Nous avons réalisé une étude des schémas phraséologiques autour des verbes modaux, qui sont très présents dans notre corpus. Nous avons réalisé une exploration automatique des modaux à l'aide de Sketch Engine. Les modaux les plus fréquents récoltés par Sketch Engine sont présentés dans la Table 5.54 ci-dessous.

word	Frequency	Freq. per million	% of concordance
can	8292	8495,97998	31,79204
will	4873	4992,87391	18,68338
'll	3986	4084,05405	15,28257
must	2037	2087,10941	7,80998
may	1887	1933,41947	7,23487
might	1407	1441,61165	5,39452
should	1280	1311,4875	4,9076
could	700	717,21973	2,68384

TABLE 5.54 – Auxiliaires modaux les plus fréquents dans le corpus d'étude

La Table 5.54 indique que les modaux les plus utilisés sont *can*, *will* et *must*. Ces modaux indiquent donc principalement la modalité déontique, c'est-à-dire la possibilité et l'obligation, qui sont au cœur de notre domaine. Nous pouvons également noter que les modaux sont utilisés fréquemment dans notre corpus, notamment si l'on compare leur fréquence d'utilisation dans la langue générale. Il est possible de faire cette comparaison en interrogeant le corpus enTenTen. Une telle interrogation nous informe que les modaux les plus fréquents en anglais général d'internet sont *will* et *can* comme le montre la Table 5.55.

word	Frequency	% of concordance
will	2708704	27,08704
can	2117469	21,17469
would	1182146	11,82146
may	854375	8,54375
could	675199	6,75199
should	665738	6,65738
must	497868	4,97868
might	269894	2,69894
'll	247594	2,47594

TABLE 5.55 – Auxiliaires modaux les plus fréquents dans le corpus web général ENTenTen

Il est intéressant de noter que les modaux sont plus fréquents dans le corpus spécialisé, notamment *must* qui apparaît en troisième position avec 7,8 % contre 4,9 % en corpus général et *will* avec 33 % contre 23 % en corpus général. Il semble que l'obligation soit légèrement plus présente dans le corpus de l'accès aux droits que dans le corpus web général. Nous avons ensuite fait des recherches afin d'identifier les collocats les plus fréquents avant et après les modaux. La Table 5.56 ci-dessous présente les collocats les plus présents avant les modaux.

word	Frequency	Freq. per million	% of concordance
you	8036	8233,68248	30,81052
You	6983	7154,77909	26,77325
they	729	746,93312	2,79503
They	558	571,72658	2,13941
we	515	527,6688	1,97454
We	473	484,63562	1,81351

TABLE 5.56 – Collocats les plus fréquents avec les auxiliaires modaux - gauche détectés automatiquement par Sketch Engine

Nous retrouvons ainsi des pronoms personnels sujets, y compris les deux pronoms les plus fréquents en corpus *you* et *they*, ainsi qu'un autre pronom qui est moins fréquent en corpus : *we*.

La Table 5.57 ci-dessous présente les collocats les plus fréquents de *will*.

word	Frequency	Freq. per million	% of concordance
need	1222	1252,06073	28,6451
get	765	783,8187	17,93249
be	683	699,80154	16,01031
have	216	221,31352	5,06329
also	206	211,06752	4,82888
n't	183	187,50173	4,28973
usually	170	174,18193	3,985
not	148	151,64074	3,46929
only	81	82,99257	1,89873
still	74	75,82037	1,73465
receive	57	58,40218	1,33615

TABLE 5.57 – Collocats les plus fréquents de *will* détectés automatiquement par Sketch Engine

Nous avons recherché par la suite les concordances à partir de *you will need to*, présentées dans la Table 5.58 ci-dessous.

Collocate	Freq	Coll. freq.
provide	135	973
prove	61	425
show	49	361
give	44	626

TABLE 5.58 – Collocats les plus fréquents de *You will need to* détectés automatiquement par Sketch Engine

À partir de ces relevés, nous avons identifié le schéma phraséologique suivant :

You will need to provide/prove/show/give [document OR information]

Ce schéma montre l'aspect conditionnel du discours, ainsi que les obligations qui incombent à l'usager. L'étude du second collocat proposé dans la Table 5.57 a mené à la recherche des collocats de *you will get* qui sont présentés ci-dessous.

word	Frequency	Freq. per million	% of concordance
letter	43	44,05778	24,85549
email	27	27,66419	15,60694
decision	13	13,31979	7,51445

TABLE 5.59 – Collocats les plus fréquents de *you will get* détectés automatiquement par Sketch Engine

Ces collocats relevés permettent de mettre en lumière le schéma phraséologique suivant :

you will get a letter/email/decision

Ce schéma phraséologique reflète également les relations bilatérales entre les usagers et l'administration.

Nous allons désormais nous intéresser au deuxième verbe modal le plus utilisé en corpus spécialisé : *can*.

word	Frequency	Freq. per million	% of concordance
get	889	910,86906	10,72118
apply	697	714,14593	8,40569
not	692	709,02293	8,34539
be	534	547,13619	6,43994
also	369	378,07726	4,45007
use	270	276,6419	3,25615
ask	260	266,3959	3,13555
help	250	256,1499	3,01495
claim	239	244,87931	2,8823

TABLE 5.60 – Collocats les plus fréquents de *can* détectés automatiquement par Sketch Engine

La Table 5.60 récapitule l'ensemble des collocats les plus fréquemment utilisés avec *can*. À partir de ces collocats relevés sur l'outil Collocations de Sketch Engine, nous avons identifié le schéma phraséologique suivant :

You can get help

Ce schéma nous fait basculer à un autre type de modalité déontique caractéristique de notre discours, soit de l'obligation à la possibilité. Le schéma reposant sur *can* informe l'usager sur la possibilité d'obtenir une aide. Le schéma phraséologique suivant est un autre exemple de l'expression de la possibilité :

You can apply for [dispositif]

Ces deux schémas indiquent que le discours est porteur d'obligation, de conditionnalité, mais aussi de possibilités pour l'usager.

5.3.1.2 Phraséologie autour des organismes d'accès aux droits

Dans la section 5.2.1.1, nous avons identifié une forte représentation de l'usager, avec une fréquente utilisation du pronom **vous** ou *you*, tandis que les organismes semblaient occuper une place moins importante dans le discours. Dans le corpus français, seulement deux organismes sont présents de façon constante et récurrente. Ces organismes sont la caisse d'allocations familiales, communément appelée Caf, et Pôle emploi [10]. Dans cette

10. Pôle emploi est devenu France Travail au 1er janvier 2024. Notre corpus ayant été récolté en 2021, nous continuons d'utiliser le nom Pôle emploi qui apparaît dans notre corpus.

section, nous nous sommes intéressée aux différents organismes et à leur place dans le corpus en analysant leur phraséologie.

5.3.1.2.1 Zoom sur la phraséologie autour de Caf

Nous nous sommes d'abord concentrée sur la phraséologie autour de l'organisme Caf. Nous avons identifié les collocations les plus fréquentes grâce à l'outil Collocations proposé par Sketch Engine, que nous présentons dans la Table 5.61 ci-dessous.

collocat	occurrences	candidats	Logdice
aider	89	341	11.45753
verser	75	181	11.45239
accorder	26	111	10.04401
m"	54	1325	9.77403
me	19	212	9.42138
automatiquement	25	674	9.22390
souvent	14	156	9.07265
relais	11	29	8.95759
directement	21	730	8.91456
particulier	12	158	8.84687

TABLE 5.61 – Collocations automatiquement détéctées du schéma [caf] [verbe]

À partir des résultats de la Table 5.61, nous avons mis en place l'expression CQL suivante :

([word="caf"]|[word="Caf"]) []{0,5} ([lemma="aider"]|[lemma="aide"])

Cette expression correspond aux occurrences de Caf (avec majuscule ou non) suivi du verbe **aider** ou du nom **aide** à une distance de 0 à 5 mots. Cette recherche nous a permis de récolter 158 occurrences de la collocation **Caf + aide**. Il semble donc que cette collocation soit bien présente dans notre corpus, comme dans les exemples ci-dessous :

(93) la Caf peut-elle m'aider financièrement ? (`caf.fr`)

(94) Votre Caf peut vous aider. (`caf.fr`)

(95) La Caf m'aide-t-elle financièrement, quel que soit le mode de garde que je choisis ? (`caf.fr`)

(96) Je souhaite employer une personne pour garder mon enfant, la Caf peut-elle m'aider à payer son salaire ? (`caf.fr`)

Nous pouvons tout d'abord observer un grand nombre de questions dans les occurrences identifiées. Sur ces 158 occurrences, nous avons identifié 59 occurrences comme étant des questions. Nous remarquons aussi qu'il est souvent précisé qu'il s'agit d'une

aide **financière**. En effet, **financièrement** est le premier collocat de notre patron phraséologique proposé par Sketch Engine, avec 15 occurrences.

Nous remarquons dans un second temps un ton neutre et détaché de la part de la Caf. Dans l'ensemble des exemples mis en avant, qui sont extraits du site de la Caf, il semble que la Caf est un tiers, et non l'organisme diffusant les informations. L'exemple 94 illustre ce phénomène avec l'utilisation du pronom possessif dans **Votre Caf**, qui semble indiquer une distance entre le site web et la Caf, alors que l'usager se trouve sur le site de l'organisme en question `caf.fr`. De la même façon, l'occurrence répétée de **« contactez votre Caf »** marque une distance entre l'usager et l'organisme. De plus, on remarque l'utilisation du pronom personnel **je** dans certaines questions. Cette utilisation renforce la distance entre le rédacteur, qui est issu de l'organisme, et l'usager. En effet, le rédacteur se substitue à l'usager et s'efface.

Nous remarquons que les deux collocations suivantes sont présentes dans la Table 5.61 : **verser** et **accorder**. Ces collocations montrent une prosodie sémantique positive avec **aide**. Nous avons identifié les schémas suivants :

« la Caf peut accorder »

« la Caf peut vous verser »

Dans ces schémas, nous retrouvons l'utilisation de **vous**, et l'utilisation de **je** est moins présente.

En conclusion, l'étude de cette collocation nous a permis d'identifier un effacement de l'organisme d'aide sociale, tout en gardant une phraséologie avec une prosodie sémantique positive, qui tend à renforcer l'association entre le nom de l'organisme et l'accord d'une **aide financière**.

5.3.1.2.2 Zoom sur la phraséologie autour de Pôle emploi

Notre étude de la phraséologie liée à l'organisme Caf a révélé l'utilisation de la première personne du singulier dans les phrases faisant référence aux organismes. Cette utilisation a également été notée pour l'organisme Pôle emploi. Nous avons étudié les collocations détectées automatiquement pas Sketch Engine comme pour l'organisme Caf. Voici les résultats que nous avons extraits dans la Table 5.62 ci-dessous.

Collocate	Freq	Coll. freq.	T-score	MI	logDice
accompagner	136	343	11,62883	8,46211	10,65701
Je	220	2159	14,66874	6,50192	10,62149
moi	78	99	8,81916	9,45276	9,98849

TABLE 5.62 – Collocats les plus fréquents de Pôle emploi détectés automatiquement par Sketch Engine

Nous pouvons tout d'abord observer que Pôle emploi **accompagne** ses usagers, tandis que la Caf **aide** ses usagers. De plus, les deuxième et troisième relevés indiquent l'usage des pronoms personnels **je** et **moi**, comme dans les relevés précédents.

(97) J'envoie un document à Pôle emploi

(98) Je m'inscris aux ateliers organisés par Pôle emploi depuis mon espace personnel : je choisis le thème, la date et le lieu du rendez-vous.

(99) Je m'actualise en déclarant un évènement.

Il s'agit généralement de phrases décrivant une action faite par l'usager. La phrase est souvent présente en titre, indiquant ainsi à l'usager que s'il souhaite s'actualiser, il faut cliquer sur le lien (Figure 5.16).

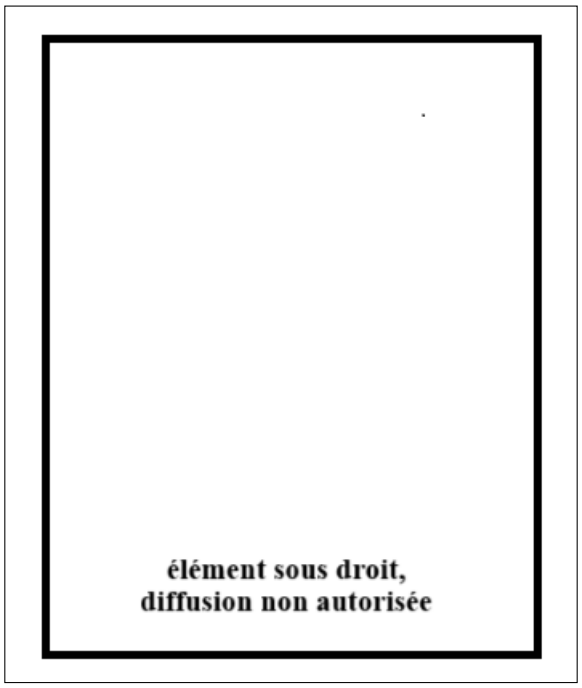


FIGURE 5.16 – Capture d'écran d'une rubrique du site France Travail (anciennement Pôle emploi)

source : `france-travail.fr`, consulté le 26 juin 2024

La capture (Figure 5.16) présente une rubrique proposée sur le site Pôle emploi. Elle montre l'utilisation du **je** pour diriger l'utilisateur, qui pourra alors cliquer sur la rubrique afin de s'informer. L'utilisation du pronom personnel **je** sur le site `service-public.fr` est limité : nous n'avons identifié que 68 occurrences, comme nous l'avions noté dans la section 5.2.1.1.

Les organismes semblent donc utiliser plus souvent le pronom personnel sujet **je**, notamment pour aider l'usager à se repérer sur le site. L'utilisation de son point de

vue permet de l'impliquer dans la recherche et dans le site web. Les organismes tels que la Caf et Pôle emploi sont en lien direct avec leurs usagers, tandis que le site `service-public.fr` ne fait que recenser les informations issues des différents organismes.

5.3.1.3 Zoom sur la collocation vous + pouvoir/devoir

Nous avons formalisé la collocation vous + pouvoir avec l'expression CQL suivante : [lemma="vous"][lemma="pouvoir"], qui compte 6 349 occurrences en corpus. La Table 5.63 présente les résultats de cette exploration.

Collocate	Freq	Coll. freq.	T-score	MI	logDice
saisir	511	716	22,51176	7,91676	11,21071
bénéficier	567	2238	23,53418	6,42261	11,07927
demander	552	2077	23,23359	6,49164	11,06789
faire	550	3370	23,02768	5,78815	10,8567
demande	782	9419	26,96948	4,81306	10,66631
aussi	297	1267	17,01656	6,31052	10,3195
votre	1080	22128	30,87472	4,04663	10,27931
former	227	279	15,01183	8,10583	10,13219

TABLE 5.63 – Collocats les plus fréquent de **vous pouvez** automatiquement détectés par Sketch Engine

Dans la Table 5.63, nous pouvons observer les principales collocations de **vous pouvez** :

— **vous pouvez saisir**
— **vous pouvez bénéficier**
— **vous pouvez demander**
— **vous pouvez faire une demande**

Ces collocations montrent deux types d'actions possibles pour l'usager. Il peut faire une demande concernant les différents droits auquel il peut accéder (**saisir, demander, faire une demande**) ou il peut recevoir une aide (**bénéficier**).

Nous avons ensuite formalisé la collocation vous + devoir avec l'expression CQL suivante : [lemma="vous"][lemma="devoir"], qui compte 7 392 occurrences en corpus. La Table 5.64 présente les résultats de cette exploration.

Collocate	Freq	Coll. freq.	T-score	MI	logDice
joindre	770	2835	27,39756	6,30354	11,26862
remplir	652	1492	25,33337	6,98966	11,23173
fournir	449	1469	20,95123	6,47392	10,69732
volet	384	415	19,5231	8,07196	10,65441
déclarer	383	746	19,43931	7,22213	10,59074
pièces	419	1895	20,15115	6,00679	10,52981
déposer	367	808	19,01221	7,04539	10,51823
faire	444	3370	20,52136	5,25985	10,40076

TABLE 5.64 – Collocats les plus fréquent de **vous devez** automatiquement détectés par Sketch Engine

Dans la Table 5.64, nous avons identifié les principales collocations de **vous devez** :

— **vous devez joindre**

— **vous devez remplir**

— **vous devez fournir**

Il semble que dans ce cas les collocations font référence à des justificatifs. Les obligations des usagers peuvent donc être liées aux justificatifs permettant de valider leur profil administratif. Nous avons également identifié un grand nombre d'occurrences de :

(100) Vous devez joindre (scanner)

On en retrouve 200 occurrences, mais il s'agit d'une phrase répétée. La précision du mot **scanner** entre parenthèses indique toutefois que les échanges sont faits de façon dématérialisée et que certains usagers peuvent avoir besoin d'explicitations.

Concernant le schéma collocationnel :

« vous devez remplir »

Il est généralement suivi de **formulaire** ou **déclaration**. Ces collocations sont donc issues du discours administratif.

Le discours de l'accès aux droits numérique présente donc des caractéristiques phraséologiques. Tout d'abord dans l'utilisation de la modalité déontique, mais aussi dans la phraséologie liée aux organismes de droits sociaux.

5.4 Caractérisation stylistique et rhétorique

5.4.1 Répétitions

5.4.1.1 Citations ou reprises mot à mot

Une précédente étude réalisée sur un corpus d'accès aux droits dans le domaine de l'aide sociale nous a permis d'identifier un phénomène de répétitions (Bouchet, 2021).

Nous souhaitons dans la présente section réaliser une typologie de ce phénomène de répétition. En effet, les répétitions sont caractéristiques des deux principales composantes de notre discours : sa spécialisation liée au discours administratif et juridique, et sa nature numérique. La nature délinéarisée des sites web, présentée dans la section 3.2 et notamment l'utilisation d'hyperliens, décrit dans la section 3.2.3, accentue la présence de phrases répétées.

Le discours de l'accès aux droits, fortement dépendant de la langue de spécialité juridique, hérite également de ses caractéristiques et répond à des critères « de clarté, d'ordre, de précision, d'impartialité et d'austérité » (Rey, 2008, p. 135). Le figement et la répétition permettent de répondre à ces critères.

La combinaison de ces deux facteurs notamment fait que les répétitions sont très présentes dans notre corpus. En utilisant le logiciel d'exploration de corpus Sketch Engine, nous avons établi une typologie des répétitions.

5.4.1.2 Éléments définitoires

Un certain nombre de segments répétés sont en fait des éléments définitoires.

(101) La prime d'activité est versée chaque mois. - 3 occurrences - (caf.fr)

(102) L'actualisation sert à déclarer sa situation tous les mois qu'elle soit professionnelle ou personnelle. - 10 occurrences - (pole-emploi.fr)

(103) Pour la rentrée 2020, l'Ars peut être versé pour chaque enfant scolarisé né entre le 16 septembre 2000 et le 31 décembre 2014 inclus, et pour chaque enfant plus jeune déjà inscrit en CP. - 15 occurrences - (caf.fr)

(104) Le certificat de résidence est valable 1 an. - 9 occurrences - (service-public.fr)

(105) Jobseeker's allowance is the benefit paid to unemployed people who are actively looking for work. 10 occurrences - (gov.uk)

(106) The tribunal is impartial and independent of government. 5 occurrences — (gov.uk)

Pour analyser ces éléments définitoires répétés, il est important de noter que l'ensemble des définitions sont issues du discours juridique. Il existe ainsi des définitions correspondantes dans les textes de loi. L'équivalent du premier exemple est issu de Titre IV : Prime d'activité (Articles R842-1 à R848-1).

> « La prime d'activité est versée mensuellement à terme échu. » (Articles R842-1 à R848-1)

Il s'agit bien ici d'une transposition d'un discours juridique en discours de l'accès aux droits destiné aux usagers. L'ensemble des autres exemples, même s'ils ne sont pas tous directement issus de textes juridiques, correspondent à des concepts juridiques.

Nous avions abordé dans la section 2.2.7.1 le discours juridique en soulignant les exigences qui pèsent sur ce type de discours, qui nécessite de la rigueur et de l'ordre. Les éléments définitoires doivent donc être cohérents. Même si le discours de l'accès aux droits n'a pas de valeur juridique propre, son interprétation peut déterminer les actions des usagers. Le figement de certains éléments définitoires peut donc être expliqué par le caractère juridique et prescriptif du discours. Toutefois, le manque de paraphrases et d'explicitations peut empêcher la compréhension pour certains usagers.

5.4.1.3 Reformulation ou reprise partielle

Même si cela n'est pas courant et même quasi inexistant dans le corpus britannique, nous avons relevé des phrases répétées présentant quelques modifications. Ces modifications sont rares et posent la question de la reformulation. Voici ces exemples :

(107) — La personne AIDANTE DOIT : — réduire : son activité salariée du secteur public ou privé — `caf.fr`

(108) — La personne aidante DOIT : — réduire : son activité salariée du secteur public ou privé — `caf.fr`

(109) Aucune récupération ne peut être exercée à l'encontre de la personne à laquelle le bénéficiaire a consenti une donation ou un legs.

(110) Aucune récupération n'est exercée à l'encontre de la personne à laquelle le bénéficiaire a consenti une donation ou un legs, quelle que soit la qualité de cette personne.

(111) Une récupération peut également être exercée à l'encontre de la personne à laquelle le bénéficiaire a consenti une donation ou un legs. — `service-public.fr`

(112) Ces sommes peuvent également être récupérées à l'encontre de la personne à laquelle le bénéficiaire a consenti une donation ou un legs.

(113) Les sommes peuvent également être récupérées à l'encontre de la personne à laquelle le bénéficiaire a consenti une donation ou un legs. — `service-public.fr`

L'ensemble de ces phrases similaires, mais non identiques, illustrent également la nature instable et fragmentée des sites web. Chaque page appartient au même site, mais est rédigée de manière indépendante. Ainsi, des phrases similaires peuvent se retrouver

sur différentes pages, bien que celles-ci ne soient pas nécessairement écrites à la même date ni par le même auteur. Elles peuvent donc faire l'objet de modifications.

Dans d'autres exemples, les modifications sont dues à une différence de législation ou de signification.

(114) If the change is about your limited capability for work and you could not reasonably have known that you should have reported the change earlier.

(115) If the change is about your disability condition and you could not reasonably have known you should have reported it earlier. — `gov.uk`

Dans l'exemple ci-dessus, les termes *limited capability for work* et *disability condition* ne se rapportent probablement pas à la même définition et il est donc utile de décliner la phrase suivant le profil de l'usager.

5.4.1.4 Répétitions liées au numérique

Le site web est un objet numérique, qui a la particularité d'être délinéarisé et ponctué de technosignes (voir section 3.2). Ces particularités sont présentes dans la phraséologie, et notamment dans certaines phrases répétées. On observe alors un phénomène de récursivité sur les sites web, qui copie du texte et le republie tel quel (Pecman, 2018).

(116) L'usage de ce service est personnel et son accès est sécurisé. 49 occurrences — `service-public.fr`

(117) Make a complaint All complaints are confidential - 8 occurrences — `gov.uk`

Ces exemples montrent la récursivité liée aux services numériques proposés par les administrations. Les services étant proposés et accessibles par plusieurs pages du site, on retrouve les mêmes phrases sur l'ensemble des pages concernées.

(118) votre Caf étudiera bien votre droit à la PreParE correspondant à votre situation. 8 occurrences — `caf.fr`

(119) You'll be contacted if your application is complex and will take longer 4 occurrences — `gov.uk`

Ces exemples montrent la récursivité, mais également le caractère asynchrone des sites web de l'accès aux droits. En effet, ces phrases sont répétées sur plusieurs pages, car les démarches réalisées en ligne, en plus d'être dématérialisées, sont asynchrones. Cela signifie qu'il existe un délai entre la réception des démarches, questions, documents, et leur traitement. Ce délai existait évidemment auparavant, notamment par l'envoi papier, mais ne se manifestait pas de la même manière.

La difficulté de naviguer et de se repérer sur le site est une caractéristique que l'on retrouve dans les répétitions.

(120) You can get help with completing the online form if you : do not feel confident using a computer or mobile device do not have internet access - 19 occurrences – `gov.uk`

(121) La nature de ces justificatifs est précisée au moment d'effectuer la démarche en ligne - 50 occurrences — `service-public.fr`

(122) Vous pouvez signaler votre changement de domicile en utilisant le service de changement d'adresse en ligne — 48 occurrences — `service-public.fr`

Les phrases répétées proposées ci-dessus illustrent la nécessité d'orienter l'usager sur les diverses pages qu'il visite. Le caractère répété des phrases permet de créer des repères pour l'usager.

Enfin, la construction des sites web avec une multitude de pages organisées en arborescence implique que les titres et éléments de repérage sont répétés sur de nombreuses pages. Les exemples ci-dessous illustrent deux types de titres présents en corpus.

(123) How and when your benefits are paid - 80 occurrences — `gov.uk`

(124) Je m'actualise avec l'actualisation enrichie - 55 occurrences — `pole-emploi.fr`

Les titres sont des éléments de repérage, mais aussi très souvent des hyperliens, qui doivent être cohérents sur l'ensemble des pages.

5.4.1.5 Intertextualité au niveau des pages pour rediriger l'utilisateur

La nature fragmentée du discours de l'accès aux droits numérique, combinée à sa structure en arborescence, entraîne la présence d'une multitude de liens sur les sites web. Ces liens ont une fonction intertextuelle, afin de rediriger l'utilisateur entre les pages. Généralement, l'intertextualité désigne en linguistique la présence dans un texte des traces d'un autre texte, c'est-à-dire les éléments issus d'autres textes (par exemple, par la présence de citations, de références bibliographiques, ou d'informations issues de sources citées). Dans le discours numérique, nous pouvons rencontrer également ce type d'intertextualité. Néanmoins, il y existe une variante de l'intertextualité spécifique au discours numérique, dans la mesure où la structure du site web consiste en plusieurs pages dont chacune peut être considérée comme un texte (cf. hypertexte), et dont chacune est à la fois ou en partie indépendante et dépendante des autres. La spécificité de cette structure explique d'ailleurs la difficulté de référencer une page web, dont seule l'URL

fournit une adresse claire. Voici un exemple de ce type d'intertextualité se manifestant par des liens hypertextes :

(125) En savoir plus Agriculture, secteur animalier De l'agriculture à nos animaux de compagnie, ce secteur d'activité englobe une multitude de métiers. - 135 occurrences - `service-public.fr`

(126) Useful links for school governors Clerking service for school governors - 3 occurrences - `york.gov.uk`

Dans les deux exemples proposés ci-dessus, un marqueur linguistique indique le changement de page ou de thème. En français, il s'agit de **En savoir plus**, qui est une routine discursive des sites web afin de signaler un changement de domaine ou de thème. L'équivalent en anglais est la routine discursive *Useful links*. Elle signale également le changement de thème ou le caractère additionnel des liens.

L'importance de l'interactivité et des liens dans le discours numérique appelle à une analyse plus fine de leurs caractéristiques et fonctions que nous présentons dans le chapitre suivant consacré aux caractéristiques discursives des sites web.

5.4.1.6 Synthèse

Les phrases répétées sont utilisées sur les sites web pour répondre à des besoins précis liés au domaine. Elles sont conditionnées par la nature du site web : numérique, fragmenté, instable.

Tout d'abord, nous avons exploré les phrases répétées liées aux éléments définitoires. Ces éléments sont issus du discours juridique ou administratif, et vulgarisé pour les usagers. Les termes jouent un rôle important dans ce discours, ces éléments étant donc répétés sur plusieurs pages afin de faciliter la compréhension et de pallier la fragmentation du discours sur les différentes pages. Nous avons également noté des reformulations opérées sur ces phrases répétées. Ces reformulations peuvent être attribuées à plusieurs facteurs. La répétition des phrases est une caractéristique saillante du discours numérique. La nature modulable des supports de rédaction, ainsi que l'exploration non linéaire possible des pages, rend la rédaction moins souple. Il est ainsi plus fréquent de retrouver les mêmes textes sur plusieurs pages ou sites. Toutefois, les pages ne sont pas toujours liées entre elles, et les mises à jour étant faites progressivement, certaines phrases peuvent subir des modifications qui ne sont pas répercutées sur l'ensemble du site. D'autres changements mineurs de ces phrases répétées sont expliqués par la multitude des profils pris en compte sur les sites.

Les répétitions sont souvent liées au caractère numérique du discours. Elles sont des éléments de repérage pour l'usager, dans les menus et les liens par exemple. Enfin, elles

permettent d'exprimer l'intertextualité du discours, mais aussi de rediriger l'usager.

Dans les deux langues, les répétitions liées au numérique ont une place importante et sont caractéristiques du corpus. Toutefois, le corpus français, composé de sites ayant une arborescence plus complexe, présente davantage de répétitions. Le discours est modifié et façonné par les contraintes du site web.

Même si elles visent à compenser la délinéarisation du discours numérique, les répétitions, en raison de leur fréquence élevée, peuvent devenir très gênantes pour l'utilisateur car elles rendent la lecture chronophage et lassante.

Chapitre 6

Caractérisation discursive des sites web de l'accès aux droits

6.1 Les sites web de l'accès aux droits en tant que genre discursif

D'après les études réalisées dans les chapitres précédents et notamment le chapitre 5, nous pouvons avancer que le site web de l'accès aux droits est un genre discursif spécifique. Ces dernières années, les rituels de communication entre les usagers et les organismes d'accès aux droits se sont stabilisés et cristallisés dans le site web de l'accès aux droits. Ce dernier présente des régularités discursives, mais il est également « relié à une pratique sociale » (Delavigne, 2022) partagée par les citoyens, comme le montrent les multiples espaces en ligne où les usagers échangent sur l'accès à ces sites, la compréhension des contenus et le fonctionnement des démarches en ligne. Ces sites partagent tous le même objectif : informer les usagers de leurs droits et leur permettre de faire des démarches administratives de façon indépendante, c'est-à-dire sans communication directe avec un agent humain.

Dans ce dernier chapitre, nous proposons d'analyser le site web d'accès aux droits en tant que genre discursif, en nous intéressant notamment à sa structure, aux hyperliens et à l'interactivité qui le caractérisent. Nous avons également étudié l'arborescence du site web et nous avons formulé des recommandations pour les recherches en corpus des discours administratifs numériques à destination des chercheurs (linguistes et terminologues notamment), ainsi que des recommandations destinées aux experts du domaine en charge de la création et de la maintenance des informations sur ce type de site.

6.1.1 Structure de la page

D'après Adam (1997), les genres déterminent en grande partie la structure des textes. C'est le cas du site web d'accès aux droits, qui possède une structure stable issue du support numérique et qui pose de nombreuses contraintes. Nous avons étudié cette structure lors de la création du corpus parcours et du schéma d'annotation manuelle qui a été réalisé à partir des codes HTML de la page, comme présenté dans la section 4.5.1. L'ensemble de ces caractéristiques forment le genre discursif de l'accès aux droits. Pour présenter ces caractéristiques, notamment en lien avec la structure des sites web et les fonctions discursives des différentes parties du texte, nous avons réalisé des schémas d'une page type à partir de nos observations et du code HTML.

6.1.1.1 Structure de la page française

Le premier schéma que nous présentons est celui d'une page issue du corpus parcours français, récolté sur `service-public.fr`.

La Figure 6.1 ci-dessous est une représentation de la structure d'une page.

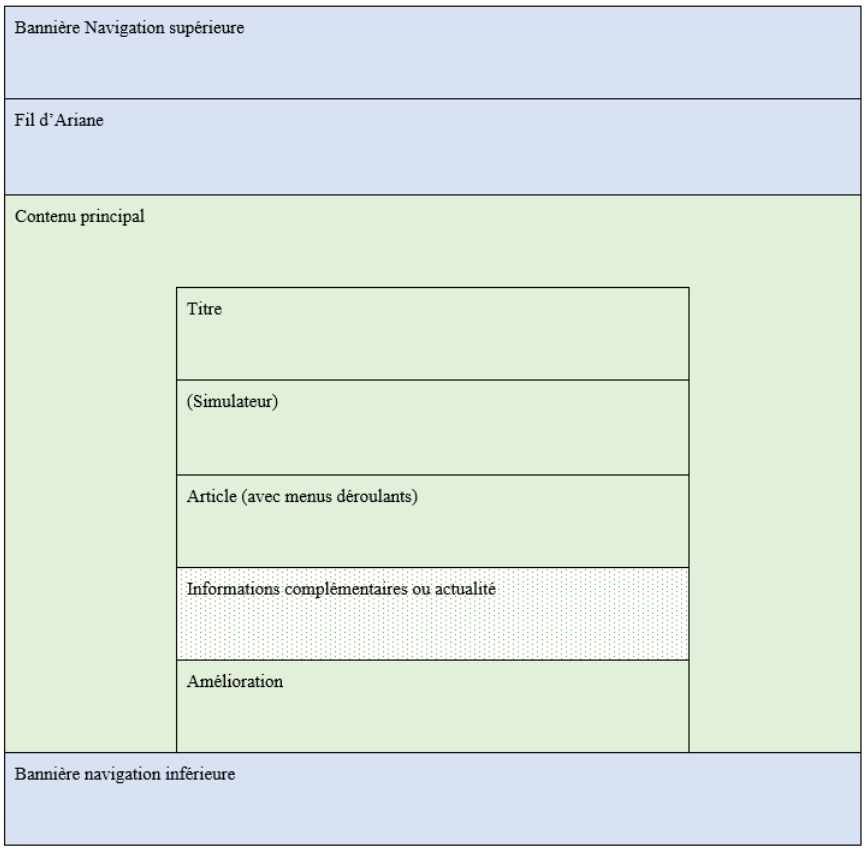


FIGURE 6.1 – Représentation graphique de la structure d'une page informative du corpus parcours — français

Nous avons représenté les sections réservées à la *navigation* en bleu et les sections réservées au *contenu* en vert. Nous avons indiqué une section hybride : il s'agit de l'avant-dernière section du *contenu* « information complémentaire ou actualité », qui peut faire partie du *contenu* ou de la *navigation*. Sur ce schéma, les sections réservées à la *navigation* sont en périphérie (en haut et en bas), tandis que le centre de la page est dédié au *contenu*. Cette organisation est classique des sites web. Le *contenu* est généralement composé d'un titre, d'un simulateur (qui n'est pas présent sur toutes les pages), puis de l'article qui se présente avec des menus déroulants. Enfin, certaines rubriques comportent des redirections vers des pages liées au contenu principal ou à l'actualité. Le *contenu* se termine toujours par une rubrique interactive portant sur l'amélioration du discours. Cette rubrique demande à l'usager de noter l'utilité de la page et d'envoyer ses retours. L'utilisateur remplit un formulaire et peut même suggérer des modifications à un rédacteur. La structure dégagée semble reprendre des codes des sites web institutionnels, ainsi que des sites web informatifs. Toutefois, nous n'avons pas trouvé de recherche exploratoire formalisant cette structure des sites web, qui nous semble pourtant caractéristique de ce genre discursif. La Figure 6.2 représente la structure d'une page de menu du corpus parcours, qui est différente d'une page informative.

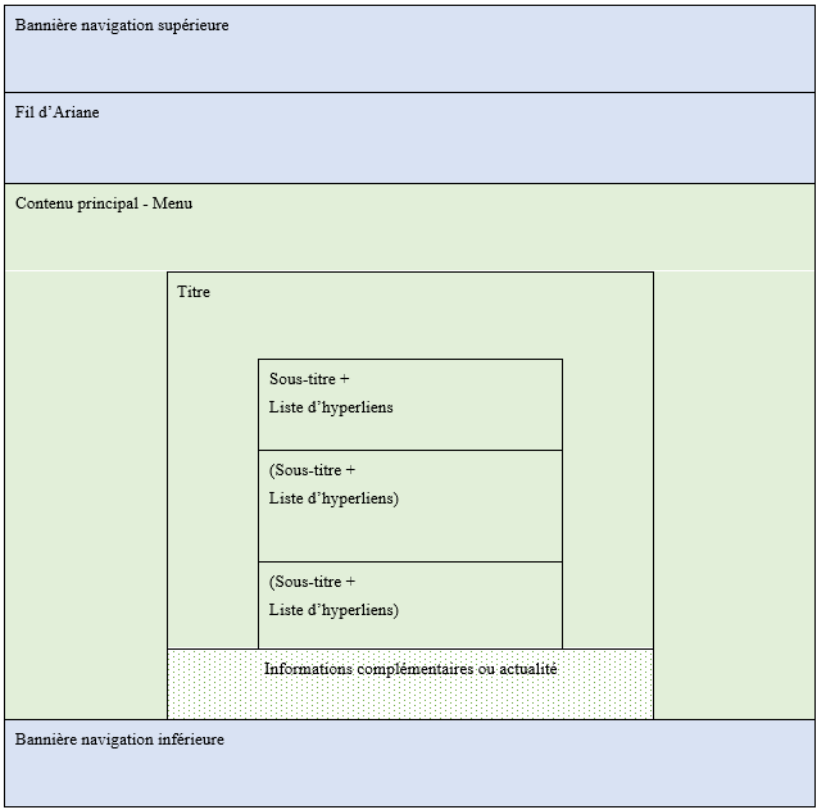


FIGURE 6.2 – Représentation graphique de la structure d'une page de menu du corpus parcours — français

La structure de la page 6.2 est similaire à celle d'une page d'information (voir Figure 6.1), mais à la place du *contenu*, on retrouve une liste de liens. Nous pouvons donc faire plusieurs observations. D'abord ce genre discursif se caractérise par l'existence de deux types de pages : pages de menus et pages informatives. Les sections *navigation* sont une constante dans ce genre discursif, puisqu'elles permettent notamment de naviguer sur le site web. Nous remarquons également le maintien de la rubrique « informations complémentaires ou actualité », qui est présente sur la plupart des pages du site. Les Figures 6.3 et 6.4 présentent deux exemples de cette rubrique.



FIGURE 6.3 – Rubrique info complémentaire de `service-public.fr`

source : `www.service-public.fr/particuliers/vosdroits/N19808`, consulté le 8 juillet 2024

Ce premier exemple présenté ci-dessus (Figure 6.3) montre une rubrique proposant des liens afin d'accéder à des informations complémentaires sur le sujet abordé. La rubrique est déroulante, la petite flèche en haut à droite indique qu'il est possible de replier la rubrique afin de ne laisser que « Voir aussi ».



FIGURE 6.4 – Rubrique actualité de `service-public.fr`

source : `www.service-public.fr/particuliers/vosdroits/N19808`, consulté le 8 juillet 2024

Le second exemple présenté ci-dessus (6.4) est une rubrique d'actualité, en une pu-

blicité ou une annonce pour une fiche disponible sur `vie-publique.fr`. Cette fiche présente du contenu informatif sur la protection sociale.

Cette structure de page semble être similaire aux structures des sites web institutionnels. Il y a évidemment des corrélations entre le genre discursif des sites web de l'accès aux droits et les autres sites web institutionnels. Néanmoins, les recherches sur la structure de ces sites semblent inexistantes.

6.1.1.2 Structure de la page britannique

Nous présentons maintenant la structure d'une page issue du corpus parcours anglais. La Figure 6.5 représente la structure classique d'une page d'un site d'accès aux droits.

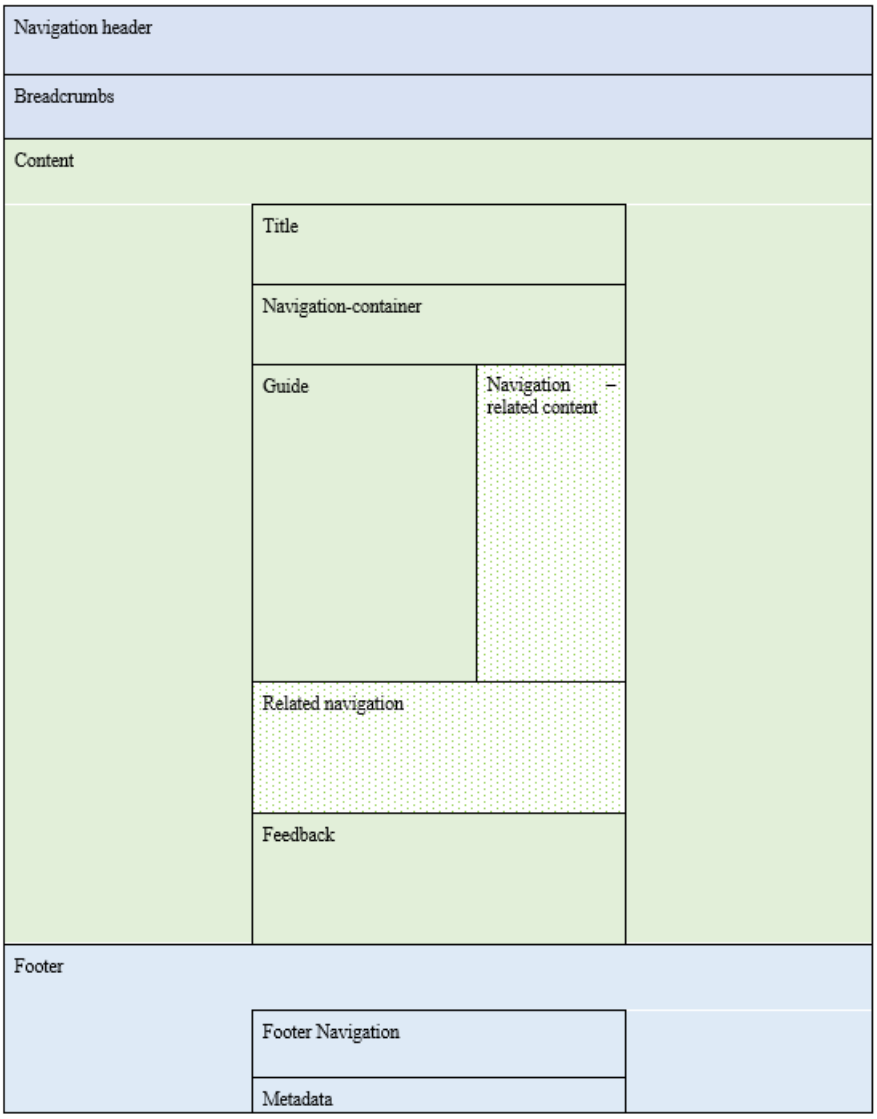


FIGURE 6.5 – Représentation graphique de la structure d'une page du corpus parcours — anglais

Cette structure présente les mêmes caractéristiques que la structure de la page française. On retrouve, en bleu, les éléments de *navigation* en haut et en bas de page, tandis que le *contenu* est au milieu de la page en vert. Deux types de rubriques sont hybrides, c'est-à-dire qu'elles appartiennent en même temps à la fonction *contenu* et à la fonction *navigation* : « Navigation — related content » et « Related navigation ». Ces deux rubriques sont similaires, mais présentent également des différences. La rubrique « Navigation — related content » propose des liens redirigeant directement sur une page informative tandis que la seconde rubrique (« Navigation related ») propose des liens redirigeant vers des pages menus. Il s'agit notamment, pour la rubrique « Navigation — related content », d'un menu permettant d'accéder à des pages sur le même thème, présenté dans la Figure 6.6 ci-dessous.

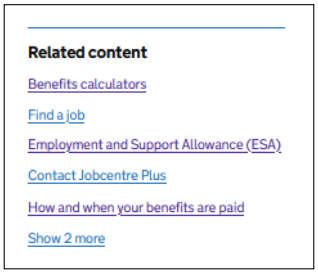


FIGURE 6.6 – Rubrique *related content* d'un site web d'accès aux droits britannique

source : `www.gov.uk/jobseekers-allowance`

La rubrique présentée dans la Figure 6.6 se trouve à droite du *contenu*. Elle propose des contenus en lien avec le thème de la page. Ces liens utilisent les couleurs conventionnelles des hyperliens : le bleu pour les hyperliens sur lequel l'utilisateur n'a pas cliqué et le violet pour les liens sur lesquels l'utilisateur a cliqué, qui sont depuis les débuts d'internet des « signes typographiques à part entière » (Mourad, 2006, p. 20). La Figure 6.7 ci-dessous est un exemple de rubrique « Navigation related ».

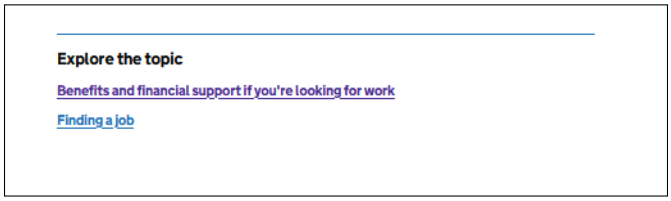


FIGURE 6.7 – Rubrique *explore* d'un site web d'accès aux droits britannique

source : `www.gov.uk/jobseekers-allowance`

La Figure 6.7 ci-dessus est un exemple de navigation liée au thème de la page. Cette rubrique propose également des pages en lien avec le sujet principal.

Nous retrouvons également deux différents types de structures qui correspondent à des menus, présentées dans les Figures 6.8 et 6.9 ci-dessous. La Figure 6.8 présente un menu dans le premier niveau de l'arborescence, il s'agit du premier menu proposé après avoir choisi un thème. Le menu est organisé en petits modules, qui présentent un hyperlien et un texte de présentation.

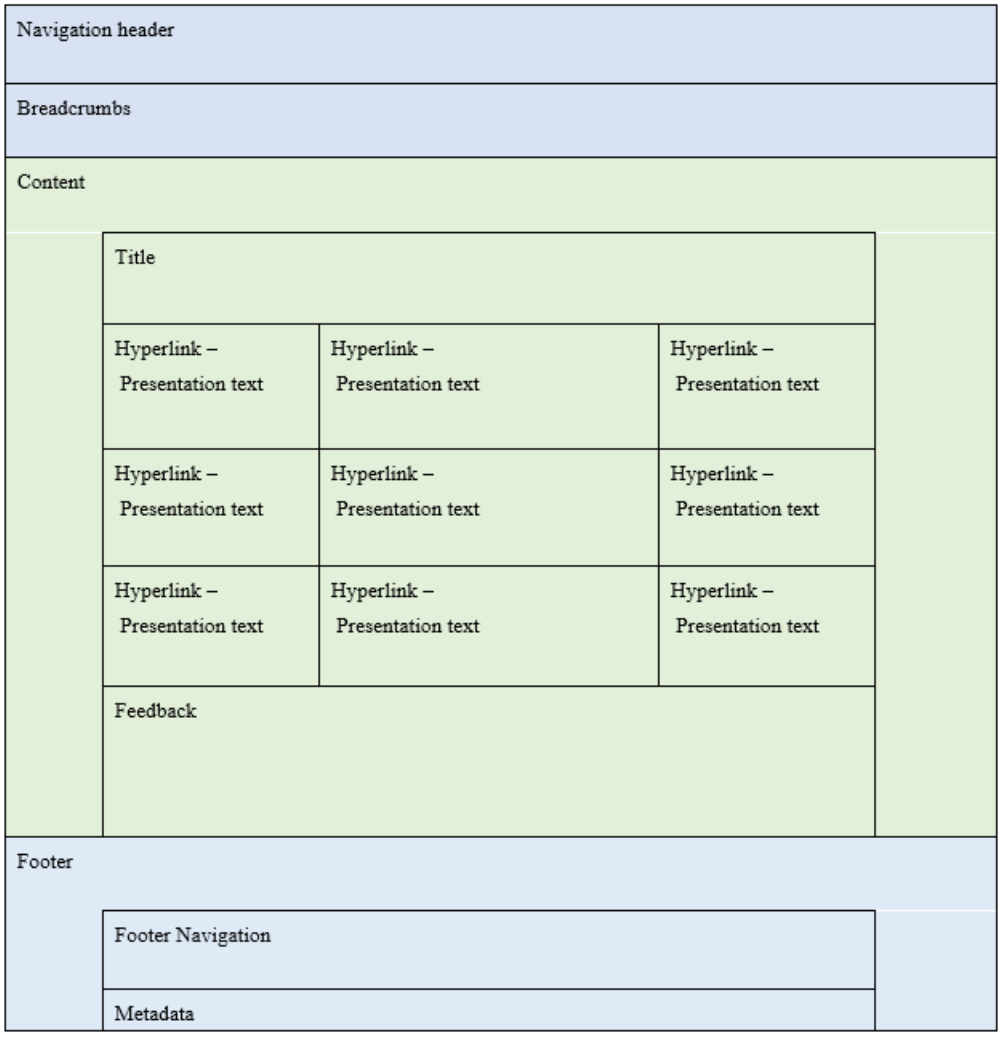


FIGURE 6.8 – Représentation graphique de la structure d'une page menu du corpus parcours — anglais

La Figure 6.9 présente un menu dans le deuxième niveau de l'arborescence, auquel l'utilisateur accède en cliquant sur l'hyperlien présenté dans la Figure 6.8.

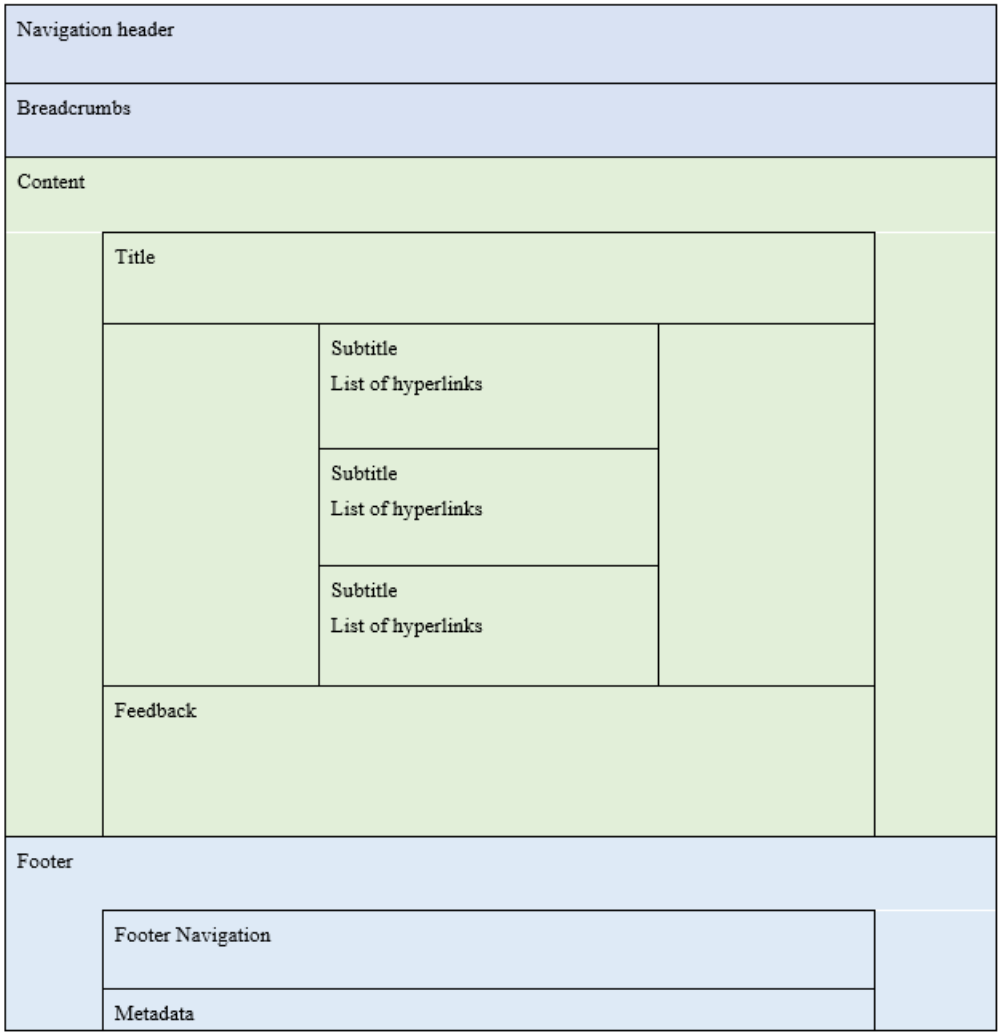


FIGURE 6.9 – Représentation graphique de la structure d'une page menu du corpus parcours — anglais

Les Figures 6.8 et 6.9 présentent la structure de deux pages de menus différentes. Les menus sont présentés différemment selon leur place dans l'arborescence. Cette différence de structure dans la page permet à l'utilisateur de se repérer plus facilement et de savoir s'il va accéder à un nouveau menu ou à une page informative.

6.1.1.3 Analyse comparative

La formalisation de la structure des pages de nos parcours nous permet de mettre en évidence les similarités dans la structure du discours.

Tout d'abord, nous retrouvons des caractéristiques communes des sites web, à savoir les bannières de navigation en bas et en haut de la page, qui permettent à l'usager de se déplacer de façon non linéaire dans l'arborescence du site. Ensuite, le fil d'Ariane ou *breadcrumbs* semble être une caractéristique du site web de l'accès aux droits, puis-

qu'il est proposé par les deux sites. Il permet de retracer le chemin fait par l'usager. Le glossaire du site *Système de Design de l'État*[1] propose une définition du fil d'Ariane :

> Le fil d'Ariane est un système de navigation secondaire qui permet à l'utilisateur de se situer sur le site qu'il consulte. (Service d'Information du gouvernement, 2024b)

Il est composé de plusieurs éléments obligatoires.

> Dans l'ordre, il se compose des éléments suivants :
> — un lien menant à la racine du site (page d'accueil) — obligatoire.
> — des liens vers les pages séparant la racine du site de la page courante — obligatoire si la hiérarchie du site comporte plus d'un niveau.
> — la page courante, seul élément non cliquable — obligatoire.

(Service d'Information du gouvernement, 2024b)

Cet article décrivant le Fil d'Ariane montre qu'il joue un rôle important sur le site.

Le fil d'Ariane ou *breadcrumbs* fait également l'objet d'une page et d'une définition similaire sur le site *Design System* du gouvernement britannique[2].

> *The breadcrumbs component helps users to understand where they are within a website's structure and move between levels.* (Gov.uk Design System team, 2024a)

On remarque dans ces deux définitions que le fil d'Ariane est construit pour l'utilisateur : il joue un rôle d'aide à la navigation et aide à la cohésion du discours, comme nous le verrons dans la section 6.2.4.

D'autres éléments semblent réguliers et stables, comme les éléments hybrides permettant d'accéder aux pages liées au contenu principal. Ces listes de liens, permettant d'accéder aux autres pages sans revenir en arrière, sont également des éléments de navigation importants.

Dans les deux corpus, nous retrouvons une section à la fonction d'amélioration, qui permet à l'utilisateur de faire des retours sur la qualité du discours et des informations. Il s'agit d'un des rares éléments réellement interactifs, c'est-à-dire qui permet de communiquer avec les rédacteurs du site. Dans l'ensemble des autres espaces où l'utilisateur peut faire des requêtes, les réponses sont générées par le site web et non par un humain.

Enfin, nous pouvons remarquer quelques différences dans la mise en page et dans les rubriques présentes. Le corpus britannique présente deux rubriques de *navigation* liées au contenu principal, tandis que le corpus français n'en présente qu'une. Le corpus

1. accessible sur `www.systeme-de-design.gouv.fr/composants-et-modeles/composants/fil-d-ariane/`
2. accessible sur `design-system.service.gov.uk/components/breadcrumbs/`

parcours français propose également des rubriques déroulantes, ce qui n'est jamais le cas pour le corpus parcours britannique, comme nous le verrons dans la section 6.2.3. Le site web britannique propose deux types de pages de menus, dans le but d'aider l'usager à se repérer, tandis que le site français n'en propose qu'un. Toutefois, les Figures 6.9 et 6.2 ont une structure très similaire.

6.2 Fonctionnalité des différentes parties du discours

6.2.1 Navigation

Afin d'étudier plus en détail chacune des parties du discours, nous proposons une exploration du sous-corpus *navigation*, c'est-à-dire des segments ayant été annotés comme étant des unités discursives servant à naviguer, comme présenté dans la section 4.5.

6.2.1.1 Analyse des structures des parties navigation dans le corpus français

Nous avons dans un premier temps réalisé un schéma de la structure des éléments de navigation sur la page, à partir des Figures 6.1 et 6.2 et du code HTML. Nous proposons dans la Figure 6.10 un schéma de la structure d'un en-tête.

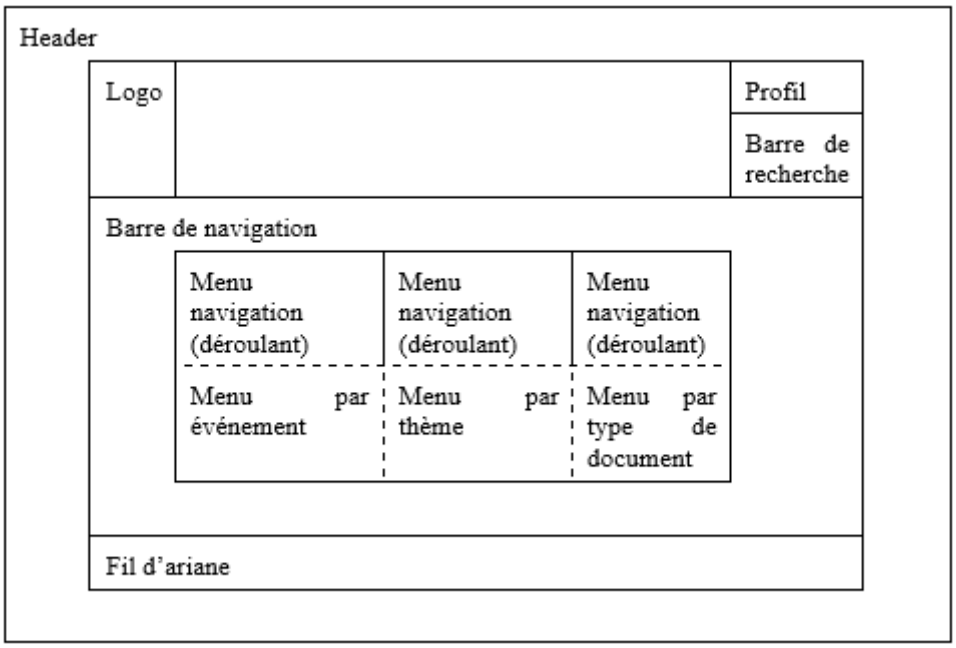


FIGURE 6.10 – Représentation de l'en-tête d'une page de site web d'accès aux droits français

La Figure 6.10 est une représentation graphique de l'en-tête d'un site web d'accès aux droits. Sa définition est la suivante :

> L'en-tête permet aux utilisateurs d'identifier sur quel site ils se trouvent. Il peut donner accès à la recherche et à certaines pages ou fonctionnalités clés. (Service d'Information du gouvernement, 2024a)[3]

L'en-tête a donc une fonction d'identification du site, mais aussi d'accès à l'arborescence et à la structure du site, en déroulant les menus disponibles. Plus précisément, ce site propose une « En-tête avec accès rapides et moteur de recherche »(Service d'Information du gouvernement, 2024a) qui permet de mettre en avant certains liens et la barre de recherche. Sur le site `service-public.fr`, duquel est issu notre corpus parcours, on remarque que les menus proposent plusieurs types de classification : par événement, par thème et par type de document. La multiplicité de ces propositions permet aux usagers d'appréhender le discours et les informations sous divers angles, suivant leurs besoins.

Nous proposons dans la Figure 6.11 un schéma de la structure d'un pied de page.

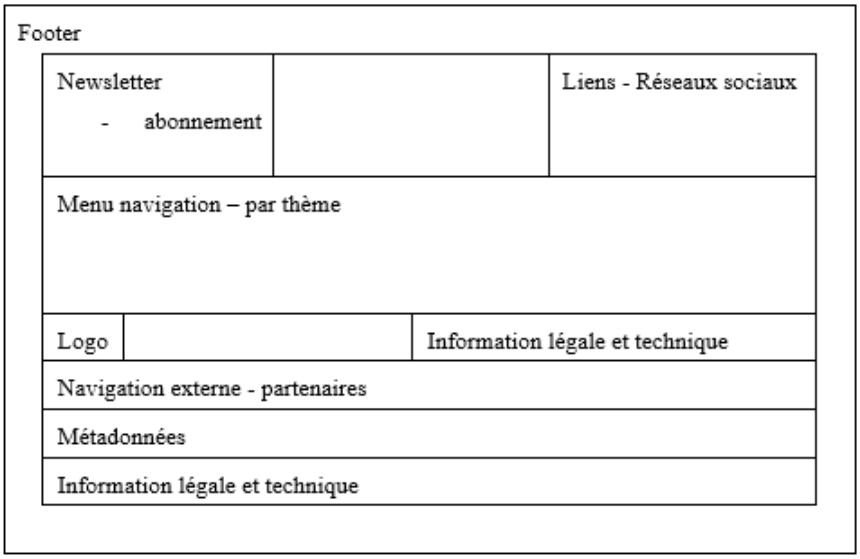


FIGURE 6.11 – Représentation du pied de page d'une page de site web d'accès aux droits français

La Figure 6.11 est une représentation graphique du pied de page d'une page issue du corpus parcours, basé sur le site `service-public.fr`. D'après le site *Système de Design* du gouvernement français, voici la définition d'un pied de page.

> Le pied de page propose des éléments d'information sur le site et une navigation secondaire pour l'utilisateur afin qu'il poursuive son parcours. Il est

3. www.systeme-de-design.gouv.fr/composants-et-modeles/composants/en-tete/

modulable selon les besoins et les exigences du site. (Service d'Information du gouvernement, 2024c)[4]

Alors que l'en-tête a un rôle d'identification initial, le pied de page a pour but d'aider l'utilisateur à poursuivre son parcours de navigation. Dans le cas des sites d'accès aux droits, le site propose un « pied de page complet », qui inclut une « liste de liens » (Service d'Information du gouvernement, 2024c). Cette liste de liens est en fait le menu par thème proposé dans l'en-tête.

6.2.1.2 Analyse des structures des parties navigation dans le corpus anglais

Nous avons procédé à la même structuration pour le sous-corpus britannique. Nous avons à nouveau identifié deux éléments issus des parties *navigation*. Le premier élément identifié est un en-tête. Voici la définition d'un *header* proposé par le site *Design System* du gouvernement britannique.

> *The GOV.UK header shows users that they are on GOV.UK and which service they are using.* (Gov.uk Design System team, 2024c)

La Figure 6.12 présentée ci-dessous est une représentation de l'ensemble des éléments de l'en-tête présents sur les pages du corpus parcours, issues du site `gov.uk`.

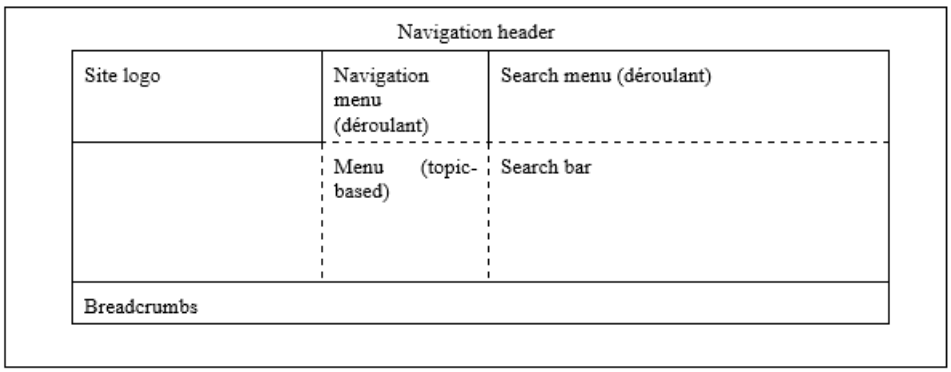


FIGURE 6.12 – Représentation de l'en-tête d'une page de site web d'accès aux droits anglais

L'en-tête a donc pour but d'aider les utilisateurs à se repérer. Dans notre corpus, nous avons plus précisément des « Header with service name and navigation » (Gov.uk Design System team, 2024c), qui indiquent qu'en plus d'inclure les informations permettant à l'utilisateur de se repérer, le service inclut des éléments permettant la navigation. Ces éléments de navigation sont un menu déroulant par thème ainsi qu'une barre de recherche déroulante. On retrouve à nouveau le fil d'Ariane ou *breadcrumbs* que nous avons défini

4. www.systeme-de-design.gouv.fr/composants-et-modeles/composants/pied-de-page/

dans la section précédente et qui aide l'utilisateur à se repérer sur le site et à retracer son parcours. Les éléments de navigation présents dans cet en-tête ne sont pas nombreux, mais permettent à l'utilisateur d'identifier le site et les services proposés. Il peut ainsi se déplacer dans l'arborescence du site en choisissant un thème ou en entrant un mot-clé dans la barre de recherche.

Le deuxième élément dédié à la *navigation* est le pied de page (Figure 6.13). Sur le site *Sytem design* du gouvernement britannique, nous trouvons la définition suivante du *footer*.

> *The footer provides copyright, licensing and other information about your service.* (Gov.uk Design System team, 2024b)

D'après ce site, les pieds de page dans notre corpus sont des « Footer with links and secondary navigation ». Le pied de page n'est donc pas un élément de navigation par défaut, puisqu'il est plutôt là pour des raisons techniques (copyright) et de reconnaissance du service. Toutefois, pour les sites web de l'accès aux droits, il semble que les webmasters ajoutent des éléments de navigation afin d'aider les utilisateurs à naviguer et se repérer sur le site. La répétition des menus par thème dans l'en-tête, puis dans le pied de page, a une fonction cohésive. Elle permet d'assurer la cohérence entre les contenus dispersés sur diverses pages et de présenter à l'utilisateur une arborescence du site stable. De plus, le menu présent dans le pied de page n'est pas déroulant, ce qui facilite son accès.

Nous proposons dans la Figure 6.13 un schéma de la structure d'un pied de page.

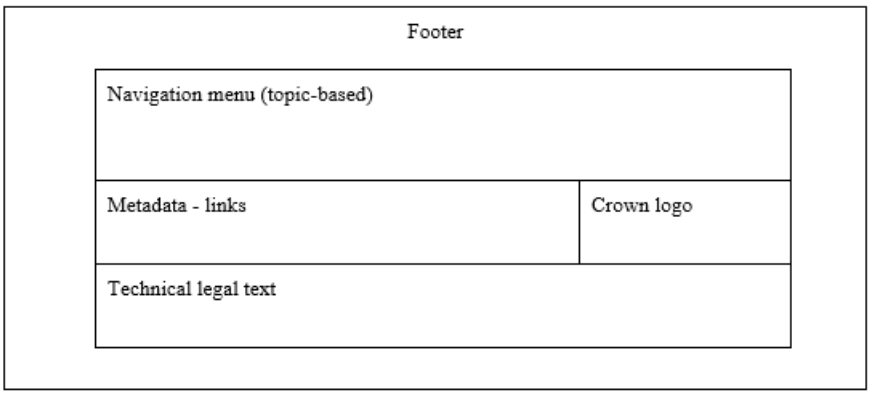


FIGURE 6.13 – Représentation de l'en-tête d'une page de site web d'accès aux droits anglais

source : `www.gov.uk/jobseekers-allowance`

6.2.1.3 Analyse comparative

Les éléments de navigation ont donc plusieurs rôles : permettre à l'utilisateur de se repérer et de naviguer sur le site afin de trouver l'information recherchée. Les éléments de navigation proposés par le site sont détaillés et complets, offrant divers moyens d'accéder aux informations : différents types de menus, une barre de recherche et la répétition de certains menus. Cette répétition semble avoir pour objectif de compenser la délinéarisation du discours liée au mode de fonctionnement du site web et aux éléments interactifs.

L'en-tête britannique est plus sobre que l'en-tête français, qui propose un foisonnement de menus et de liens, comme le montrent les captures d'écran ci-dessous.

La première Figure ci-dessous (6.14) est l'en-tête du site `service-public.fr`.



FIGURE 6.14 – Capture d'écran de l'en-tête d'une page du site `service-public.fr`

source : `service-public.fr`, consulté le 10 juillet 2024

La deuxième Figure (6.15) est l'en-tête du site `gov.uk`.



FIGURE 6.15 – Capture d'écran de l'en-tête d'une page du site `gov.uk`

source : `www.gov.uk/jobseekers-allowance`, consulté le 10 juillet 2024

Ces deux captures d'écran, ainsi que les représentations graphiques présentées dans les Figures 6.12 et 6.10 montrent la différence entre les deux systèmes français et britannique. En France, le système est complexe linguistiquement et légalement. Le domaine adopte une stratégie de foisonnement de l'information pour permettre son accessibilité, avec la multiplication de sites, des termes et des moyens d'accès à l'information. Ces caractéristiques sont visibles dans la Figure 6.14, qui montre une barre de navigation conséquente, avec trois menus déroulants, plusieurs parties du site accessibles via des liens, mais aussi une icône pour se connecter à un espace en ligne.

À l'inverse, l'en-tête britannique est plus sobre, comme le montre la Figure 6.15. En effet, il n'y a que trois éléments présents dans l'en-tête britannique. Cette sobriété reflète

le système britannique qui propose un service numérique unique et met en œuvre une accessibilité par la simplification et la réduction des démarches, des dispositifs, des sites, etc.

La terminologie choisie est également différente, pour indiquer le menu déroulant par exemple, le site britannique choisit de décrire le type d'objet utilisé (un menu), tandis que le site français a recours à une description du contenu et non de l'objet (« fiches pratiques par thème »). Les deux ont recours au même signe typographique pour indiquer le menu déroulant : la flèche vers le bas.

Les pieds de page présentent, pour leur part, de nombreuses similitudes, comme le montrent les deux Figures ci-dessous. La première Figure (6.16) est le pied de page du site `service-public.fr`.



FIGURE 6.16 – Capture d'écran du pied de page d'une page du site `service-public.fr`

source : `service-public.fr`, consulté le 10 juillet 2024

La deuxième Figure (6.17) est le pied de page du site `gov.uk`.

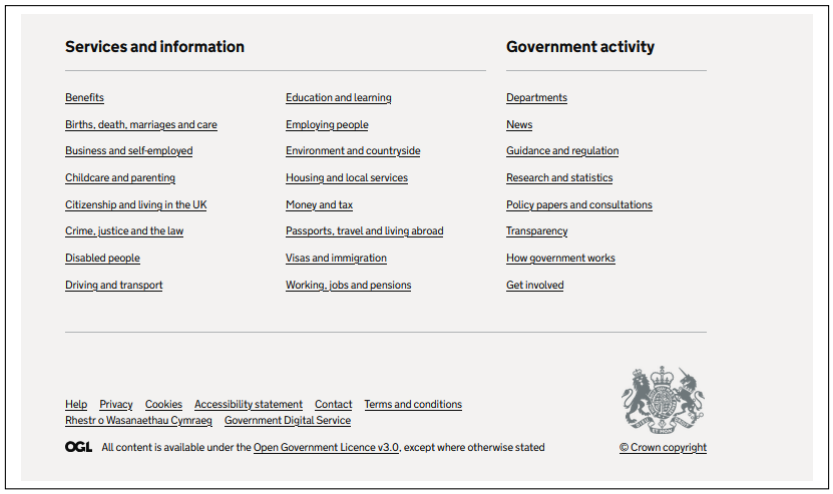


FIGURE 6.17 – Capture d'écran du pied de page d'une page du site `gov.uk`

source : `www.gov.uk/jobseekers-allowance`, consulté le 10 juillet 2024

Ces deux captures d'écran, ainsi que les représentations graphiques présentées dans les Figures 6.13 et 6.11 présentent des similitudes. Dans les deux pieds de page, on retrouve des informations techniques et légales, notamment sur les licences. Les deux pieds de page présentent également une répétition du menu par thème proposé dans l'en-tête, ainsi que des liens additionnels permettant de naviguer. Dans la partie inférieure, les deux pieds de page présentent des liens correspondant aux métadonnées.

La France présente un pied de page plus complexe et fourni, qui indique une présence sur les réseaux sociaux, propose une newsletter à laquelle il est possible de s'abonner, et inclut une liste de services gouvernementaux complémentaires (`vie-publique.fr` ou encore `entreprendre.service-public.fr`). Le pied de page britannique est, quant à lui, plus sobre et présente moins d'informations. Cela peut s'expliquer, encore une fois, par la multiplicité des sources d'information en France (comme la newsletter du site ou encore les publications sur les réseaux sociaux) tandis que le gouvernement britannique cherche à centraliser ses informations.

En conclusion, ces deux éléments de navigation (l'en-tête et le pied de page) constituent des éléments discursifs stables dans les deux langues étant donné qu'ils se répètent sur chaque page et qu'ils ont pour objectif d'aider l'utilisateur à se repérer. Ils reflètent, dans chacune des langues, les caractéristiques du domaine telles que nous les avons décrites, sur le plan sociétal, politique et juridique (dans le chapitre 1) et sur le plan linguistique (dans le 5).

6.2.2 Liens

Pour étudier les liens, nous avons mis en place et annoté un petit corpus spécialisé dans chaque langue. Ces corpus sont créés à partir de trois scénarios de navigation et présentés dans la section 4.5.1 du chapitre 4. À partir de l'annotation de ces corpus, nous avons pu extraire des sous-corpus dédiés aux liens, et notamment aux liens que nous appelons *internes* et qui sont les liens qui redirigent sur les pages du site sur lequel se trouve l'utilisateur. Il s'agit d'une large majorité des liens. Les liens externes représentent systématiquement moins de 5 % des liens recensés en corpus. Nous souhaitons concentrer notre étude sur les hyperliens internes, afin d'identifier s'il existe des caractéristiques spécifiques aux hyperliens dans ce genre discursif. Nous compléterons cette analyse discursive du discours d'escorte des liens par une étude de l'arborescence dans la section 6.3 de ce chapitre. Pour rappel, le discours d'escorte désigne le texte qui compose l'URL, c'est-à-dire l'extrait de texte cliquable (voir section 3.2.3).

6.2.2.1 Analyse des liens en corpus français

Dans un premier temps, nous avons réalisé une analyse croisée des liens en corpus afin d'identifier des caractéristiques saillantes au niveau statistique dans l'utilisation des liens dans le genre discursif français. Pour ce faire, nous avons comparé le corpus parcours aux sous-corpus suivants :

— l'ensemble des segments ayant la fonction *hyperlien interne* dans le sous-corpus *navigation*,
— l'ensemble des segments ayant la fonction *hyperlien interne* dans le sous-corpus *contenu*,
— l'ensemble du sous-corpus *navigation*,
— l'ensemble du sous-corpus *contenu*.

Rappelons la portée des mouvements rhétoriques *contenu* et *navigation* et des deux types de liens :

— *contenu* :
  Il s'agit d'une partie contenant les informations principales de la page, précédemment identifiée pour former le premier corpus.
— *navigation* :
  Il s'agit de parties du texte qui ont pour fonction principale d'aider l'utilisateur à naviguer sur le site.
— *hyperlien interne* :
  Il s'agit de liens redirigeant sur le site sur lequel l'utilisateur se trouve déjà.
— *hyperlien externe* :

Il s'agit d'un lien redirigeant l'utilisateur sur un site externe.

Rappelons que, comme montré dans la Figure 4.16, les liens internes ont ensuite été divisés en deux catégories : *page*, s'ils redirigent vers une autre section de la même page, puis *site*, s'ils redirigent vers une autre page du site.

Ainsi, la construction de ces 4 sous corpus à partir du corpus parcours entier nous permet d'examiner la place des liens suivant la fonction discursive étudiée. Nous n'avons ici sélectionné que les liens internes, sans distinction entre les liens qui redirigent sur la même page ou sur une autre page du site.

L'exploration de ces sous-corpus a permis de mettre en lumière les caractéristiques spécifiques des segments correspondant à un hyperlien et portant la fonction *interne*, par rapport aux autres segments ne comportant pas d'hyperliens. Le corpus et les sous-corpus ont été annotés automatiquement par catégories syntaxiques, nous avons donc étudié les catégories syntaxiques suivantes : les noms, les verbes et les pronoms. Nous avons également exploré les acronymes et sigles présents dans les trois corpus, car ils font partie des caractéristiques spécialisées du discours étudié, comme indiqué dans la section 5.1.4. L'objectif de cette première étude statistique est d'identifier les éventuelles caractéristiques propres aux discours d'escorte des hyperliens. Pour plus de lisibilité, nous avons divisé les statistiques en deux Tables 6.1 et 6.2.

Corpus Français			
Fonctions	Liens internes	Liens internes dans Navigation	Liens internes dans Contenu
Mots occ./%	23 171 / 33 %	14 588 / 100 %	5 847 / 14,5 %
Noms occ./%	9 483 / 41 %	7800 / 53 %	1 976 / 33,7 %
Verbes occ./%	1 759 / 7,5 %	1 234 / 8,4 %	635 / 11 %
Pronoms occ./%	286 / 1,2 %	157 / 1 %	150 / 2,5 %
Sigles occ./%	134 / 0,57 %	77 / 0,5 %	58 / 0,9 %

TABLE 6.1 – Statistiques des hyperliens présents dans le corpus parcours en français

Corpus Français			
Fonctions	Navigation	Contenu	Corpus complet
Mots occ./%	14 588 / 21 %	40 330 / 58 %	69 567 / 100 %
% de liens dans la section	100 %	14,5 %	33 %
Noms occ./%	7 800 / 53 %	12 386 / 30 %	22 430 / 32 %
Verbes occ./%	1 234 / 8,4 %	6 451 / 16 %	8 960 / 12 %
Pronoms occ./%	157 / 1 %	2 501 / 6,2 %	3 012 / 4,3 %
Sigles occ./%	77 / 0,5 %	334 / 0,8 %	435 / 0,6 %

TABLE 6.2 – Statistiques des sections *contenu* et *navigation* dans le corpus parcours en français

Les Tables 6.1 et 6.2 présentent un récapitulatif des statistiques collectées à partir de

l'annotation automatique en catégories syntaxiques de Sketch Engine. Pour chaque catégorie étudiée sont indiqués le nombre d'occurrences, puis le pourcentage correspondant. Notre première observation a porté sur le nombre de mots total de chaque sous-corpus. Comme nous pouvons le voir dans la Table 6.2, dans le corpus français, les segments ayant pour fonction la *navigation* sont tous des hyperliens. Ces hyperliens sont le plus souvent disponibles sous forme de menus comme nous pouvons le voir dans les captures d'écran ci-dessous. La première capture d'écran 6.18 est un menu déroulant issu de l'entête de la page. Ce menu permet à l'utilisateur d'accéder à l'arborescence du site. Il est placé au-dessus du contenu principal.

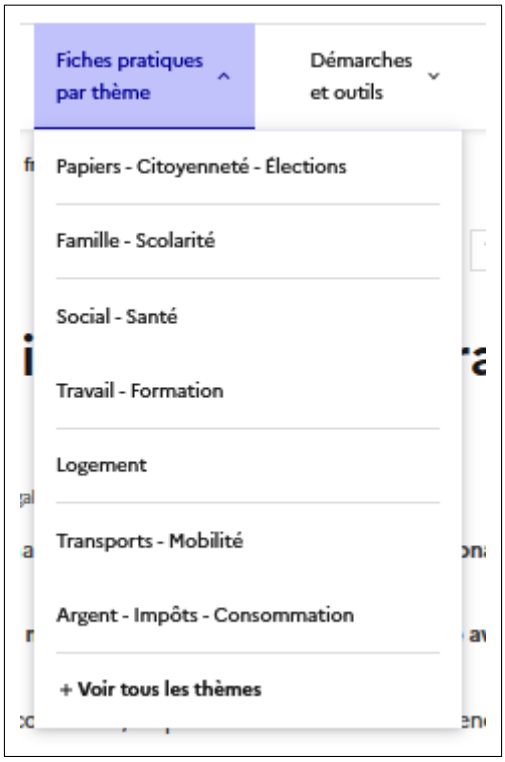


FIGURE 6.18 – Capture d'écran d'un menu déroulant

source : `www.service-public.fr/particuliers/vosdroits/F34717`, consulté le 8 juillet 2024

La deuxième capture, présentée dans la Figure 6.19, est un menu statique issu du pied de page.

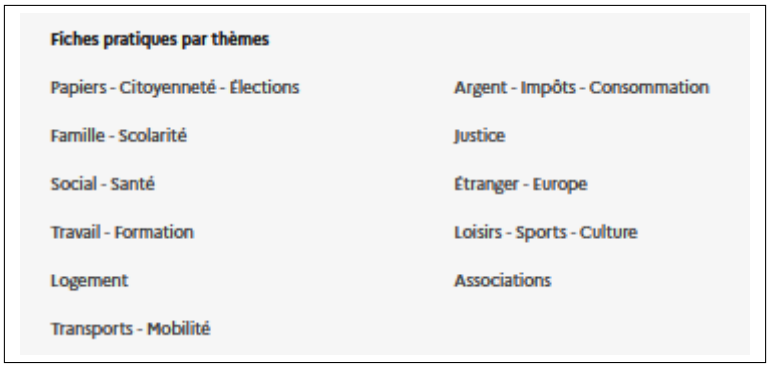


FIGURE 6.19 – Capture d'écran d'un menu statique

source : `www.service-public.fr/particuliers/vosdroits/F34717`, consulté le 8 juillet 2024

La Figure 6.19 propose le même menu que la Figure 6.18, mais celui-ci apparaît en bas de page et n'a pas d'élément interactif déroulant. Comme indiqué dans la section 5.4.1.5, cette répétition a pour objectif d'aider l'utilisateur à se repérer sur le site web. Comme nous pouvons le voir, ces deux menus sont essentiellement composés de termes relatifs aux différents thèmes proposés. La Figure 6.20 est un menu statique issu du pied de page. Il s'agit d'un menu différent de celui présenté dans la Figure 6.19.

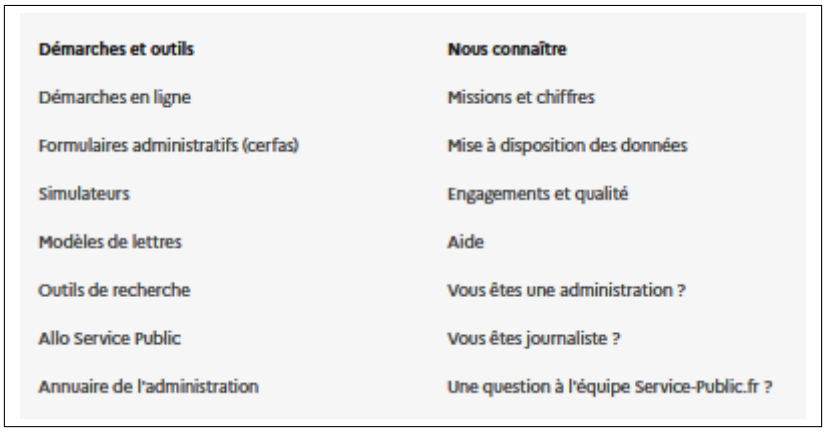


FIGURE 6.20 – Capture d'écran d'un menu statique en bas de page

source : `www.service-public.fr/particuliers/vosdroits/F34717`, consulté le 8 juillet 2024

Le menu illustré par la Figure 6.20 permet d'accéder à différentes parties du site. Il est composé de termes, mais aussi de questions qui permettent aux utilisateurs d'identifier les pages réservées à leurs profils. Il est donc raisonnable d'envisager que la fonction *navigation* joue un rôle important sur le site web.

La fonction *contenu*, quant à elle, correspond à 58 % du corpus total. Il s'agit de la fonction principale du texte présent sur les pages collectées. Elle fait l'objet d'une

caractérisation linguistique dans le chapitre 5. Les hyperliens sont peu présents dans ce sous-corpus, puisqu'ils représentent 14,5 % de l'ensemble du sous-corpus (contre 100 % de liens dans la *navigation* par exemple). Ils prennent également une forme différente, comme nous pouvons le voir dans la Figure 6.21 ci-dessous qui présente des hyperliens issus du *contenu*.

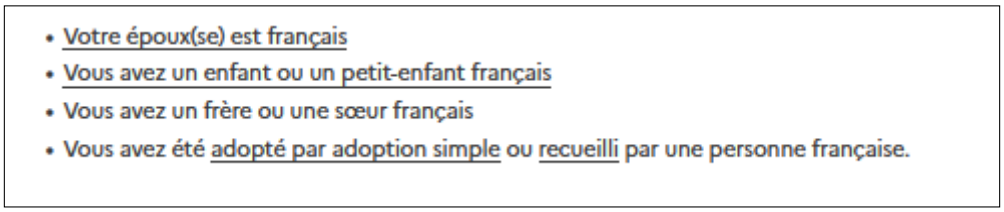


FIGURE 6.21 – Capture d'écran d'hyperliens dans le *contenu*

source : `www.service-public.fr/particuliers/vosdroits/F34717`, consulté le 8 juillet 2024

Dans la Figure 6.21, nous pouvons observer une liste de profils. Les éléments soulignés sont des hyperliens qui redirigent vers d'autres pages du site. Ces pages sont consacrées au sujet indiqué en texte, afin d'expliciter le discours. En cliquant par exemple sur le premier lien (« votre époux(se) est français »), l'utilisateur arrive sur la page « Naturalisation par mariage » qui est donc adaptée à son profil.

La Figure 6.22 illustre d'autres types de liens dans le *contenu*.

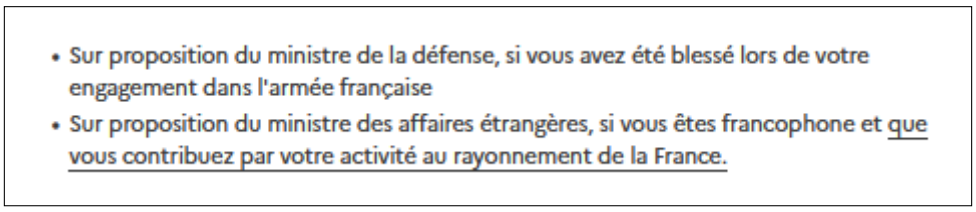


FIGURE 6.22 – Capture d'écran d'hyperliens dans le *contenu*

source : `www.service-public.fr/particuliers/vosdroits/F34717`, consulté le 8 juillet 2024

Dans la Figure 6.22, on peut voir que certains liens peuvent être accompagnés de syntagmes verbaux plus longs.

Sur l'ensemble du corpus, indépendamment de la fonction des segments, les hyperliens représentent 33 % du texte. Cela signifie que près d'un tiers du discours présent sur le site est cliquable. La Figure 6.23 ci-dessous propose une représentation de la place des liens dans le discours de l'accès aux droits en France.

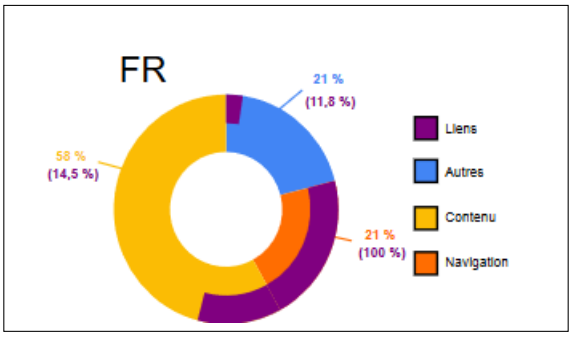


FIGURE 6.23 – Représentation graphique du pourcentage de liens dans le corpus français

La Figure 6.23 a été réalisée à partir des Tables 6.1 et 6.2. Les pourcentages entre parenthèses correspondent au taux de liens dans chaque partie, 14,5 % pour le *contenu*, 100 % pour la *navigation*, et 11,8 % pour le reste du corpus. On voit donc que les liens ont une grande place dans le corpus, notamment dans les segments *navigation*.

Ces données montrent que les liens jouent un rôle important dans le discours numérique de l'accès aux droits. Tout d'abord, ils permettent à l'utilisateur de se déplacer dans l'arborescence du site, ce qui est indispensable. Au-delà de cette utilisation pratique, une grande partie du discours présenté à l'utilisateur est un hyperlien, ce qui signifie que l'utilisateur a sans arrêt le choix de cliquer et par conséquent d'interrompre sa lecture. De plus, si les liens correspondent à des caractéristiques discursives spécifiques, ils peuvent influencer le discours numérique de l'accès aux droits. Nous avons donc réalisé une exploration discursive et linguistique des segments étiquetés comme liens, qui sont présentés dans la Table 6.1.

La première catégorie syntaxique qui nous a intéressée est celle des noms. Les menus des sites web étant majoritairement composés de noms, il nous semblait intéressant de vérifier si les syntagmes présents dans les liens étaient majoritairement des noms. Nos analyses statistiques ont permis d'identifier une grande proportion de noms dans les hyperliens. Nous avons compté 32 % de syntagmes nominaux en moyenne, et jusqu'à 53 % lorsqu'il s'agit de segments ayant la fonction de *navigation*. En comparant ces résultats au pourcentage de syntagmes nominaux dans les segments ayant la fonction *contenu* (y compris, mais ne se limitant pas aux hyperliens), qui est de 30 %, on observe que les noms sont utilisés beaucoup plus fréquemment dans la *navigation*, ce qui confirme nos observations initiales. De plus, nous avons relevé 53 % de noms dans les liens *navigation*, tandis qu'il n'y en a que 33,7 % dans les liens *contenu*. Nous avons relevé 41 % de noms dans l'ensemble des liens du corpus, ce qui est légèrement plus élevé que dans le corpus complet (32 %). À l'inverse, nous observons un pourcentage de syntagmes verbaux inférieur dans les hyperliens, environ 7,5 % dans le corpus d'hyperliens

contre 12 % dans le corpus complet. Comme précédemment, la plus grande différence est avec les segments étiquetés comme étant du *contenu*, qui affiche 16 % de syntagmes verbaux et ceux étiquetés comme étant de la *navigation* qui affiche 8,4 %. Les hyperliens issus des segments étiquetés comme *contenu* présentent également un pourcentage de syntagmes verbaux plus élevé avec 11 %, contre 8,4 % pour les hyperliens issus des segments *navigation*. Cela peut être dû au fait que les hyperliens issus de la fonction *contenu* sont souvent intégrés à des phrases, comme le montre la Figure 6.22.

L'observation des pronoms montre un taux très bas de pronoms dans les hyperliens (1,2 %), tandis que le *contenu* indique 6,2 % et le corpus complet 4,3 %. Enfin, d'après nos relevés, les acronymes et sigles sont présents dans l'ensemble des catégories, pour chacune des fonctions. Toutefois, comme pour les verbes et les pronoms, il semble y avoir une proportion légèrement plus élevée d'acronymes et de sigles dans les segments issus du *contenu*.

Ces résultats peuvent donc indiquer que les hyperliens sont principalement composés de syntagmes nominaux, mais aussi que les hyperliens issus des segments *contenu* utilisent des catégories syntaxiques plus variées pour s'adapter à leur environnement. Les hyperliens du corpus parcours français présentent donc certaines caractéristiques spécifiques.

6.2.2.2 Analyse des liens en corpus britannique

Nous avons ensuite réalisé une étude similaire sur le corpus anglais. Les Tables 6.3 et 6.4 présentent les données analogues pour l'anglais.

Corpus Britannique			
Fonctions	Liens internes	Liens internes dans Navigation	Liens internes dans Contenu
Mots occ./%	8 045 / 46 %	4 976 / 96 %	2 655 / 32 %
Noms occ./%	3 710 / 46 %	2 335 / 46 %	1 120 / 42 %
Verbes occ./%	756 / 9 %	393 / 7,8 %	304 / 11%
Pronoms occ. /%	110 / 1,3 %	35 / 0,7 %	77 / 3 %
Sigles occ./%	232 / 2,8 %	118 / 2,3 %	24 / 0,9 %

TABLE 6.3 – Statistiques des hyperliens présents dans le corpus parcours anglais

Corpus Britannique			
Fonctions	Contenu	Navigation	Corpus complet
Mots occ./%	8 120 / 48 %	5 138 / 30 %	17 154 / 100 %
% de liens dans la section	32 %	96 %	46 %
Noms occ./%	3 141 / 38,6 %	3 074 /59 %	6 862 / 40 %
Verbes occ./%	1 436 / 17,6 %	507 / 9,8 %	2 356 / 13,7 %
Pronoms occ./%	658 / 8 %	41 / 0,8 %	890 / 5,2 %
Sigles occ./%	184 / 2,2 %	135 / 2,6 %	344 / 2 %

TABLE 6.4 – Statistiques des sections *contenu* et *navigation* en anglais britannique

Comme pour le corpus français, pour chaque catégorie étudiée, nous indiquons le nombre d'occurrences, puis le pourcentage correspondant. Tout d'abord, nous observons que les segments ayant la fonction *navigation* ne sont pas toujours des hyperliens internes. Il est important de noter qu'ici nous n'avons pas étudié les hyperliens *externes*, qui pourraient donc constituer les 4 % restant. Nous avons noté une grande proportion de segments dédiés à la *navigation*. En effet, un tiers du texte collecté (34 %) est dédié à la *navigation*, ce qui correspond à une partie non négligeable du corpus. Les segments du texte dédiés à la *navigation* sont principalement des menus.

Les liens sont plus présents dans les segments *navigation* (96 %), et moins présents dans les segments étiquetés comme étant du *contenu* (32 %). Cela semble correspondre aux différentes fonctions, les parties consacrées à la *navigation* étant de nature classificatoire et ayant pour rôle d'aider l'utilisateur à se déplacer. Nous pouvons l'observer dans la Figure 6.24 ci-dessous qui présente un menu en bas de page du site `gov.uk`.

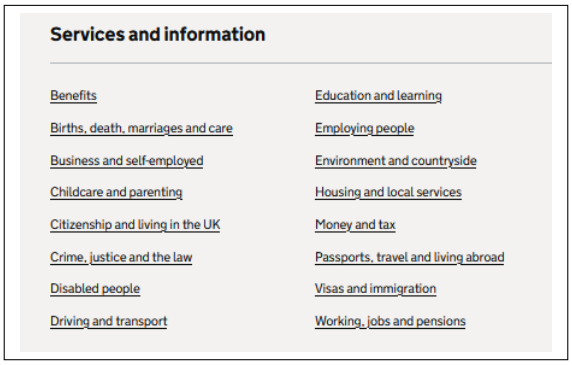


FIGURE 6.24 – Capture d'écran d'un menu

source : `https://www.gov.uk/british-citizenship`, consulté le 8 juillet 2024

Ce menu est composé majoritairement de termes du domaine de l'accès aux droits. Chaque thème correspond à une branche de l'arborescence.

Les parties consacrées au *contenu* sont plutôt informatives et les liens présents dans

ces parties font plus état de l'intertextualité des textes ou du besoin de vulgarisation de certaines informations (voir section 3.2.3). La Figure 6.25 présente un extrait du *contenu* avec des hyperliens.

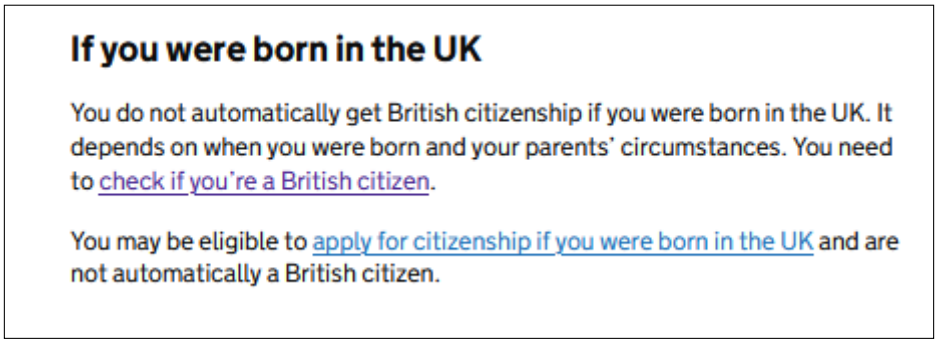


FIGURE 6.25 – Capture d'écran d'hyperliens dans le contenu

source : `https://www.gov.uk/british-citizenship`, consulté le 8 juillet 2024

Les hyperliens conservent les couleurs conventionnelles (bleu pour un lien sur lequel on a pas encore cliqué, puis violet une fois que la page a été visitée). Ce sont également des syntagmes plus longs que les termes observés dans les menus des segments *navigation*. Les hyperliens s'adaptent donc à leur environnement discursif.

Sur l'ensemble du corpus, les liens représentent 46 % du texte, ce qui signifie que près de la moitié du texte disponible à l'utilisateur est cliquable. Les liens jouent donc un rôle essentiel dans le discours de l'accès aux droits britannique. Au-delà de leur aspect structurant et classificatoire, ils sont incontournables pour l'utilisateur, qu'il décide ou non de cliquer dessus. La Figure 6.26 ci-dessous propose une représentation de la place des liens dans le discours de l'accès aux droits au Royaume-Uni.

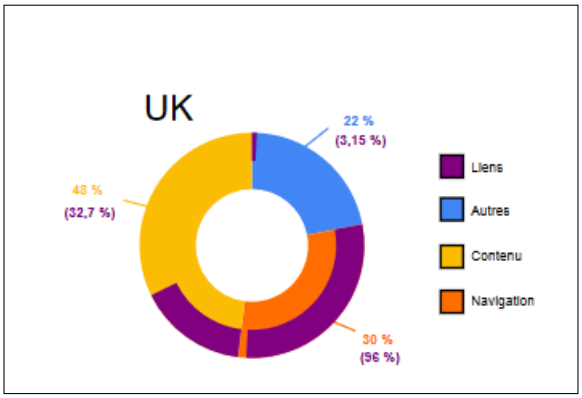


FIGURE 6.26 – Représentation graphique du pourcentage de liens dans le corpus britannique

La Figure 6.26 a été réalisée à partir des Tables 6.3 et 6.4. Les pourcentages entre parenthèses correspondent au taux de liens dans chaque partie, 32 % pour le *contenu*, 96

% pour la *navigation*, et 3,5 % pour le reste du corpus. Les liens ont une grande place dans le corpus, notamment dans les segments *navigation*.

Tout comme dans le corpus français, nous nous sommes intéressée tout d'abord aux noms, qui peuvent avoir un rôle structurant dans un site web. Ils sont très présents dans les liens issus des segments de *navigation* (46 %), mais aussi très présents dans les liens issus des segments *contenu* (46 %). Dans le corpus entier, les noms représentent 40 % du texte, ce qui est nettement supérieur au corpus de référence, l'enTenTen, qui affiche 28 % de noms. La présence élevée de noms peut être caractéristique des langues de spécialité, qui disposent d'une terminologie importante. La deuxième catégorie syntaxique que nous avons étudiée est celle des verbes, qui représente une partie moins importante des hyperliens. Pour cette catégorie, nous pouvons voir une disparité entre les segments étiquetés comme étant de la *navigation* et les segments étiquetés comme étant du *contenu* (respectivement 9,8 % et 17,6 %). Cette disparité est réduite lorsqu'il ne s'agit que des liens de chacune de ces catégories (respectivement 7,8 % et 9 %), mais reste tout de même présente. Enfin, la dernière catégorie syntaxique qui nous a intéressés est celle des pronoms. Pour cette catégorie, il existe une disparité importante entre les segments *navigation* avec une quasi-absence de pronoms (0,8 %) et *contenu* avec une présence relativement élevée (8 %), que nous avons abordée dans la section 5.2.1. Les hyperliens de la section *navigation* présentent quant à eux 0,7 % de pronoms contre 3 % dans les hyperliens de la section *contenu*. À nouveau, les hyperliens issus du *contenu* s'adaptent à l'environnement discursif dans lequel ils apparaissent. Pour compléter cette analyse des catégories syntaxiques, nous avons également relevé les sigles, dont l'utilisation dans les hyperliens de la section *navigation* (2,3 %) est plus importante que dans les hyperliens de la section *contenu* (0,9 %). Toutefois, ces résultats sont à nuancer, car l'ensemble des sigles présents dans la *navigation* sont :

— *the UK*,
— *the EU*.

Ces résultats ne semblent donc pas significatifs.

En conclusion, les pronoms sont largement moins présents dans les segments étiquetés comme étant de la *navigation*. Toutefois, les noms et les verbes semblent être présents de façon similaire dans la *navigation* et dans le *contenu*. Enfin, les liens sont bien moins présents dans les sections de *contenu*, et très présents dans les sections *navigation*.

6.2.2.3 Analyse contrastive des liens

Les relevés réalisés sur les corpus français et britannique montrent des résultats similaires au niveau des catégories syntaxiques. Le discours d'escorte des hyperliens est bien plus marqué par la nominalisation que le reste du discours, ce qui ne semble pas surpre-

nant étant donné la nature classificatoire des menus. Nous observons toutefois quelques différences. Toute d'abord, la proportion d'hyperliens est plus élevée dans le corpus issu des sites web du Royaume-Uni (46 %) que dans le corpus français (33 %). Toutefois, les valeurs restent relativement proches. Le corpus français contient beaucoup de mots et le *contenu* y occupe une plus grande part : 58 % du texte dans le corpus français, contre 47 % dans le corpus britannique. À l'inverse, la partie *navigation* est moins présente dans le corpus français (21 %) et plus présente dans le corpus britannique (30 %). Cette différence pourrait expliquer un taux d'hyperliens légèrement plus élevé dans le corpus britannique, qui est plus marqué par les menus. En conclusion, même si nous observons les mêmes tendances dans les deux langues : une proportion élevée de texte cliquable, avec un taux plus élevé de syntagmes nominaux, certaines différences persistent. La disparité entre la taille des corpus montre également que les pratiques discursives sont différentes dans les deux pays.

6.2.2.4 Spécificités discursives relevées dans le sous-corpus liens français

Nous allons maintenant présenter les spécificités discursives relevées dans le sous-corpus liens français.

Pour explorer le corpus de *liens internes*, nous avons tout d'abord fait une recherche autour des noms. La Table 6.5 ci-dessous présente un récapitulatif des noms les plus fréquents dans le corpus composé d'hyperliens.

word	Frequency	Freq. per million	% of concordance
ligne	156	6187,52975	1,64505
secteur	131	5195,93844	1,38142
logement	130	5156,27479	1,37087
fonction	120	4759,63827	1,26542
étranger	114	4521,65635	1,20215
travail	99	3926,70157	1,04397
France	93	3688,71966	0,9807
thème	87	3450,73774	0,91743
naturalisation	87	3450,73774	0,91743
administration	86	3411,07409	0,90689
Aides	84	3331,74679	0,8858
Santé	70	2776,45566	0,73816
Assurance	67	2657,4647	0,70653
Hébergement	64	2538,47374	0,67489

TABLE 6.5 – Noms les plus fréquemment utilisés dans le sous-corpus liens français

Nous avons également recherché les n-grammes les plus fréquents qui sont présentés dans la Table 6.6 ci-dessous. Ils illustrent les séquences faisant partie des phrases ou titres qui font office d'hyperliens.

Item	**Frequency**
la fonction publique	104
à l' étranger	89
le secteur privé	83
Afficher le thème	80
de naturalisation française	64
Demande en ligne	63
en ligne de	62
ligne de naturalisation	62
votre ville ou	56

TABLE 6.6 – N-grammes les plus fréquemment utilisés dans le sous-corpus de liens français

Notre première observation est que le corpus composé d'hyperliens est très marqué par les répétitions. Nous relevons par exemple 83 répétitions de **« (secteur privé) »**, ou encore 81 occurrences de **« dans le secteur privé »**. Plus surprenant, les noms figurant en première position ne semblent pas, ou peu, faire référence au thème de nos parcours de navigation : naturalisation, allocation logement ou acte de naissance. Le cinquième résultat proposé dans la Table 6.6 est le suivant :

de naturalisation française

Il semble correspondre au thème du parcours qui concerne la naturalisation. Il apparaît ici, car une des pages du parcours de navigation naturalisation comporte un segment répété 64 fois. Dans la Figure 6.27 présenté ci-dessous, nous pouvons voir un exemple de l'hyperlien en contexte : « Demande en ligne de naturalisation française ».



FIGURE 6.27 – Capture d'écran d'hyperliens dans le contenu

source : `https://www.service-public.fr/particuliers/vosdroits/F2213`, consulté le 10 décembre 2021

La Figure 6.27 est un encadré qui indique où envoyer sa demande de naturalisation suivant le lieu d'habitation. Il apparaît 64 fois en corpus. À noter que cet encadré n'existe pas sur la version du site en juillet 2024, car les demandes de naturalisation sont maintenant centralisées sur la plateforme Anef.

Nous pouvons également relever un n-gramme qui semble faire référence aux parcours spécifiques :

à l'étranger

Dans ce cas, il ne s'agit pas de répétitions, mais plutôt d'un syntagme classificatoire, qui revient régulièrement dans le menu du site, et qui figure donc dans les segments d'hyperliens présentés ici. Nous avons ainsi relevé les occurrences suivantes :

— **Assurance maladie et santé d'un Français à l'étranger**,
— **Conduire à l'étranger**,
— **Voyager à l'étranger**,
— **Vivre à l'étranger**

Ce syntagme classificatoire existe également pour le parcours **état civil**. Nous avons relevé les menus suivants :

— **Actes d'état civil**,
— **Changement d'état civil**,
— **Registre d'état civil électronique (RECE)**.

Ces relevés nous montrent le rôle classificatoire des hyperliens, qui permettent d'identifier les différentes parties de l'arborescence du site. La répétition de la catégorie **étranger** ou **état civil** permet de garder une cohérence et une stabilité pour l'usager qui reconnaît qu'il s'agit d'une même branche de l'arborescence.

L'analyse des autres relevés présents sur les Tables 6.5 et 6.6 montre qu'ils ne sont pas directement liés aux thèmes des parcours, puisqu'ils renvoient à des situations plus générales. On retrouve une forte fréquence du segment **Dans la fonction publique**, qui ne semble liée à aucun parcours. L'exploration du corpus et de ces relevés nous montre que l'ensemble des hyperliens présents sur les pages ne sont pas en liens directs avec le thème de la page. En effet, ils ont pour fonction de proposer des passerelles vers d'autres thèmes ou d'autres profils, que l'on peut retrouver sur le site. Une autre caractéristique saillante de ces tableaux est que la terminologie occupe une place importante.

Nous avons poursuivi avec l'étude des verbes les plus fréquents. La Table 6.7 présente les 20 verbes les plus présents dans le corpus.

word	Frequency	Freq. per million	% of concordance
privé	141	5592,57496	8,01592
Fermer	138	5473,58401	7,84537
Voir	105	4164,68348	5,9693
Demande	102	4045,69253	5,79875
Afficher	82	3252,41948	4,66174
est	78	3093,76487	4,43434
Accéder	34	1348,56418	1,93292
faire	29	1150,24591	1,64866
connecter	26	1031,25496	1,47811
êtes	25	991,59131	1,42126
Aller	24	951,92765	1,36441
montant	21	832,9367	1,19386
Ouvrir	20	793,27304	1,13701

TABLE 6.7 – Verbes les plus fréquemment utilisés dans le sous-corpus liens français

Certains relevés ne sont pas pertinents, notamment **Demande** et **privé** qui ne nous intéressent pas ici puisqu'ils ne sont pas employés en tant que verbe, mais en tant que nom et participe passé à valeur adjectivale respectivement. En analysant les autres relevés toutefois, nous pouvons observer qu'au moins 7 verbes font référence au processus de navigation :

— Fermer,
— Voir,
— Afficher,
— Accéder,
— Connecter,

— Aller,
— Ouvrir.

Une exploration en corpus permet de confirmer que ces verbes font référence au processus de navigation. Cette terminologie questionne les limites du discours spécialisé de l'accès aux droits numérique. Il convient de se demander si ces parties du discours spécialisé font partie du domaine, par exemple avec le schéma phraséologique suivant :

Afficher le thème [thème]

Il s'agit d'un schéma possédant 82 occurrences dans le corpus parcours et qui permet à l'utilisateur de se repérer. Une partie de la terminologie des discours numériques est issue d'internet avec le verbe **Afficher**, tandis que l'autre partie qui concerne les thèmes relève de la terminologie du domaine de l'accès aux droits ou de l'administration.

L'étude des hyperliens dans le discours permet donc d'identifier une terminologie et une phraséologie hybride, issue de deux domaines qui se rassemblent dans le discours numérique de l'accès aux droits. De plus, la terminologie et la phraséologie générale d'internet semble interagir avec la terminologique du discours spécialisé : les catégories notionnelles et les termes clés cherchent à être plus saillants à travers les menus, les liens et les répétitions, le discours devient dialogique (au lieu de monologal) donnant lieu à une phraséologie spécifique.

L'exploration des pronoms en corpus confirme que le pronom **vous** est le plus utilisé, avec 61 occurrences. De plus, certains liens réutilisent des schémas répétés avec question rhétorique telle que « Vous êtes une administration ? » qui signifie que les administrations sont concernées par ce lien, comme nous l'avons vu dans 5.4.1. Les autres occurrences relevées ne semblent pas caractériser les hyperliens comme ayant une utilisation spécifique des pronoms.

Enfin, l'exploration des acronymes nous a semblé essentielle, car ils représentent une caractéristique du discours de l'accès aux droits et des discours institutionnels en général. Les acronymes ont fait l'objet d'une étude dans la section 5.1.4 qui nous a permis d'identifier leur utilisation comme une spécialisation du discours de l'accès aux droits sur les sites web. Nous souhaitons à présent observer leur présence dans le sous-corpus d'hyperliens du corpus parcours. La Table 6.8 ci-dessous présente l'ensemble des sigles et acronymes présents dans le corpus.

word	Frequency	Freq. per million	% of concordance
MSA	27	1070,91861	16,77019
APL	14	555,29113	8,69565
RSA	11	436,30017	6,8323
OQTF	9	356,97287	5,59006
AEEH	9	356,97287	5,59006
JDC	9	356,97287	5,59006
CRDS	8	317,30922	4,96894
IFI	8	317,30922	4,96894
ARE	8	317,30922	4,96894
DPU	8	317,30922	4,96894
PCH	8	317,30922	4,96894
AAH	8	317,30922	4,96894
IVG	8	317,30922	4,96894
CSG	8	317,30922	4,96894
ALS	6	237,98191	3,72671
ALF	6	237,98191	3,72671
CDD	1	39,66365	0,62112
CNF	1	39,66365	0,62112
RAR	1	39,66365	0,62112
SDF	1	39,66365	0,62112
RECE	1	39,66365	0,62112
VAE	1	39,66365	0,62112

TABLE 6.8 – Acronymes et sigles présents dans le corpus parcours français

L'analyse de la Table 6.8 nous permet d'identifier que les acronymes et sigles présentés sont similaires à ceux présentés dans la table 5.42. Il n'y a pas donc d'utilisation particulière des sigles et acronymes dans les liens. Il est néanmoins intéressant de noter que l'ensemble des sigles et acronymes présentés ici font référence à des dispositifs ou organismes en lien avec l'administration et l'accès aux droits. En d'autres termes, nous n'avons pas identifié de sigles ou d'acronymes faisant référence aux sites web, comme *PDF* dans le corpus présenté dans la section 5.1.4. La focalisation sur le parcours utilisateur explique cette absence du sigle PDF qui fait partie de la terminologie générale d'internet. L'utilisation des acronymes ou sigles est bien présente dans les hyperliens. Toutefois, ils ne semblent pas faire l'objet d'une utilisation particulière par rapport au reste du corpus.

6.2.2.5 Spécificités discursives relevées dans le sous-corpus liens britannique

Nous allons maintenant présenter les spécificités discursives relevées dans le sous-corpus d'hyperliens anglais. Pour explorer le corpus de liens internes, nous avons tout d'abord fait une recherche autour des noms. La Table 6.9 ci-dessous présente un récapitulatif des noms les plus fréquents dans le sous-corpus d'hyperliens.

word	Frequency	Freq. per million	% of concordance
UK	133	15555,55556	3,58491
Search	92	10760,23392	2,47978
people	78	9122,80702	2,10243
GOV	75	8771,92982	2,02156
Government	64	7485,38012	1,72507
Housing	61	7134,50292	1,6442
support	59	6900,5848	1,5903
cookies	52	6081,87135	1,40162
Citizenship	52	6081,87135	1,40162
page	46	5380,11696	1,23989
tax	46	5380,11696	1,23989
death	45	5263,15789	1,21294
Benefits	45	5263,15789	1,21294
Childcare	42	4912,2807	1,13208

TABLE 6.9 – Noms les plus fréquemment utilisés dans le sous-corpus liens britannique

En ce qui concerne le sous-corpus liens britannique, l'analyse des relevés montre que les noms les plus utilisés dans le corpus sont un mélange entre la terminologie de l'accès aux droits et de l'administration avec les relevés suivants :

— *UK*
— *Governement*
— *Housing*
— *Citizenship*
— *tax*
— *death*
— *Benefits*
— *Healthcare*

et la terminologie des sites web avec les exemples suivants :

— *Search*
— *GOV*
— *support*
— *cookies*
— *page*

Dans tous les cas, les exemples nous montrent que la terminologie est très présente dans les hyperliens, ce qui semble cohérent avec leur rôle classificatoire.

Nous avons ensuite exploré les verbes présents dans le sous-corpus de liens britannique, que nous présentons dans la Table 6.10 ci-dessous.

word	Frequency	Freq. per million	% of concordance
living	117	13684,21053	15,47619
parenting	40	4678,36257	5,29101
learning	39	4561,40351	5,15873
is	39	4561,40351	5,15873
Employing	39	4561,40351	5,15873
Get	39	4561,40351	5,15873
transport	34	3976,60819	4,49735
works	33	3859,64912	4,36508
get	32	3742,69006	4,2328
Find	27	3157,89474	3,57143
Send	23	2690,05848	3,04233
apply	20	2339,18129	2,6455
involved	18	2105,26316	2,38095
Reject	18	2105,26316	2,38095
Accept	18	2105,26316	2,38095
're	13	1520,46784	1,71958
doing	12	1403,50877	1,5873
were	11	1286,54971	1,45503

TABLE 6.10 – Verbes les plus fréquemment utilisés dans le sous-corpus lien britannique

La grande majorité des verbes présents et répertoriés dans la Table ci-dessus semblent être liés au domaine de l'accès aux droits. Nous retrouvons seulement quelques verbes non spécialisés ou qui ne relèvent pas de domaines connexes (*is*, *get*, *'re* (pour *are*), *doing* et *were*) et quelques verbes spécifiques à la nature numérique du discours. Concernant les premiers, en observant les occurrences de *'re* en corpus, nous pouvons noter qu'il s'agit d'un extrait du schéma conditionnel *if you're [profil]*. Il s'agit donc d'un verbe faisant partie d'une unité spécialisée.

Concernant les seconds, nous avons identifié les verbes spécifiques à la nature numérique du discours suivants :

— *Find*
— *Send*
— *Reject*
— *Accept*

Ces verbes d'action semblent très présents et montrent, encore une fois, la place importante accordée à l'usager dans ce genre discursif, puisque ces verbes servent à indiquer à l'utilisateur les actions disponibles via le lien.

L'analyse des verbes anglais comme celle des verbes français révèle l'existence d'une terminologie double : de l'accès aux droits et des sites web. L'analyse du sous-corpus de liens montre que le discours numérique de l'accès aux droits présente bien une terminologie à part entière, issue du discours administratif, juridique et institution-

nelle, mais aussi du discours d'internet et des sites web.

L'analyse des pronoms reste en accord avec les analyses réalisées dans la section 5.2.1. En effet, le pronom principalement présent est *you*, notamment dans le schéma *if you're [profil]*, mais aussi dans des schémas présentant de la modalité. L'utilisation des pronoms ne semble donc pas caractéristique des liens, mais plutôt respecter les normes discursives du discours en général.

Enfin, nos analyses se sont intéressées aux acronymes et sigles qui sont caractéristiques de la terminologie du discours. Ils sont répertoriés dans la Table 6.11 ci-dessous.

word	Frequency	Freq. per million	% of concordance
UK	133	15555,55556	91,09589
EU	5	584,79532	3,42466
II	1	116,95906	0,68493
LHA	1	116,95906	0,68493
SMI	1	116,95906	0,68493
TV	1	116,95906	0,68493
VAT	1	116,95906	0,68493
EEA	1	116,95906	0,68493
NR	1	116,95906	0,68493
ILR	1	116,95906	0,68493

TABLE 6.11 – Acronymes et sigles présents dans le corpus parcours britannique

Malgré la présence de sigles dans le sous-corpus de liens, 133 des 146 occurrences relevées par notre outil sont le sigle *UK* et 5 occurrences sont le sigle *EU*. Il semble donc que les acronymes et sigles ne sont pas caractéristiques des liens. Les sigles les plus présents sont issus de la langue générale et non du discours spécialisé étudié. Nous retrouvons tout de même une occurrence de *LHA* pour *Local Housing Allowance* et *SMI* pour *Support for Mortgage Interest* qui relève de notre domaine d'étude. Les occurrences du sigle *UK* relève du rôle classificatoire des liens, car ils permettent de se repérer autour de différentes problématiques de la catégorie *living in the UK*. Ici, la répétition de *in the UK* dans l'ensemble des menus et des liens permet de garder une cohérence discursive pour l'utilisateur.

6.2.2.6 Analyse comparative

Les analyses réalisées sur les sous-corpus anglais et français permettent de mettre en avant certaines caractéristiques. Tout d'abord, dans les deux corpus, il semble y avoir une grande proportion de termes dans les hyperliens. Cette terminologie est issue de deux domaines : l'accès aux droits et les sites web qui, ensemble, forment le domaine de l'accès aux droits numérique.

Les acronymes et sigles, caractéristique présente dans les deux langues, ne sont pas retrouvés dans le corpus de liens en anglais. Toutefois, on y retrouve des schémas conditionnels, ce qui n'est pas le cas du corpus français.

Dans les deux corpus, les liens ont un rôle classificatoire et proposent des phrases et schémas répétés pour aider l'utilisateur à se repérer.

6.2.3 Interactivité

Nous explorons dans cette section le corpus parcours annoté de façon à nous permettre de distinguer divers éléments interactifs, notamment les *formulaires*, où l'utilisateur peut transmettre des informations aux serveurs du site web, les *éléments éphémères* tels que les pop-up et les informations d'actualités, et les *menus déroulants*, c'est-à-dire les textes sur lesquels l'utilisateur doit cliquer afin de faire apparaître du texte.

Nous présentons les statistiques des différents types d'éléments interactifs en français (Table 6.12) et en anglais (Table 6.13).

	Corpus français				
	Interactivité	Interactivité + formulaire	Interactivité + déroulant	Interactivité + éphémère	Corpus complet
Mots	3 034	1 579	995	460	69 567
%	4,3 %	2,2 %	1,3 %	0,6 %	100 %

TABLE 6.12 – Statistiques des éléments interactifs présents dans le corpus parcours français

	Corpus Britannique				
	Interactivité	Interactivité + formulaire	Interactivité + déroulant	Interactivité + éphémère	Corpus complet
Mots	823	459	243	121	17 154
%	4,7 %	2,6 %	1,4 %	0,7 %	100 %

TABLE 6.13 – Statistiques des éléments interactifs présents dans le corpus parcours britannique

Il est intéressant de noter que dans les Tables 6.12 et 6.13, les pourcentages d'éléments interactifs sont plutôt similaires. Nous retrouvons presque les mêmes valeurs dans les sous-corpus, avec une majorité d'éléments interactifs de type *formulaire*, puis *déroulant* et enfin *éphémère*. Les valeurs de chaque catégorie sont très proches. Cette étude pourrait être approfondie sur un corpus plus conséquent. Toutefois, nous pouvons avancer l'hypothèse que le pourcentage d'éléments interactifs est stable, car ils relèvent de convention liée aux sites web, et plus spécifiquement aux sites web de l'administration. Il est possible que les bonnes pratiques en ce qui concerne l'introduction d'éléments interactifs sur les sites web et/ou les contraintes techniques influencent ces statistiques. Il

est cependant important de noter que l'organisation des pages britanniques et françaises relève de principes différents. Les menus ne se déroulent pas de la même manière sur le site `gov.uk` et sur le site `service-public.fr`. Sur le site britannique, les menus déroulants n'obstruent pas la vue sur la page, tandis que sur le site français, ces derniers s'affichent par-dessus les éléments de la page. Les éléments interactifs, même s'ils jouent un rôle important, pourraient relever de conventions liées au site web et non au discours.

6.2.4 Cohérence et cohésion dans les textes numériques

La cohérence et la cohésion ne sont pas des caractéristiques saillantes du discours numérique. Nous avons noté à plusieurs reprises que le discours était délinéarisé et fragmenté par nature (3.2) par des liens, mais aussi par sa structure en arborescence. De plus, dans notre analyse, nous avons mis en avant un certain nombre de résultats montrant une certaine instabilité et fragmentation du discours. Cette fragmentation se retrouve au niveau terminologique comme présenté dans la section 5.1.3.1 et au niveau conceptuel, dans le corpus français notamment, comme montré dans la partie 5.1.5.

Toutefois, nous pouvons observer que certains éléments assurent la cohérence et la cohésion du discours numérique de l'accès aux droits. Tout d'abord, les phrases répétées présentées dans la section 5.4.1 de ce chapitre se présentent comme un procédé de cohésion, car elles permettent d'aboutir à une forme de cohérence textuelle. Elles assurent une information adéquate sur l'ensemble des pages par rapport à la visée communicative du site web. De plus les liens, qui peuvent favoriser la fragmentation, sont également un facteur de cohérence dans le discours numérique. En effet, même s'ils incitent, voire obligent, l'utilisateur à changer de page et donc à ne pas voir l'ensemble du texte sur un même visuel, ils assurent tout de même que les pages sont accessibles aux endroits propices.

La terminologie est également un élément de cohésion. Bien que la variation crée une instabilité, la terminologie demeure un point d'ancrage fort dans le domaine avec ses contextes définitoires et riches en connaissances qui structurent les liens entre les termes et les connaissances du domaine. Enfin, les discours numériques présents sur les sites web ne sont pas faits pour être appréhendés entièrement. Contrairement à un texte ou à un livre qui présente un fil discursif linéaire, le site web a pour but d'être consulté, non comme un tout, mais par extraits.

La cohérence et la cohésion du discours peuvent également être envisagées dans son interactivité. L'interactivité, notamment par les hyperliens et les menus déroulants, est présente dans le texte de façon régulière et systématique. L'étude de l'interactivité réalisée dans le corpus parcours montre que les liens avec une fonction de *contenu* proposent

une phraséologie et une terminologie similaires au texte non interactif. La cohésion linguistique est donc respectée et la fragmentation n'est que visuelle. La phrase peut être interrompue par l'ouverture d'une nouvelle page, mais il s'agit alors de la volonté de l'utilisateur de compléter les informations. De plus, les unités discursives dédiées à la *navigation* possèdent leurs caractéristiques linguistiques propres et sont donc facilement identifiables. La cohésion entre les pages est assurée par le fil d'Ariane, qui propose une représentation de la place de la page dans l'arborescence. Les Figures 6.28 et 6.29 présentées ci-dessous sont une capture d'écran des fils d'Ariane de nos deux sites principaux.

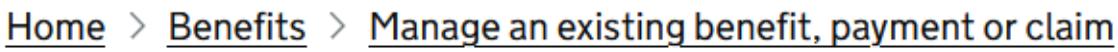


FIGURE 6.28 – Capture d'écran du fil d'Ariane issu de `gov.uk`

source : `www.gov.uk/child-tax-credit`, consulté le 10 juillet 2024



FIGURE 6.29 – Capture d'écran du fil d'Ariane issu de `service-public.fr`

source : `/www.service-public.fr/particuliers/vosdroits/F1427`, consulté le 10 juillet 2024

Ces fils d'Ariane peuvent être vus comme des reconstructions d'un fil discursif. Néanmoins, il ne s'agit pas forcément du fil discursif de l'utilisateur puisque le fil d'Ariane permet seulement de remonter dans l'arborescence et ne suit pas le parcours de l'utilisateur. Cet élément, qui permet à l'utilisateur de se repérer, s'apparente plutôt à une numérotation de page.

Le genre discursif du site web de l'accès aux droits repose sur des unités discursives spatialement fixes, qui permettent de se repérer et de se déplacer. La familiarité acquise avec la répartition spatiale des unités discursives permet de garder une cohérence sur l'ensemble des pages, malgré leur caractère fragmenté.

De ce fait, nous pouvons nous demander si la numérotation des pages web ou un code similaire à la numérotation dans les tables des matières pourrait assurer une plus grande cohérence discursive et informationnelle, tant pour les rédacteurs que pour les usagers. Elle pourrait être associée à la fonction « Plan du site » qui est un élément stable du site se trouvant en bas des pages, et proposant parfois la possibilité d'être affiché sur une nouvelle page (ex. `https://www.service-public.fr/P10003`). En effet, le parcours de lecture repose sur la structure qui est en partie une expérience

visuelle, et dans le cas du discours numérique, la fragmentation du discours est éminemment une fragmentation de type visuel. Le texte numérique peut être parcouru de haut en bas, mais chaque clic suspend les séquences de pages. C'est à l'utilisateur de mémoriser, s'il le peut, son propre parcours. Le fil d'Ariane, s'il met en avant la structure globale du site, n'est pas personnalisé et ne permettra pas à l'usager de remonter vers les informations précédemment consultées. Ce dernier peut toutefois s'attacher à la possibilité que lui offrent les navigateurs d'ouvrir chaque page dans un onglet différent. Cette fonctionnalité permet de garder l'ensemble des pages ouvertes, mais ne permet pas de numéroter le parcours ou de se repérer dans le site. Cette situation rend la structure du discours numérique séquentielle, et il est fort probable que la numérotation des pages, associée à l'arborescence sous forme de table des matières, pourrait créer une structure plus facilement appréhendable.

6.3 Une arborescence ou un labyrinthe

L'ensemble des observations réalisées dans le chapitre 5 et la section 6.1, ainsi que les connaissances relatives aux sites web présentées dans la section 3.1.2.1 nous permettent d'établir que le site web est construit en arborescence. Cette arborescence, d'après les menus de navigation présentés dans la section 6.2.1, est explorable notamment par thème, comme présenté dans les pieds de page 6.13 et 6.11.

En passant par une exploration par menu, c'est-à-dire en choisissant de naviguer à partir de la page d'accueil, l'usager a accès à la structure du site par thème et peut affiner sa recherche au fil des menus. Cette arborescence s'appuie notamment sur la terminologie du domaine. Nous proposons dans la Figure 6.30 un schéma de cette arborescence.

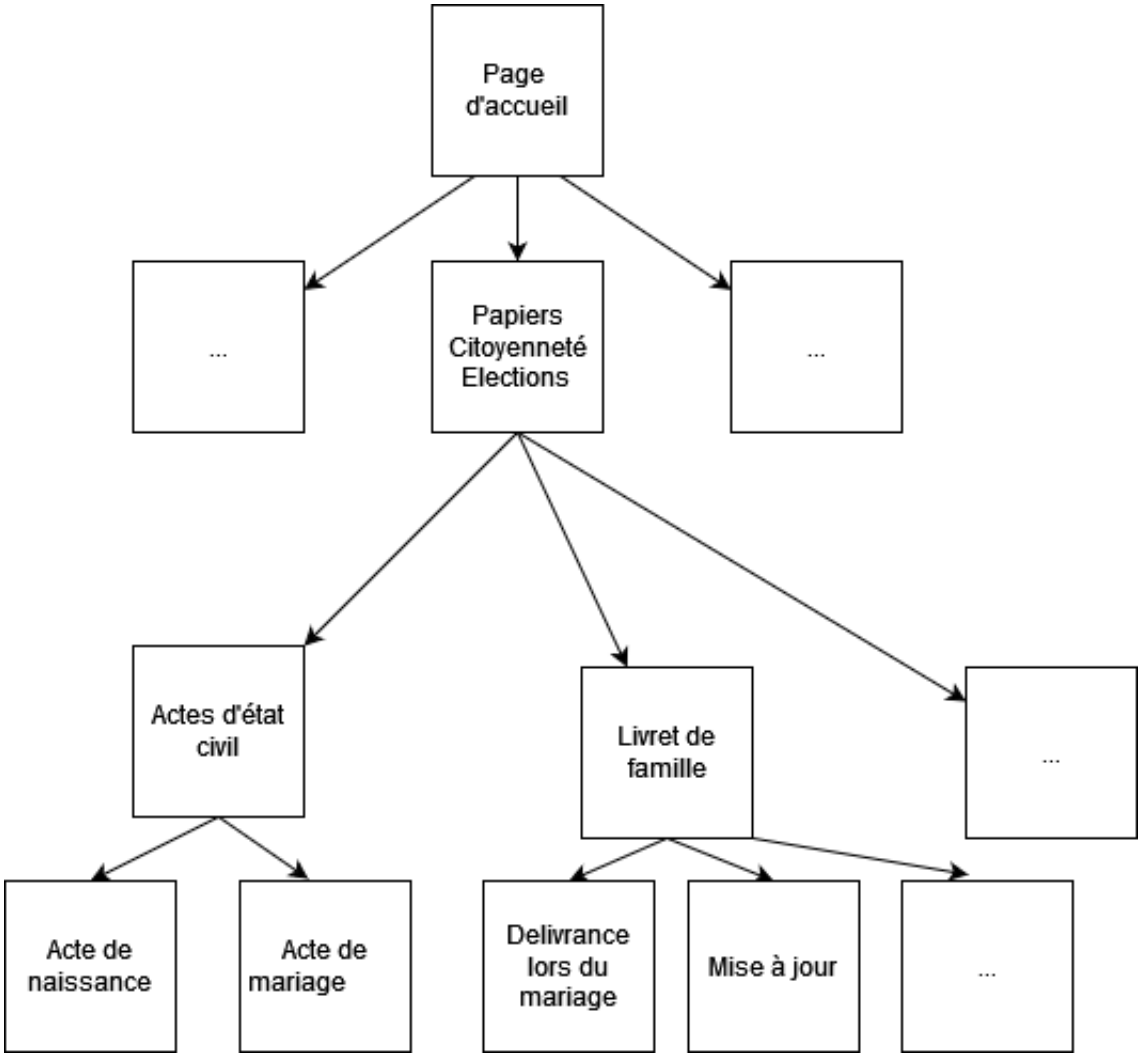


FIGURE 6.30 – Représentation d'une partie de l'arborescence du site `service-public.fr`

Cette représentation graphique de l'arborescence permet de voir la progression des-

cendante des pages. Chaque menu choisi est plus précis et resserre les possibilités de navigation. Cela permet d'obtenir les informations recherchées par l'utilisateur, mais également de se repérer sur les différentes pages. La Figure 6.30 ne représente qu'une partie de l'arborescence, comme l'indiquent les cases avec « [...] ». Le citoyen peut aussi retracer sa navigation grâce au fil d'Ariane. Cette structure en arborescence offre à l'utilisateur la possibilité de naviguer de manière linéaire en suivant un parcours cohérent, guidé par les termes des menus. Il convient de souligner que cette organisation n'est accessible que depuis la page d'accueil et que les liens permettant de descendre dans l'arborescence de façon linéaire sont loin d'être les seuls liens disponibles sur la page, comme nous l'avons vu dans la section 6.2.2 consacrée à leur étude. La Figure 6.31 représente les liens horizontaux de l'arborescence, c'est-à-dire les liens qui permettent de naviguer entre deux pages informatives sans passer par des menus.

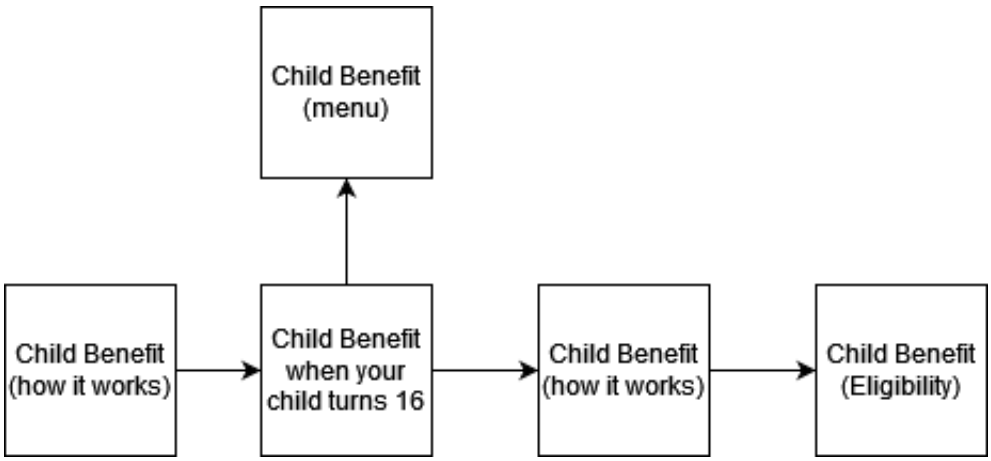


FIGURE 6.31 – Représentation d'une partie de l'arborescence du site `gov.uk`

Cette Figure représente le parcours de l'usager sur le site, autrement dit, son parcours de navigation. Chaque page est dotée d'un très grand nombre de liens qui permettent non seulement de se déplacer de façon linéaire dans l'arborescence, mais aussi de passer de page en page sans repasser par un menu. Cette navigation horizontale permet par exemple d'expliciter certains concepts en proposant des pages liées explicatives. Dans la Figure 6.31, nous pouvons par exemple voir que les pages « *Child Benefit (how it works)* et « *Child Benefit when your child turns 16* » sont liées. Cet hyperlien donne la possibilité à l'utilisateur d'avoir des précisions adaptées à son profil. Toutefois, la navigation horizontale pose la question du fil discursif et de la cohérence du parcours de navigation de l'usager. Dans la Figure 6.31, chaque carré représente une page web possédant une URL différente et s'affichant de façon indépendante. Cette arborescence représente un parcours de navigation possible pour une personne cherchant des informations sur le dispositif *Child Benefit*. Dans ce cas, l'utilisateur n'est pas descendu de façon linéaire dans

l'arborescence mais il est arrivé directement sur la page *Child Benefit (how it works)* par un moteur de recherche.

L'ensemble de ce parcours n'est pas enregistré, ce qui signifie que chaque fois que l'utilisateur clique sur un lien, la page initiale disparaît et laisse place à une nouvelle page. De plus, le fil d'Ariane ne lui permet pas de se repérer dans la mesure où il permet uniquement de remonter dans l'arborescence. L'utilisateur ne peut donc pas recréer son parcours à moins de retourner en arrière à l'aide de son navigateur. Le discours est alors interrompu, puis repris, dans un ordre aléatoire, que le rédacteur ne peut contrôler. Le réseau d'hyperliens sur les sites `service-public.fr` et `gov.uk` est dit « fermé » (Storrer, 2000) puisqu'ils ne mènent qu'à des pages internes au site. L'utilisateur a donc affaire à une série de textes finie, et ne peut pas être redirigé sur un nouveau site qui lui présentera de nouvelles informations[5]. Ces caractéristiques nous mènent à envisager le discours numérique natif diffusé sur les sites web comme un labyrinthe, où l'usager peut se perdre en cherchant l'information dont il a besoin. La situation est d'autant plus complexe si l'on prend en compte les sites web dans leur entièreté, en tant que supports de communication multimodaux. Ils sont alors semblables à une « mosaïque » (nous avons déjà cité Maingueneau (2016, p. 14) qui décrit des sites web en tant qu'« une mosaïque plus ou moins dense de modules hétérogènes »).

Nous comparons donc le discours sur les sites web de l'accès aux droits à un labyrinthe, où l'utilisateur ne peut saisir son parcours dans son ensemble. Il accède plutôt au discours et aux informations par fragments, qui disparaissent progressivement au fur et à mesure de sa navigation. Chaque parcours est éphémère, tout comme les pages, qui sont régulièrement modifiées par les rédacteurs. De plus, la variation terminologique que nous avons analysée dans la section 5.1.3.1 peut renforcer cette impression de labyrinthe. Par exemple, les dispositifs *contribution-based JSA*, *income-based JSA* ou encore *New style JSA* se côtoient sur plusieurs pages. Ils peuvent facilement être confondus par les utilisateurs, d'autant plus que certains variants dérivés comme *income-related JSA* existent. La Figure ci-dessous illustre la co-occurrence de ces variants.

La Figure 6.32 présente les trois premiers résultats de la recherche « JSA ». On remarque une co-occurrence des différents variants dans les menus qui peut être déroutante pour les usagers. De plus, on observe une hiérarchie parmi les pages proposées avec certaines pages qui semblent être des menus ou pages principales, comme *Jobseeker's Allowance (JSA)*, et d'autres pages qui semblent être plus basses dans l'arborescence comme *How it works*.

5. Il existe dans notre corpus quelques liens externes mais ces derniers ne concernent pas les pages informatives.

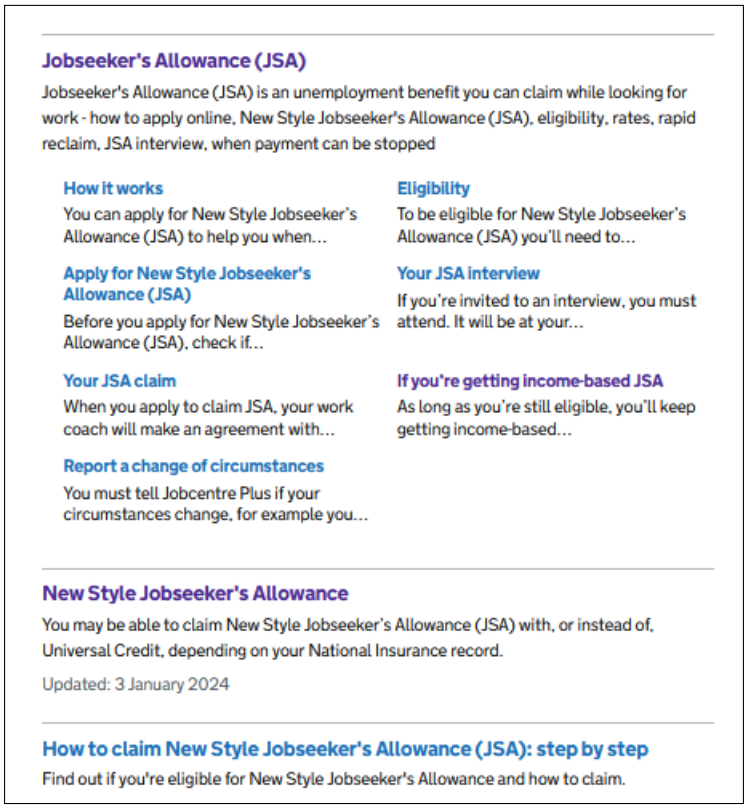


FIGURE 6.32 – Capture d'écran des résultats de recherche de *JSA*

source : `www.gov.uk/search/all?keywords=jsa&order=relevance`, consulté le 17 juillet 2024

La variation terminologique renforce donc l'impression de labyrinthe. De la même façon, la multiplicité des dispositifs peut pousser l'usager à se perdre, comme le montrent ces multiples pages consacrées au droit de séjour des étudiants. La Figure 6.33 est la capture d'écran d'un menu concernant le droit au séjour étudiant.



FIGURE 6.33 – Capture d'écran du menu concernant le droit au séjour des étudiants étrangers

source : `www.service-public.fr/particuliers/vosdroits/N110`, consulté le 17 juillet 2024

La Figure 6.33 propose 6 pages, chacune correspondant à des statuts et situations différents pour l'étudiant étranger. L'ensemble de ces 6 pages ne sont présentées ensemble que dans ce menu. Pour l'utilisateur qui choisit de faire une recherche correspondante dans la barre de recherche avec le terme « carte de séjour étudiant », la Figure 6.34 ci-dessous illustre l'ensemble des pages qui lui sont proposées.

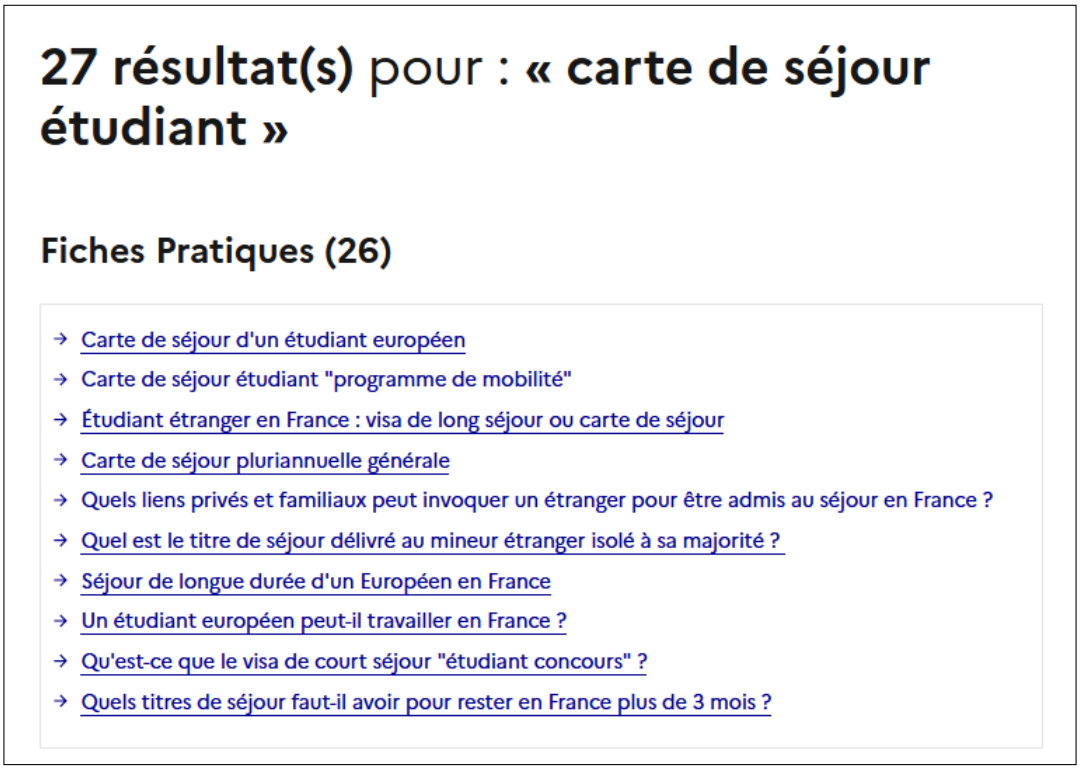


FIGURE 6.34 – Capture d'écran des résultats de recherche de **carte de séjour étudiant**

source : `www.service-public.fr/particuliers/vosdroits/N110`, consulté le 17 juillet 2024

Dans la Figure 6.34, différentes pages concernant les cartes de séjour étudiantes sont proposées. Toutefois, seuls les trois premiers résultats sont des pages présentes dans le menu proposé dans la Figure 6.33. Suivant son mode de navigation du site, l'utilisateur aura accès à différentes informations. De plus, suivant le terme choisi pour la recherche par mot-clé, les résultats seront différents. La Figure 6.35 ci-dessous présente les résultats pour la recherche avec le variant « titre de séjour étudiant ».

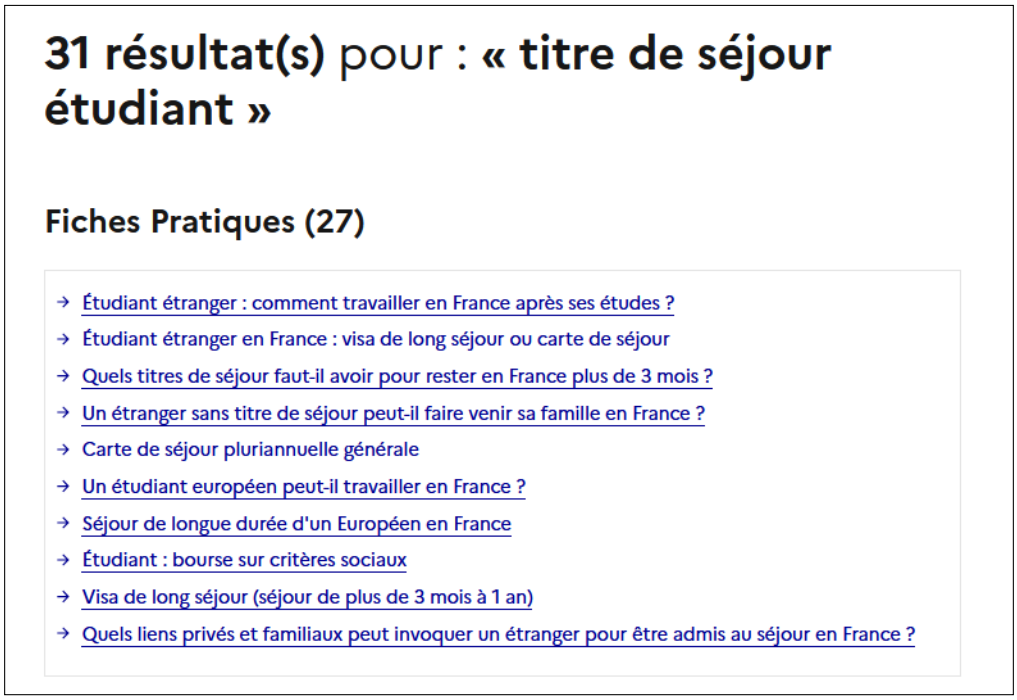


FIGURE 6.35 – Capture d'écran des résultats de recherche de **titre de séjour étudiant**

source : `www.service-public.fr/particuliers/vosdroits/N110`, consulté le 17 juillet 2024

Les résultats proposés dans la Figure 6.35 présentent des différences avec la recherche précédente (Figure 6.34). Ces différences sont dues à l'utilisation du variant **titre de séjour** à la place de **carte de séjour**.

Les Figures 6.33, 6.34 et 6.35 proposent toutes des pages différentes autour du même concept. Ces pages peuvent parfois être accessibles entre elles, en naviguant horizontalement, comme nous l'avons montré dans la Figure 6.31, mais ce n'est pas toujours le cas. Les informations auxquelles accède l'usager ne sont pas stables, et elles ne permettent pas de construire un parcours consultable dans son entier. De plus, si l'utilisateur se déplace de page en page pour y récolter ses informations, c'est que chaque page ne présente pas l'ensemble des informations utiles pour chaque usager. L'ensemble de ces caractéristiques font du site web un labyrinthe où l'utilisateur peut perdre ses repères.

6.4 Recommandations

L'ensemble de nos analyses nous a permis d'élaborer des recommandations. Ces recommandations sont applicables dans deux domaines. Dans un premier temps, elles sont destinées aux recherches en corpus sur les discours administratifs numériques. Nos interrogations méthodologiques et technologiques nous ont permis de mettre en œuvre

une méthodologie novatrice et efficace dans l'analyse d'un discours administratif numérique, que nous pensons reproductible pour d'autres discours similaires. Dans un second temps, nous avons élaboré des recommandations à destination des experts rédigeant ou produisant le discours étudié. Nos analyses ont permis de mettre en lumière plusieurs problématiques importantes du domaine qu'il nous semble intéressant d'aborder.

6.4.1 Recommandations méthodologiques pour les recherches en corpus sur les discours administratifs numériques

La recherche et la méthodologie développées dans cette thèse ont été conçues en tenant compte des spécificités des discours administratifs numériques. Notre objectif est d'introduire ce discours dans la sphère académique afin que davantage de recherches soient menées, notamment en linguistique de corpus. C'est pourquoi nous proposons des recommandations méthodologiques pour les recherches en corpus sur les discours administratifs numériques à partir de nos résultats.

6.4.1.1 Étudier les éléments ne faisant pas partie du contenu

L'adoption d'une méthode inductive pour étudier le corpus a permis de mettre en avant plusieurs caractéristiques saillantes du discours numérique de l'accès aux droits. Une de ces caractéristiques, documentée dans l'état de l'art (section 3.2) et confirmée dans l'étude du corpus, est la délinéarisation du discours et le caractère particulier du site web avec ses différentes pages.

L'étude des discours spécialisés se limite généralement au contenu des sites web, c'est-à-dire à la partie informative du site. Toutefois, nos analyses nous ont conduits à choisir d'étendre le champ de nos recherches aux éléments discursifs ne faisant pas partie du contenu. Quelques chercheurs se sont intéressés aux menus en tant que catégorisation linguistique (Perea, 2012), mais nous avons trouvé peu de recherches en linguistique de corpus et analyse des discours spécialisés intégrant cette composante. Pourtant, notre étude indique qu'il est indispensable de les intégrer afin de réaliser une caractérisation complète du discours, car ces éléments permettent de prendre en compte les éléments interactifs. Ils permettent également une analyse de la structure des sites, ce qui est essentiel pour garantir une arborescence facilement navigable pour l'utilisateur. Nos analyses ont mené à la description du site web en tant que labyrinthe, où le fil d'Ariane, en tant qu'élément caractéristique des sites web, apporte la preuve de leur nature labyrinthique. Les éléments ne faisant pas partie du contenu informatif sont essentiels non seulement pour la reconstruction du fil discursif, mais également pour la compréhension du discours. Certains chercheurs ont intégré cette dimension à leur re-

cherche, à l'image de (Stanfill, 2015) qui élabore la *discursive interface analysis* qui consiste à étudier les fonctionnalités de la page, ses menus, son organisation, ses caractéristiques. En s'appuyant sur les théories de Foucault (1972) concernant le discours, Stanfill étudie la manière dont l'interface est construite, notamment par rapport à son objectif, qui, dans le cas du discours administratif numérique, est sa diffusion aux usagers. C'est la raison pour laquelle nous recommandons la prise en compte d'éléments ne faisant pas partie du contenu dans les recherches en corpus sur les discours administratifs numériques. Nos travaux nous ont permis de déterminer qu'ils sont essentiels et qu'ils doivent faire l'objet de recherche dans le domaine de la terminologie et de l'analyse de discours. En effet, proposer une terminologie claire et compréhensible dans l'arborescence du site est un élément clé de la vulgarisation du discours administratif numérique. Cela permet aux usagers non experts d'accéder dès les débuts de leur navigation à une terminologie stabilisée, que l'on retrouve dans le fil d'Ariane et qui permet à l'utilisateur de se repérer.

6.4.1.2 S'appuyer sur les éléments HTML pour appréhender le discours

Nos travaux prônent également la coopération entre les éléments structurels discursifs et les éléments structurels informatiques, comme le code HTML qui met en forme la page. Le travail de rédacteur web allie deux compétences : la mise en page informatique et la rédaction. En nous intéressant au code HTML, nous avons pu prendre une approche adaptée à l'annotation par mouvements rhétoriques du texte qui ne reposait pas sur le lecteur, mais qui émanait du texte lui-même. Ces éléments structurent la page web et surtout, ils lui confèrent ses caractéristiques numériques. C'est pourquoi il s'agit d'éléments importants que le chercheur doit apprendre à lire et intégrer à ses recherches, notamment dans la création du corpus.

De plus, chaque site présente une convention de mise en page HTML stable, qui peut permettre d'extraire de façon automatique certains éléments. Ces conventions HTML pourraient permettre d'élaborer une méthode d'annotation automatique des unités discursives en nous appuyant sur les balises HTML. À partir de ces balises, nous pourrions réaliser une analyse systématique des sites web sur de grands corpus en prenant en compte leur structure. Nous pourrions ainsi étendre le nombre de parcours de navigation, mais aussi transposer notre méthodologie à d'autres types de discours issus de sites web. Ces parcours de navigation pourraient être enregistrés à partir de l'activité réelle des usagers, puis annotés automatiquement à partir du code HTML.

En conclusion, les éléments HTML constituent un outil précieux pour le chercheur s'intéressant aux discours administratifs numériques, mais également aux discours issus de sites web en général.

6.4.2 Recommandations à destination des experts rédigeant ou produisant le discours étudié

Nous proposons ensuite des recommandations à destination des experts rédigeant ou produisant le discours étudié.

6.4.2.1 Remettre la terminologie et la phraséologie au centre des préoccupations

La simplification du discours administratif ne peut se faire sans une clarification de la terminologie pour les rédacteurs. En effet, les multiples guides de style indiquant qu'il est nécessaire de faire des phrases simples, concises, avec du vocabulaire simple (Service, 2024) ne prennent pas en compte la complexité et la spécialisation des domaines du droit et de l'administration qui régissent le discours. Les questions de simplification et de vocabulaire peuvent être abordées par une approche en langue de spécialité, qui propose des outils concrets pour aider à la compréhension de concepts spécialisés.

L'exploration terminologique et phraséologique a permis de mettre en lumière diverses instabilités dans le discours, et notamment dans l'utilisation de nombreux variants. Ces analyses ont pour but de participer à la caractérisation du discours de l'accès aux droits, qui peut constituer une première étape de simplification. Comme nous l'avons expliqué dans le chapitre 1, la transformation numérique du discours de l'accès aux droits s'est faite sans action de planification sur le long terme, tout du moins au niveau discursif. Il est donc important de répertorier l'ensemble des pratiques actuelles, afin d'identifier les points d'amélioration.

Il est évident que le gouvernement français s'intéresse à la terminologie. La direction interministrielle de la transformation publique (DITP) propose par exemple une émission le 14 mai 2024 portant sur le « jargon » administratif. Certains communicants de l'émission parlent également de « mots pivots » ou encore tout simplement « termes ». Cela montre qu'il s'agit d'une préoccupation d'actualité en France. Le gouvernement britannique a quant à lui pris le tournant du *Plain language* en proposant des conventions de rédactions, mais aussi des actions de simplification par le biais du Government Digital Service. Toutefois, ces travaux se font sans l'appui de méthodologies issues de la recherche en langue de spécialité. La volonté de diminuer la présence de termes et les tournures compliquées sont des stratégies de simplifications sensées et compréhensibles, mais qui pourraient être accompagnées d'outils méthodologiques adaptés, que proposent la terminologie et la phraséologie.

Il est tout d'abord indispensable de faire un état des lieux du discours actuel, que nous avons amorcé dans cette thèse, afin de recenser les pratiques terminologiques et phraséologiques, avant de procéder à sa simplification. Dans le discours médical, Delavigne

(2022) remarquent que les termes ne posent pas toujours problème pour les usagers, car au fur et à mesure de leur familiarisation avec le discours, ils reprennent les termes et les explicitent (Delavigne, 2022, p. 36). La simplification du discours de l'accès aux droits et de sa terminologie pourrait passer, comme nous l'avons montré dans la section 5.1.5, par une ontologie des principaux concepts. De plus, une présentation claire des termes et de leurs variants à l'aide de fiches, accompagnées des arborescences des concepts désignant les dispositifs d'aide, pourrait être utilisée pour guider les utilisateurs dans leur recherche par mot-clé. Enfin, il est indispensable que les éléments interactifs, comme la barre de recherche, prennent en compte ces variations terminologiques afin de présenter des éléments stables, quel que soit le variant choisi.

Le langage est au cœur des problématiques de simplification et la recherche en lange de spécialité peut proposer des réponses adaptées aux problématiques de vulgarisation du discours.

6.4.2.2 Créer des outils d'aide à la navigation

Afin de recomposer les discours, nous pensons qu'il serait utile de créer un outil qui permette de prélever des extraits de textes utiles à l'utilisateur afin de les garder lors de sa navigation. À la manière du presse-papier, l'ensemble des extraits copiés peuvent être présentés dans un onglet ou un document afin que l'utilisateur puisse créer sa propre page et avoir accès à l'ensemble des informations sélectionnées, sous la forme d'un bloc-notes. Cet outil pourrait pallier la délinéarisation du discours posé par le site web.

6.4.2.3 Mettre en place des collaborations avec la recherche en sciences du langage

Cette troisième recommandation est basée sur un article de Labasse (2022) que nous souhaitons mettre en avant car il répond aux enjeux de vulgarisation du discours administratif de l'accès aux droits. Cet article nous a particulièrement intéressée, car il aborde les guides de styles utilisés dans l'administration française. Malgré d'apparents efforts afin de simplifier la communication des administrations, pour l'auteur, ces guides ne font pas preuve de fondements scientifiques solides et utilisent des connaissances « empiriques » (Labasse, 2022, p. 66). S'il est vrai qu'une série de fiches *Simplifier le langage administratif* a été publiée en 2021, celles-ci relèvent plutôt de « bonnes pratiques », comme l'indique la page web (Equipe Sciences comportementales et Direction interministerielle de la transformation publique, 2024). Ces fiches proposent bon nombre de conseils de rédaction qui ne semblent pas s'appuyer sur des recherches réalisées en sciences du langage ou en sciences comportementales.

Cette absence de fondement scientifique se confirme par le manque de collaboration avec la recherche publique française pour l'élaboration de ces discours. Nous recom-

mandons donc la création de collaboration entre chercheurs et organismes d'accès aux droits afin de proposer des solutions de simplification.

6.4.2.4 Multiplier les genres discursifs pour faciliter la compréhension

Enfin, nous pensons qu'une multiplicité des genres discursifs facilite la compréhension et l'accessibilité au discours. La centralisation sur un seul média (le site web) ne peut être efficace, car elle présente des obstacles de compréhension pour certains usagers. Toutefois, cette multiplicité des genres discursifs ne doit pas mener à une multiplicité de plateformes telle qu'elle existe aujourd'hui. Les informations devraient être disponibles sur une plateforme unique, qui pourrait disposer de plusieurs URL, mais qui mènerait à un seul espace. Cette plateforme unique présenterait également un design stable et reconnaissable, comme le site `gov.uk`, sans opérer de changements réguliers comme le fait actuellement le site `service-public.fr`, ce qui peut déranger les usagers.

L'apport d'éléments visuels comme l'ontologie proposée dans la section 5.1.5 peut constituer une aide, tout comme les vidéos proposées par certains organismes. Ces outils devraient être rassemblés sur une plateforme unique, afin de permettre aux usagers de choisir le moyen d'accès aux informations.

De plus, la possibilité de communiquer avec des experts semble essentielle afin de comprendre le discours, et les multiples dispositifs mis en place ne pourront pas remplacer l'efficacité d'une explication d'expert. Nous préconisons de réinstaurer la possibilité de discuter avec des experts. Si la volonté des gouvernements est d'intervenir le moins possible dans les démarches, l'intervention de professionnels experts peut être accompagnée d'interventions d'usagers-experts, notamment par le biais de forums, qui existent déjà de façon informelle, mais qui pourrait gagner en efficacité s'ils étaient proposés par les organismes d'accès aux droits. Ces forums existent déjà sur certains sites, comme le forum ameli pour les assurés[6]. Le forum présente un avantage important pour garantir la compréhension du discours : les messages restent disponibles de façon pérenne, ce qui permet à plusieurs utilisateurs ayant la même question de ne pas multiplier les communications (en envoyant un mail personnalisé par exemple). De plus, cela instaure une politique de transparence des communications qui peut aider les usagers à faire confiance aux divers services d'accès aux droits.

6. `forum-assures.ameli.fr`

Chapitre 7

Conclusion

En prenant pour objet d'étude le discours numérique de l'accès aux droits, cette thèse avait un double objectif. D'une part, il s'agissait d'élaborer une méthodologie d'analyse propre au discours numérique en s'appuyant sur une approche outillée et descriptive. D'autre part, elle visait à adopter une approche discursive, terminologique et phraséologique pour contribuer à la caractérisation du discours numérique de l'accès aux droits. L'ensemble de nos recherches ont été conduites de façon comparative entre le discours de l'accès aux droits britannique et son équivalent français. L'intérêt de cette recherche est également double : non seulement le discours d'accès aux droits a la spécificité d'être un discours nativement numérique, mais il s'agit également d'un discours au cœur de la vie de tous les locuteurs, le seul et principal accès aux droits de tous les citoyens.

Pour ce faire, nous avons retracé les étapes de création de ce genre discursif ainsi que les caractéristiques relevées par la communauté scientifique. Les recherches précédemment réalisées sur la conditionnalité administrative (Willmann, 2012) et numérique (Mazet, 2019), mais aussi sur la charge administrative (Moynihan et al., 2015) et l'État plateforme (Colin et Verdier, 2015), nous ont permis d'identifier les enjeux et les points de difficulté de la communication dans ce domaine, ainsi que leurs répercussions sur le discours. Les recherches en langue de spécialité et en vulgarisation ont permis d'identifier le discours comme étant hybride, entre spécialisation et vulgarisation. En effet, la spécialisation des connaissances issue des discours administratif et juridique confère à ce discours des caractéristiques évidentes d'un discours spécialisé. D'un autre côté, les efforts de vulgarisation et d'explication, visant à diffuser les informations aux usagers et à transmettre les connaissances spécialisées liées à leurs droits, confèrent au discours des caractéristiques propres à la vulgarisation. Ce double statut lui donne une place et des caractéristiques uniques. Les recherches sur l'histoire du discours administratif lié aux évolutions technologiques se sont avérées non seulement utiles mais indispensables pour l'interprétation des données observées en corpus. Au Royaume-Uni, nous avons

noté une tendance à la centralisation des services. En 2011, nous avons recensé la création du Government Digital Service, un service unique dédié à la transition numérique, incluant notamment la gestion des sites dédiés à l'accès aux droits. Depuis 2013, le gouvernement britannique introduit un dispositif d'aide centralisé, le *Universal Credit*, qui rassemble plusieurs dispositifs d'aide. En France, le site `service-public.fr`, responsable de la communication numérique entre les administrations et les usagers, est créé en 2000. Dans les années qui suivent, le gouvernement français ainsi que les organismes multiplient la création de sites (l'administration numérique des étrangers en France en 2014, Welcome to France en 2017).

L'ensemble de ces définitions ont permis de concevoir une méthodologie adaptée au corpus spécialisé tout en intégrant sa dimension numérique et vulgarisée. Pour cela, notre méthodologie s'est développée en deux temps et sur deux plans. Dans un premier temps, les recherches en linguistique de corpus (Biber, 1993; Sinclair, 1991; McEnery et Wilson, 2001) ont permis de définir les critères de création d'un corpus de grande taille afin de réaliser des explorations discursives, terminologiques et phraséologiques. Dans un second temps, les recherches en analyse de discours (Swales, 1990; Moreno et Swales, 2018; Biber et al., 2007) et en analyse de discours numérique (Paveau, 2015) ont permis d'établir un nouveau cadre méthodologique pour étudier un discours spécialisé numérique issu de sites web. La méthodologie des *moves* (ou mouvements rhétoriques) conçue par Swales (1990) ainsi que l'étude du concept de parcours utilisateur (Beauvisage, 2004) ont contribué à l'élaboration d'une méthodologie et d'un corpus complémentaire. L'élaboration de cette méthodologie en deux temps a permis de réaliser une analyse sur plusieurs niveaux. Ces différents niveaux (lexical, syntaxique, rhétorique, phraséologique, etc.) ont permis d'obtenir une vision d'ensemble d'un discours jusqu'ici peu documenté sur le plan linguistique.

De plus, pour chaque langue nous avons construit un corpus qui se compose de trois sous-corpus correspondant aux trois thématiques différentes : droits sociaux, immigration et citoyenneté. Ce choix a permis d'aboutir à une caractérisation plus globale du discours administratif et de poser un double regard contrastif sur les données issues des deux langues et de ces trois thématiques.

À partir des corpus récoltés, nous avons réalisé une étude systématique des termes formés par des groupes nominaux complexes (par ex. *New Style Jobseeker's Allowance*, *Personal Independance Payment* ; **allocation d'aide au retour à l'emploi**, **allocation de logement sociale**). Nous avons également analysé les termes simples (ex. *allowance*, *credit*, **allocation**, **prestation**) et composés (par ex. *standard allowance*, *housing payment* ; **allocation logement**, **prestations familiales**) qui caractérisent également le discours administratif. L'analyse de l'ensemble de la terminologie administrative a montré

une forte tendance à la variation, notamment observable dans les groupes nominaux complexes. Pour la France, nos analyses ont ainsi mis en avant une tendance à l'utilisation des groupes nominaux complexes, et un phénomène de variation terminologique liée aux principaux dispositifs d'aide et de droit au séjour (par exemple, les termes suivants ont été relevé pour désigner le dispositif d'aide pour les personnes à la recherche d'un emploi : **allocations chômage**, **assurance chômage**, **allocation d'aide au retour à l'emploi**, **ARE**, **allocation de retour à l'emploi**). Pour le Royaume-Uni, nous avons également mis en lumière une tendance à la variation terminologique à travers les groupes nominaux complexes désignant les dispositifs d'aide dans le domaine de l'aide sociale (par ex. les termes suivants ont été relevés pour désigner le dispositif d'aide pour les personnes à la recherche d'un emploi : *Jobseeker's Allowance*, *New Style Jobseeker's Allowance*, *JSA*, *contribution-based Jobseeker's Allowance*, *income-based Jobseeker's Allowance*, *unemployment benefits*). Ce phénomène de variation contribue à une instabilité terminologique dans les deux langues. En revanche, la terminologie du droit au séjour britannique est stabilisée et provient majoritairement du discours juridique. La terminologie issue du thème citoyenneté est quant à elle stable et bien établie dans les deux langues. Ces caractéristiques permettent de déterminer qu'une clarification de la terminologie pourrait améliorer la vulgarisation du discours, notamment pour les dispositifs d'aide relatifs aux droit sociaux dans les deux pays, ainsi que pour le droit au séjour en France. Le phénomène de variation est également accentué par l'utilisation de nombreux sigles et acronymes dans les deux langues. Ces acronymes et sigles renforcent la spécialisation du discours, car l'usager n'a plus accès au terme initial qui est souvent plus transparent. Les explorations nous ont amenée à étudier un grand nombre de concepts et de nombreux variants dans les deux langues. Par conséquent, la première partie de nos résultats propose des commentaires concis et denses pour chaque concept et terme, indispensables pour dégager les problématiques terminologiques. Notre méthodologie nous a non seulement permis d'analyser un nombre de concepts clés du discours administratif, tels que **justificatifs**, **visas**, **titre de séjour**, mais également de relever une terminologie et phraséologie hybride, composées également des termes et motifs propres à internet (par ex. **mon espace**, **mon compte**). De plus, une analyse ontologique a été proposée pour illustrer l'apport possible de l'analyse terminologique et ontologique à la diffusion d'informations vers les usagers. Cette approche classificatoire, qui repose sur nos analyses fines en corpus, pourrait être une solution efficace pour aider les usagers à accéder plus facilement aux connaissances du domaine et, par conséquent, à leurs droits.

L'exploration semi-automatique du corpus a également permis de mettre en avant des caractéristiques syntaxiques, notamment au niveau de l'utilisation des pronoms. Les analyses ont mis en évidence la fréquente utilisation du pronom **vous** en français ou *you*

en anglais, ainsi que les schémas phraséologiques qui l'accompagnent. Ces pronoms montrent la place centrale de l'usager qui est la figure principale de ce discours. L'exploration des pronoms a également révélé l'utilisation singulière des pronoms **je** et **mon**, qui jouent un rôle classificatoire sur certains sites en permettant à l'utilisateur de définir son profil. Ces deux pronoms sont également souvent utilisés avec les termes **compte** ou **espace**, caractéristiques du volet numérique de ces sites web. Dans le corpus anglais, nous avons mis en avant le ton particulier adopté par les municipalités à travers leur utilisation du pronom personnel *we*. L'analyse de ces pronoms, ainsi que les fondements théoriques abordés dans l'état de l'art ont mené à l'étude des schémas conditionnels de type *if you are* ou **si vous êtes** qui jouent également un rôle classificatoire en permettant de définir le profil de l'usager. Enfin, une caractérisation stylistique et rhétorique a révélé le rôle des répétitions en discours. La présence de nombreuses phrases répétées est due au caractère délinéarisé du genre discursif étudié et peut être vue comme une tentative de recomposer une cohérence discursive à travers les différentes pages.

Le dernier chapitre de cette thèse est consacré à la caractérisation discursive du discours par le biais d'un corpus de petite taille annoté en mouvements rhétoriques. Les annotations, réalisées à partir du code HTML, ont permis de dégager la structure générale des pages web de ce genre discursif. Trois éléments structurels sont ressortis de cette analyse : l'en-tête et le pied de page, qui sont des éléments stables, et la partie principale, qui est un élément modulable et personnalisable. Les éléments stables présentent principalement des listes de liens pour aider l'usager à s'orienter sur le site. Ces listes de liens sont composées de syntagmes nominaux et contiennent peu de phrases. L'en-tête et le pied de page font partie des unités identifiées comme étant dédiées à la navigation sur le site et se composent presque exclusivement de liens hypertextes dans les deux langues. Les éléments modulables présentent du contenu informatif ainsi que quelques liens hypertextes, mais dans une proportion moins élevée (environ 14 % en français et 17,5 % en anglais). Le fil d'Ariane est également un élément linguistique et numérique ayant pour but de garantir une cohérence dans le parcours de l'utilisateur. L'étude des mouvements rhétoriques, des liens hypertextes et de la structure des pages a ensuite permis de mettre en avant la nature labyrinthique du site web. Même si ce dernier possède une arborescence, les pages présentent une abondance d'hyperliens en circuit fermé qui peuvent perdre l'usager. Le site web s'apparente donc parfois à un labyrinthe dans lequel l'usager doit trouver son chemin afin d'accéder aux informations. À cette nature labyrinthique de la partie discursive et textuelle des sites web que nous avons étudiée s'ajoute la nature multimodale inhérente de ceux-ci, si bien que le site web est communément comparé à une mosaïque.

L'ensemble de nos résultats ont permis la rédaction de deux types de recommanda-

tions. Dans un premier temps, ces recommandations portent sur la méthodologie pour les recherches en corpus sur les discours administratifs numériques. Elles suggèrent d'étudier l'ensemble des éléments discursifs présents sur les sites en ne se limitant pas au contenu, mais en intégrant les éléments HTML susceptibles d'apporter des clés d'analyse supplémentaires. Il s'agit de tenir compte de la structure particulière des sites web. Elle porte également sur la nécessité de prendre en compte le parcours utilisateur, en complément de l'analyse en corpus. Dans un second temps, nous avons émis des recommandations à destination des experts rédigeant ou produisant le discours étudié. Ces recommandations suggèrent que la terminologie et la phraséologie constituent des outils théoriques pertinents pour la simplification du discours numérique. Cette simplification pourrait être réalisée en collaboration avec les chercheurs et chercheuses en sciences du langage, notamment en France. Enfin, nous avons également indiqué qu'une multiplication des genres discursifs pouvait faciliter la compréhension du discours étant donné que la centralisation sur un seul média (le site web) présente des obstacles de compréhension pour certains usagers. Cette multiplication des genres discursifs ne doit pas s'accompagner d'une multiplication des plateformes comme cela existe aujourd'hui. Au contraire, nous pensons nécessaire de proposer une centralisation sur une plateforme unique de l'ensemble des informations nécessaires et suffisantes présentées dans divers genres discursifs.

Le cheminement réalisé dans cette thèse nous a permis de répondre à nos questions de recherche, auxquelles nous allons apporter des réponses claires et succinctes dans cette conclusion.

— Comment élaborer une méthodologie d'analyse pour dégager les caractéristiques, à la fois terminologiques et phraséologiques, d'un discours nativement numérique et au cœur de la vie de tous les locuteurs ?

Il est nécessaire d'élaborer une méthodologie en deux étapes. Dans un premier temps, un grand corpus récolté selon une méthode plus classique permet de mettre en avant les caractéristiques du discours, comme la variation terminologique, la conditionnalité et les phrases répétées. Dans un second temps, un petit corpus spécialisé centré sur le parcours utilisateur et annoté en fonctions discursives permet de dégager les caractéristiques d'un discours nativement numérique, notamment dans ses éléments interactifs et navigationnels.

— En combinant une approche outillée et descriptive, quelles problématiques terminologiques et phraséologiques peuvent être mises en lumière à travers l'analyse

de ce discours ?

Les résultats des analyses réalisées en corpus ont mis en lumière plusieurs problématiques. Tout d'abord, l'étude des termes principaux du domaine montre qu'il existe, dans les deux langues, un phénomène de variation autour de certains dispositifs.

Cette variation est accentuée par le caractère numérique de la page web et par les différentes fonctions de chaque partie de la page (titre, sous-titre, contenu, menu déroulant, etc.). Des termes généraux sont utilisés dans les titres (*Jobseeker's Allowance* ou **allocation chômage**), tandis que les sigles sont utilisés dans les explications détaillées (*JSA* ou **Are**). Les analyses ont également révélé des problématiques phraséologiques. Nous avons identifié un grand nombre de phrases répétées sur différentes pages des sites. Ces phrases répétées sont caractéristiques d'un discours nativement numérique et jouent différents rôles, notamment en participant à la cohésion du discours délinéarisé par la construction en arborescence du site. L'étude des pronoms a permis de mettre en avant des schémas conditionnels qui servent à définir le profil de l'usager ainsi que ses droits. Ces schémas conditionnels seraient issus du discours administratif et ne sont *a priori* pas caractéristiques d'un discours numérique.

Enfin, l'étude de la phraséologie a permis de dégager certaines problématiques autour des organismes d'accès aux droits. En français, nous avons identifié une phraséologie et des collocations distinctes entre les deux organismes, Pôle emploi [1] et la Caf. La Caf utilise plus souvent le pronom **vous**, tandis que Pôle emploi préfère **je**. Dans les deux cas, une prosodie sémantique positive se dégage autour de ces organismes, qui cherchent tous deux à impliquer l'usager. Au Royaume-Uni, les municipalités adoptent un ton singulier en utilisant le pronom *we*. Là où les autres sites étudiés se positionnent rarement en sujet, les municipalités montrent leur implication dans les démarches et décisions. Dans les deux langues, la modalité est très présente avec les collocations **vous devez** ou **vous pouvez** en français et les modaux *can*, *will* et *must* en anglais.

— Comment une approche discursive, terminologique et phraséologique peut-elle contribuer à caractériser efficacement le discours administratif numérique ?

L'approche adoptée a permis de révéler des problématiques terminologiques et phraséologiques claires. L'identification des points de difficulté dans le discours est la première étape pour une simplification efficace. À notre connaissance, les efforts de simplification précédents n'avaient pas inclus d'état des lieux du discours. En observant l'utilisation de la terminologie en contexte, nous avons pu identifier l'utilisation des variants issus de différentes législations sur la même page ou dans le même texte. De plus, l'approche par mouvements rhétoriques a permis d'identifier des éléments de cohérence

1. Depuis le 1er janvier 2024, Pôle emploi est devenu France Travail.

et de cohésion comme les hyperliens, le fil d'Ariane, les phrases répétées, mais aussi des éléments de délinéarisation (éléments interactifs, construction en arborescence). L'ensemble de ces approches ont permis de recenser des caractéristiques essentielles du discours. Rappelons néanmoins que pour parvenir à cette caractérisation, notre étude a montré également la nécessité de combiner l'analyse en corpus avec des recherches sur les aspects socio-historiques des services administratifs. Cette approche permet d'interpréter les données de manière appropriée et d'aboutir ainsi à une caractérisation précise et de qualité du discours.

— Comment peut-elle apporter des éléments concrets pour aider à la création de ressources à destination des différents acteurs et usagers ?

L'approche adoptée a permis deux avancées essentielles à la création de ressources adaptées aux besoins des usagers. Toute d'abord, nous avons identifié une méthodologie adaptée aux discours administratifs numériques qui pourra être étendue à d'autres types de sites et de discours. Cette méthodologie permettra ainsi de généraliser la caractérisation d'autres discours administratifs numériques. Nous avons également identifié qu'il pourrait être possible de mettre en place une annotation automatique des parcours utilisateur afin de réaliser des études à plus grande échelle. Dans un second temps, les analyses et conclusions en terminologie et phraséologie pourraient permettre de réaliser un guide de style pour les rédacteurs, ou encore de mettre en place des ressources ontologiques à destination des usagers comme nous l'avons proposé dans la section 5.1.5. Ces approches ont donc apporté des éléments d'analyses clairs et concrets qui pourront servir de socle à la création de ressources.

— Quels sont les types de locuteurs concernés par ce discours ?

L'étude de la spécialisation et de la vulgarisation des discours a mené à une redéfinition de l'expertise. L'expertise est un rôle qu'endosse un individu face à un non-expert (Lelebina et Sardas, 2011). Cet individu a acquis des connaissances qui lui permettent de maîtriser un sujet (Ericsson, 2014). Dans le cas de l'accès aux droits, nous avons identifié plusieurs types d'expertises, portées par plusieurs types d'experts. Les professionnels possèdent un certain type d'expertise (législateurs, rédacteurs, fonctionnaires et médiateurs). Les militants et usagers possèdent un autre type d'expertise issue de leur expérience, parfois qualifiée de contre-expertise (Trombert, 2013). Ces deux expertises se complètent et permettent d'obtenir une connaissance transversale du domaine. Les locuteurs et les experts du discours de l'accès aux droits numérique sont donc multiples.

Cette thèse a proposé une vision d'ensemble du discours grâce à une méthodologie

innovante. Néanmoins, cette approche comporte plusieurs limites auxquelles nous avons dû faire face. Le corpus réalisé à partir de parcours utilisateur représente une étude pilote destinée à valider la méthodologie. En effet, nous n'avons réalisé l'étude que sur quatre parcours par langue. De plus, les parcours utilisateur étant simulés, il serait plus intéressant d'obtenir des parcours réels, relevés auprès d'utilisateurs. Il nous paraissait toutefois plus raisonnable de réaliser une première étude afin d'évaluer les résultats obtenus en utilisant ce type de corpus. L'annotation manuelle de ces corpus a également été une opération chronophage, qui a nécessité une attention particulière lors de nos travaux de recherche. Cette annotation étant créée à partir du code HTML, il pourrait être envisageable d'automatiser le processus d'annotation afin de diminuer le risque de biais et d'accélérer le processus.

Les limites de nos recherches se trouvent également dans le manque de données concernant le discours de l'accès aux droits prénumérique. Nous avons trouvé peu de recherches documentant ce discours avant sa transformation numérique. Cela peut être expliqué par l'explosion de la linguistique de corpus qui a eu lieu en même temps que la transition numérique, mais également par la place de l'oral dans le discours prénumérique. En effet, la communication écrite entre l'administration et ses usagers s'est largement développée, tandis que la communication orale était beaucoup plus fréquente dans le discours de l'accès aux droits prénumérique. Une étude diachronique du discours permettrait d'étudier son évolution et de confirmer quelles caractéristiques sont apparues avec le numérique et quelles caractéristiques sont inhérentes au domaine. Néanmoins, nous avons décidé de commencer par une caractérisation de l'existant que nous considérions comme un point de départ essentiel.

L'étude de ce corpus présente également des limites inhérentes au discours numérique natif. Les données étudiées sont éphémères et en constante évolution, si bien que les textes ne sont pas stables. Les rédacteurs modifient régulièrement les textes. Ces modifications peuvent être dues à de nouvelles législations, comme nous l'avons vu pour Pôle emploi qui est devenu France Travail en 2024. Elles sont parfois dues à un changement dans le guide de style ou les consignes de rédaction. Enfin, elles peuvent tout simplement témoigner d'un effort de simplification pour un texte qui n'était pas clair. Il existe bien sûr un système permettant de consulter la date de modification de la page, mais aucune information n'apparaît sur l'extrait de texte modifié. Le corpus étudié est quant à lui figé dans le temps. Il existe donc des disparités entre nos recherches, qui sont basées sur un corpus récolté en 2021, et le discours utilisé en 2024. D'une façon plus générale, la transformation et la réécriture constante des textes présents sur les sites rendent difficile la description du discours.

Enfin, notre étude s'intéresse à un genre discursif, ce qui ne nous permet pas de com-

prendre la langue de spécialité dans son ensemble. Nous n'avons étudié que le discours diffusé par les organismes d'accès aux droits. Toutefois, comme nous l'avons montré dans la section 1.3, les usagers sont également engagés dans la diffusion de ce discours et peuvent être également experts. Les observations réalisées dans le cadre de cette recherche se limitent aux sites web et ne donnent pas d'éléments concernant la diffusion et l'utilisation de la terminologie et de la phraséologie dans la vie des usagers, qui sont les premiers concernés par ce discours.

Nous proposons maintenant quelques perspectives de recherche pouvant répondre aux limites énoncées précédemment. Dans un premier temps, la récolte du corpus à partir de parcours utilisateur pourrait être améliorée. Une étude visant à enregistrer l'activité des usagers sur les sites pourrait permettre de construire des parcours utilisateur plus représentatifs. Il serait alors possible de multiplier les parcours et d'agrandir le corpus. En France, nous pourrions envisager par exemple une collaboration avec des centres sociaux où les usagers rencontrent des écrivains publics ou naviguent indépendamment sur internet pour obtenir des informations. Au Royaume-Uni, ces études pourraient être menées dans des associations comme *Citizen Advice* ou encore dans les bibliothèques municipales. Ces données pourraient également permettre de comparer les parcours utilisateur des experts et des non-experts. Dans un deuxième temps, nous pourrions automatiser une partie de l'annotation par mouvements rhétoriques à partir des balises HTML afin d'annoter de façon plus efficace.

Afin de compléter les analyses réalisées sur le genre discursif étudié dans cette thèse, il pourrait être intéressant d'étudier le discours employé par les usagers. La récolte de corpus à partir de forums ou de groupes Facebook pourrait apporter des informations précieuses sur l'utilisation du discours par les usagers. Nous pourrions ainsi comparer le discours disséminé par les administrations et le discours utilisé par les usagers, notamment l'usage des sigles et des variants.

Enfin, l'ontologie réalisée dans la section 5.1.5 pourra servir de modèle à la réalisation d'ontologies autour de plusieurs concepts. Ces ontologies permettront aux usagers ou aux divers professionnels d'acquérir une connaissance globale des concepts et de leur terminologie.

Les recherches réalisées dans cette thèse ont apporté de la visibilité à ce discours en proposant une vision d'ensemble de ses problématiques. Il reste essentiel de continuer cette caractérisation afin de proposer des actions de simplification efficaces et pertinentes.

Bibliographie

(2000). Arrêté du 4 décembre 2000 portant agrément de la convention du 1er janvier 2001 relative à l'aide au retour à l'emploi et à l'indemnisation du chômage et du règlement annexé à cette convention. [208]

(2019). Décret n° 2019-797 du 26 juillet 2019 relatif au régime d'assurance chômage. [208]

(2021a). Article L314-2 - Code de l'action sociale et des familles. [29]

(2021b). Décret N° 2021-313 du 24 mars 2021 relatif à la mise en place d'un téléservice pour le dépôt des demandes de titres de séjour. [18, 29]

(2022). The Jobseeker's Allowance (Schemes for Assisting Persons to Obtain Employment) (Amendment) Regulations 2022. Publisher : King's Printer of Acts of Parliament. [214]

(2023). Code de l'entrée et du séjour des étrangers et du droit d'asile. [29]

Adam, J.-M. (1997). Genres, textes, discours : Pour une reconception linguistique du concept de genre. *Revue belge de Philologie et d'Histoire 75*(3), 665–681. [4, 47, 272]

Adam, J.-M. (1999). *Linguistique Textuelle : Des Genres de Discours Aux Textes*. Nathan Paris. [46]

Adam, J.-M. (2001). Types De Textes Ou Genres De Discours ? Comment Classer Les Textes Qui "Disent De Et Comment Faire" ? *Langages* (141), 10–27. [5, 47, 49]

Adam, J.-M. (2006). Texte, contexte et discours en questions. *Pratiques 129*(130), 21–34. [46]

Adler, M. (2012). The Plain Language Movement. *The Oxford Handbook of Language and Law* (April 2018), 1–20. [69]

Agar, J. (2016). *The Government Machine : A Revolutionary History of the Computer*. The MIT Press. [20]

Algan, Y. et T. Cazenave (2017). *L'État En Mode Start-Up*, Volume 2e éd. [32]

Anderson, W. J. (2003). *A corpus linguistic analysis of phraseology and collocation in the register of current European Union administrative French*. Thesis, University of St Andrews. Accepted : 2014-06-25T13 :57 :55Z. [6, 95]

Angermüller, J. (2007). L'analyse du discours en Europe. *Analyse du discours et sciences humaines et sociales*, 9–22. [44]

Authier, J. (1982). La mise en scène de la communication dans des discours de vulgarisation scientifique. *Langue Française* (53), 34–47. [5, 66, 75, 87]

Azzopardi, S. (2010). Présentation : La linguistique « de » corpus au-delà des champs disciplinaires : questions et enjeux transversaux. *Cahiers de praxématique* (54-55), 11–24. [100, 101]

Baekgaard, M., D. P. Moynihan, et M. K. Thomsen (2021). Why Do Policymakers Support Administrative Burdens ? The Roles of Deservingness, Political Ideology, and Personal Experience. *Journal of Public Administration Research and Theory 31*(1), 184–200. [31]

Bakhtine, M. (1984). Les genres du discours. In *Esthétique de Ma Création Verbale*. Paris. [4, 47, 48]

Bakhtine, M. (1995). Dialogisme et analyse du discours. *Paris, Bertrand-Lacoste*. [86]

Baroni, M. et M. Ueyama (2006). Building general- and special-purpose corpora by Web crawling. In *Proceedings of the 13th NIJL International Symposium, Language Corpora : Their Compilation and Application*. [106]

Bathia, V. (2004). *Worlds of Written Discourse : A Genre-Based View*, Volume 148. Bloomsbury Academic. [45, 48, 74, 125]

Beacco, J.-C. et S. Moirand (1995). Autour des discours de transmission des connaissances. *Langages 29*(117), 32–53. [46, 48]

Beaudet, C., A. Condamines, C. Leblay, et A. Picton (2016). Rédactologie et didactique de l'écriture professionnelle : un chantier terminologique à mettre en place. *Pratiques. Linguistique, littérature, didactique* (171-172). [123]

Beauvisage, T. (2004). *Sémantique Des Parcours Des Utilisateurs Sur Le Web*. Ph. D. thesis, Université Paris X, Paris. [120, 121, 322]

Bennett, F. (2012). Universal Credit : Overview and gender implications. In *Social Policy Review 24*, Chapter Social Policy Review 24, pp. 15–34. Policy Press. [28]

Bensaïd, A. (2022). Parcours utilisateur et UX : Créez votre scénario d'usage ! [124]

Benveniste, E. (1966). *Problèmes de Linguistique Générale, Tome I : 1939-1964*. Gallimard. [45]

Bhargava, S. et D. Manoli (2012). Why are benefits left on the table ? Assessing the role of information, complexity, and stigma on take-up with an IRS field experiment. *ACR North American Advances*. [31]

Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing 8*(4), 243–257. [104, 109, 110, 322]

Biber, D., C. Ulla, et T. A. Upton (2007). *Discourse on the Move : Using Corpus Analysis to Describe Discourse Structure*. [122, 322]

Bibié-Emerit, L. (2016). La notion de lieu de corpus : Un nouvel outil pour l'étude des terrains numériques en linguistique. *Corela* (14-1). [83, 84, 108]

Bonelli, E. T. (2010). Theoretical overview of the evolution of corpus linguistics. In *The Routledge Handbook of Corpus Linguistics*, pp. 14–28. Routledge London and New York. [xvii, 100, 103, 109]

Bonhomme, M. (2015). La problématique des genres de discours dans la communication sur Internet. *Tranel-Travaux neuchâtelois de linguistique 63*, 31–47. [83, 84]

Bouchet, M. (2023). Repenser l'analyse de corpus numériques issus de sites web dynamiques. In *Actes Des 25èmes Rencontres Jeunes Chercheurs Sciences Du Langage*, Paris. [104, 109]

Boudier, F., F. Bensebaa, et A. Jablanczy (2012). L'émergence du patient-expert : une perturbation innovante. *Innovations 39*(3), 13–25. [55]

Bowker, L. (2010, March). CHAPTER 9. The Contribution of Corpus Linguistics to the Development of Specialised Dictionaries for Learners. In P. A. Fuertes-Olivera (Ed.), *Specialised Dictionaries for Learners*, pp. 155–170. DE GRUYTER. [105]

Bozio, A. et J. Sultan (2021). *La réforme du Universal Credit au Royaume-Uni*. report, Institut des politiques publiques (IPP). [64]

Bracken, M. (2015). Government as a Platform : The next phase of digital transformation. [26]

Brigo, F., W. M. Otte, S. C. Igwe, F. Tezzon, et R. Nardone (2015). Clearly written, easily comprehended ? The readability of websites providing information on epilepsy. *Epilepsy & Behavior 44*, 35–39. [70]

Bronckart, J.-P. (1996). *Activité langagière, textes et discours : pour un interactionisme socio-discursif*. Sciences des discours. Lausanne : Delachaux et Niestlé. [47]

Burden, B. C., D. T. Canon, K. R. Mayer, et D. P. Moynihan (2012). The Effect of Administrative Burden on Bureaucratic Perception of Policies : Evidence from Election Administration. *Public Administration Review 72*(5), 741–751. [30]

Burger, M. (2018). Entre affordances et multimodalité : De nouveaux enjeux pour l'analyse des discours du digital. *Cahiers du Centre de Linguistique et des Sciences du Langage 55*(55), 3–24. [45]

Cabré, M. T. (1999). *Terminology : Theory, Methods, and Applications*. John Benjamins Publishing. [89]

Cacchiani, S. (2018). Webpage usability and utility : Citizens' rights and the law on Gov.Uk. (12), 192–205. [88, 93, 251]

Central Digital and Data Office (2019). Government Design Principles. [32]

Cestnik, B. et A. Kern (2014). Administrative And Digital Literacy : The Legend Or Myth In E-Government. In *Electronic Government and Electronic Participation*, pp. 216–223. IOS Press. [61, 63]

Chappoz, Y. et P.-C. Pupion (2012). Le New Public Management. *Gestion et management public 1/n°2*(2), 1. [23]

Charaudeau, P. et D. Maingueneau (2002). *Dictionnaire d'analyse Du Discours*. Seuil. [45, 86]

Chartier, R. (2005). De l'écrit sur l'écran. In *Les écritures d'écran : histoire, pratiques et espaces sur le Web*, Aix-en-Provence. [80, 81, 85]

Chelle, E. (2019). *Gouverner Les Pauvres : Politiques Sociales et Administration Du Mérite.* (Presses universitaires de Rennes. ed.). [28]

Cheok, S. M., L. M. Hoi, S.-K. Tang, et R. Tse (2022). Crawling Parallel Data for Bilingual Corpus Using Hybrid Crawling Architecture. *Procedia Computer Science 198*, 122–127. [106]

Cimade, L. (2021). Dématérialisation des demandes de titre de séjour : De quoi parle-t-on ? www.lacimade.org/dematerialisation-des-demandes-de-titre-de-sejour-de-quoi-parle-t-on/, Consulté le 23 juin 2024. [18]

Clerc, I. (2022). *Communication écrite État-citoyens : Défis numériques, perspectives rédactologiques*. Presses de l'Université Laval. [123, 124]

Clot, N. F. (2017). Va, vis et deviens : L'expérience utilisateur au coeur de l'évolution des services. In *Développer l'accueil En Bibliothèque, Un Projet d'équipe*, Volume 41 of *Boîte à Outils*. Presses de l'enssib. [121]

Colin, N. et H. Verdier (2015, May). *L'Age de La Multitude. Entreprendre et Gouverner Après La Révolution Numérique*. [25, 26, 32, 55, 321]

Condamines, A. (2005). Linguistique de corpus et terminologie. *Langages 157*(1), 36–47. [44, 74]

Condamines, A. (2010). Variations in terminology : Application to the management of risks related to language use in the workplace. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 16*(1), 30–50. [90]

Conseil d'Administration de l'ANAS (2022). Paye tes allocs ! Déclaration de l'Anas du 4 avril 2022. [35]

Conseil Départemental de l'Accès au Droit (2018). Qu'est-ce que l'accès au droit ? [14, 16]

Coutherut, M. (2014). Les textes procéduraux en anglais : un genre homogène ? [49]

Coutherut, M. (2016). *Les Textes Procéduraux En Anglais : Création d'une Échelle de Prototypicalité*. These de doctorat, Sorbonne Paris Cité. [49]

Coutherut, M. (2021). Ellipses dans les recettes de cuisine en anglais : Un micro-genre procédural concis pour une réalisation plus efficace. In *Réduction – Densification – Elision. Formes Réduites et Leurs Fonctions, Div. Abb & Tabellen, 92-108.* [49]

Cowie, A. (1994). *Phraseology* (Asher, R. E. ed.). Oxford : Oxford University Press. [94]

Davis, B. H. et J. Brewer (1997). *Electronic Discourse : Linguistic Individuals in Virtual Space*. SUNY Press. [81]

De Maeyer, J. et A. E. Holton (2016). Why linking matters : A metajournalistic discourse analysis. *Journalism 17*(6), 776–794. [87]

Defenseur des droits (2022). Dématérialisation des services publics : trois ans après, où en est-on ? Technical report, République Française. [18]

Delavigne, V. (2003). Quand le terme entre en vulgarisation. *Terminologie et Intelligence artificielle*, 80–91. [67]

Delavigne, V. (2017). Term usage and socioterminological variation. *Multiple perspectives on terminological variation 18*, 31. [89]

Delavigne, V. (2022). La notion de domaine en question À propos de l'environnement. *Neologica 2022, n° 16. Néologie et environnement*, 27–59. [46, 56, 74, 271, 318, 319]

Deseilligny, O. (2022). Tapas, oreilles et diktats : négocier l'écriture de l'information à La Croix. *Quaderni 107*(3), 39–55. [209]

d'Etat à la réforme de l'Etat et à la simplification et Le Secrétariat générale pour la modernisation de l'action publique, S. (2015). La Simplification en marche - Synthèse. Technical report, Premier ministre. [72]

Direction interministerielle de la transformation publique et Direction de l'information légale et administrative (2024). Les Bonnes pratiques - Un formulaire simple à remplir, compréhensible et accessible pour tous. Technical report. [1]

Döring, M. (2018). How-to Bureaucracy : Administrative Literacy of Citizens. Technical report. [7, 61, 62]

Drahy, J. (2004). *Le droit contre l'Etat ? : Droit et défense associative des étrangers : l'exemple de la CIMADE*. Logiques Juridiques. L'Harmattan. [18]

Dubois, V., M. Paris, et P.-É. Weill (2018). Des chiffres et des droits. Le data mining ou la statistique au service du contrôle des allocataires. *Revue des politiques sociales et familiales 126*(1), 49–60. [31]

Dury, P. (2019). La 'médicalisation' de la terminologie du textile en anglais, une analyse diachronique basée sur corpus. *Textes et contextes* (13-2). [89]

Dury, P. (2021). L'obsolescence terminologique dans le domaine de la pharmacologie. *Linx. Revue des linguistes de l'université Paris X Nanterre* (82). [89]

Dury, P. et P. Drouin (2010). L'obsolescence des termes en langues de spécialité : une étude semi- automatique de la « nécrologie » en corpus informatisés, appliquée au domaine de l'écologie. [206]

Dury, P. et S. Lervad (2008). La variation synonymique dans la terminologie de l'énergie : approches synchronique et diachronique, deux études de cas. *LSP and professional communication (2001-2008) 8*(2). Number : 2. [6]

Dury, P. et A. Picton (2009). Terminologie et diachronie : vers une réconciliation théorique et méthodologique ? *Revue française de linguistique appliquée XIV*(2), 31–41. [89]

Dwyer, P. (2004). Creeping Conditionality in the UK : From Welfare Rights to Conditional Entitlements ? Technical Report 2. [28]

Dwyer, P., L. Scullion, K. Jones, J. McNeill, et A. B. R. Stewart (2020, March). Work, welfare, and wellbeing : The impacts of welfare conditionality on people with mental health impairments in the UK. *Social Policy & Administration 54*(2), 311–326. [28]

Dwyer, P. et S. Wright (2014). Universal Credit, ubiquitous conditionality and its implications for social citizenship. *Journal of Poverty and Social Justice 22*(1), 27–35. [28]

Eisenlauer, V. (2013). *A Critical Hypertext Analysis of Social Media : The True Colours of Facebook*. London : Bloomsbury Academic, An imprint of Bloomsbury Pub. Plc. [87, 88]

Elberse, J. E., J. F. Caron-Flinterman, et J. E. W. Broerse (2011). Patient–Expert partnerships in research : How to stimulate inclusion of patient perspectives. *Health Expectations 14*(3), 225–239. [55]

Ellimoottil, C., A. Polcari, A. Kadlec, et G. Gupta (2012). Readability of websites containing information about prostate cancer treatment options. *The Journal of urology 188*(6), 2171–2176. [70]

Equipe Sciences comportementales et Direction interministerielle de la transformation publique (2024). Simplifier les documents administratifs. [319]

Ericsson, K. A. (2014). Expertise. *Current Biology 24*(11), R508–R510. [53, 59, 327]

Eshkol-Taravella, I. et N. Grabar (2016). Reformulation à l'oral et dans le forum Web. *SHS Web of Conferences 27*, 11003. [108]

Fairon, C. (2000). GlossaNet : Parsing a web site as a corpus. *Lingvisticæ Investigationes 22*(1-2), 327–340. [83, 106]

Farmer, D. J. (1995). *The Language of Public Administration de David John Farmer* (University Alabama Press ed.). Tuscaloosa, Alabama. [65]

Flowerdew, L. (2010). Using corpora for writing instruction. In *The Routledge Handbook of Corpus Linguistics*, pp. 444–457. Routledge London. [105]

Fox, M. L. (2010). Directgov 2010 and beyond : Revolution not evolution. Rapport d'étude. [21]

François, F., L. Brouwers, H. Naets, et C. Fairon (2014). AMesure : Évaluation du niveau de difficulté des textes. In *Proceedings of TALN 2014 (Volume 2 : Short Papers)*, Marseille. Ministère de la Communauté française. [73]

François, T., A. Müller, B. Degryse, et C. Fairon (2018). AMesure, une plateforme web pour soutenir la rédaction simple de textes administratifs. *Repères-Dorif 16*. [68, 73]

Freixa, J. (2006). Causes of denominative variation in terminology. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication-Terminology / International Journal of Theoretical and Applied Issues in Specialized CommunicationTerminology 12*(1), 51–77. [6, 89, 90]

Frouillou, L., C. Pin, et A. van Zanten (2019). Le rôle des instruments dans la sélection des bacheliers dans l'enseignement supérieur. La nouvelle gouvernance des affectations par les algorithmes. *Sociologie 10*(2), 209–215. [31]

Garcés-Conejos Blitvich, P. et P. Bou-Franch (2019). Introduction to Analyzing Digital Discourse : New Insights and Future Directions. In *Analyzing Digital Discourse : New Insights and Future Directions*, pp. 3–22. Cham : Springer International Publishing. [81, 82]

Gatto, M. (2014). *Web As Corpus : Theory and Practice*. A&C Black. [103, 106, 107]

Gélédan, F. (2016). Spectres du léviathan : L'État à l'épreuve de la simplification administrative (2006-2015). *Revue Francaise d'Administration Publique 157*(1), 33–48. [72]

Gémar, J.-C. (1990). Les fondements du langage du droit comme langue de spécialité. Du sens et de la forme du texte juridique. *Revue générale de droit 21*(4), 717–738. [63]

Genette, G. (1982). *Palimpsestes : la littérature au second degré* (Seuil ed.). [87]

Genevois, B. (2009). Le GISTI : Requérant d'habitude ? La vision du Conseil d'Etat. In *Défendre La Cause Des Étrangers*, Paris, pp. 65–79. [18]

Ghliss, Y. (2019). Les photo-discours WhatsApp : Éléments d'analyse d'une affordance d'une application mobile. *Corela. Cognition, représentation, langage*. [84]

Gilmont, J.-F. (2004). *Une introduction à l'histoire du livre et de la lecture : du livre manuscrit à l'ère électronique*. Editions du CEFAL. [80, 81]

Girardi, C. (2007). Htmcleaner : Extracting the relevant text from the web pages. In *Building and Exploring Web Corpora : Proceedings of the 3rd Web as Corpus Workshop, Incorporating Cleaneval (WAC3)*, Volume 4, pp. 141–143. [107]

Girault, Y. (2017). La dématérialisation au service de la performance de l'administration. *Gestion & Finances Publiques* (6), 81–83. [31]

GISTI (2023). La préfecture du Bas-Rhin condamnée pour la dématérialisation illégale des demandes de titre de séjour [Actions collectives]. [18, 54]

Gledhill, C. et N. Kübler (2016). What can linguistic approaches bring to English for Specific Purposes ? *Archivio Storico Pugliese* (69), 65–95. [5, 51, 74, 94]

Gledhill, C., H. Martikainen, A. Mestivier, et M. Zimina (2019). Towards a Linguistic Definition of 'Simplified Medical English' : Applying Textometric Analysis to Cochrane Medical Abstracts and Their Plain Language Versions. *LCM - La Collana / The Series 9788879169*, 91–114. [69]

Gordon, L. K. (1975). Bureaucratic Competence and Success in Dealing with Public Bureaucracies. *Social Problems 23*(2), 197–208. [5, 61]

Gov.uk Design System team (2024a). Breadcrumbs. [279]

Gov.uk Design System team (2024b). Footer. [283]

Gov.uk Design System team (2024c). Header. [282]

Graber, D. (1981). Political Languages. In *Handbook of Political Communication*. London : Sage Publications. [64]

Graillot, A. (2024). IA, simplification des formulaires, France Services : Gabriel Attal annonce sa feuille de route pour « débureaucratiser » les démarches administratives. [1]

Granger, S. et M. Paquot (2008). Disentangling the phraseological web. In *Phraseology. An Interdisciplinary Perspective*. Amsterdam : John Benjamins Publishing Company. [6, 94]

Gries, S. T. (2008). Phraseology and linguistic theory. In *Phraseology. An Interdisciplinary Perspective*. Amsterdam : John Benjamins Publishing Company. [94]

Guiddens, A. (1998). *The Third Way. The Renewal of Social Democracy.* [28]

Gué, V. (2023). Qu'est-ce qu'un chapô et comment en écrire un ? [209]

Habert, B. (2000). Des corpus représentatifs : De quoi, pour quoi, comment. *Cahiers de l'Université de Perpignan 31*, 11–58. [103, 104, 109]

Hallett, R. W. et J. Kaplan-Weinger (2010). *Official Tourism Websites : A Discourse Analysis Perspective*. Channel View Publications. [83]

Halliday, M. A. K. et R. Hasan (1976). *Cohesion in English*. Number 9 in English language series. London New York : Routledge. [93]

Halliday, M. A. K. et J. J. Webster (2006). *Language of Science : Volume 5*. Collected Works of M.A.K. Halliday. London : Bloomsbury Publishing. [216]

Heaton, R. (2021). Government as a Platform (GaaP) milestones - Government Digital Service. [26]

Herring, S. (1999). Interactional coherence in CMC. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*, Maui, HI, USA, pp. 13. IEEE Comput. Soc. [80, 81]

Herring, S. C. (1992). Gender and Participation in Computer-Mediated Linguistic Discourse. In *Annual Meeting of the Linguistic Society of America*, Philadelphie, pp. 1–12. ERIC. [80]

Heurley, L. (2001). Cinq approches différentes du texte procédural. *Pratiques 111*(1), 39–64. [49]

Holler, R. et N. Tarshish (2022). Administrative Burden in Citizen-State Encounters : The Role of Waiting, Communication Breakdowns and Administrative Errors. *Social Policy and Society*, 1–18. [31]

Humbley, J. (2016). Catherine Resche (dir.), Terminologie et domaines spécialisés, Approches plurielles. Paris : Classiques Garnier, Rencontres 143, Série Linguistique 2, 2015. *ASp. la revue du GERAS 70*, 127–132. [64]

Humbley, J. (2018). *La néologie en terminologie*. Lambert-Lucas. [89, 92, 228]

Humbley, J. et A. Picton (2017). Introduction. In *Multiple Perspectives on Terminological Variation*, pp. 1–10. John Benjamins. [89]

Jakob, J. (2022). Supporting digital discourse ? The deliberative function of links on Twitter. *New Media & Society 24*(5), 1196–1215. [87]

Jakubíček, M., A. Kilgarriff, V. Kovář, P. Rychlỳ, et V. Suchomel (2013). The TenTen corpus family. In *7th International Corpus Linguistics Conference CL*, Lancaster University, pp. 125–127. [101, 106, 107]

Jayaratne, Y. S. N., N. K. Anderson, et R. A. Zwahlen (2014). Readability of websites containing information on dental implants. *Clinical Oral Implants Research 25*(12), 1319–1324. [70]

Jeannot, G. (2020). Vie et mort de l'État plateforme. *Revue française d'administration publique N°173*(1), 165. [25]

Johnson, A. (2018). How we built a tool to check our style guide. `https/::wearecitizensadvice.org.uk/how-we-built-a-tool-to-check-our-style-guide-74ab6df1ab1` Consulté le 17 juillet 2024. [73]

Jouet, E., L. G. Flora, et O. Las Vergnas (2010). Construction et reconnaissance des savoirs expérientiels des patients. *Pratiques de Formation - Analyses 2010*(58-59), olivier_lv. [56]

Kennedy, G. (2014). *An Introduction to Corpus Linguistics*. Studies in Language and Linguistics. New York : Routledge. [99, 100, 103]

Kettl, DF. (2008). The next government of the United States : Why our institutions fail us and how to fix them. [32]

Kilgarriff, A., V. Baisa, J. Bušta, M. Jakubícek, V. Kovár, J. Michelfeit, P. Rychlý, et V. Suchomel (2014). The Sketch Engine : Ten Years on. *Lexicography 1*(1), 7–36. [141]

Kincaid, J. P., R. P. Fishburne Jr, R. L. Rogers, et B. S. Chissom (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. [70]

Krichane, S. et Y. Rochat (2018). Pour une analyse des discours sur le jeu vidéo : l'exemple des « cinématiques ». *Décadrages. Cinéma, à travers champs* (39), 62–92. [83]

Krieg-Planque, A. (2020). Quand la communication publique travaille son expression. *Politiques de communication 14*, 3–34. [64, 68]

Krieg-Planque, A. et C. Oger (2010). Discours institutionnels. Perspectives pour les sciences de la communication. *Mots. Les langages du politique* (94), 91–96. [65]

Kübler, N., A. Mestivier, M. Pecman, et M. Zimina (2016, June). Exploitation quantitative de corpus de traductions annotés selon la typologie d'erreurs pour améliorer les méthodes d'enseignement de la traduction spécialisée. Volume 2, pp. 731. [90]

Kübler, N., M. Pecman, et A. Mestivier (volanschi) (2017). Quand le corpus met K.O. la norme terminologique : le défi permanent des GN complexes anglais en traduction spécialisée. [90]

Labasse, B. (2022). Entre expression et communication, les épineuses évolutions de la parole administrative. *Communication écrite État-citoyens : Défis numériques, perspectives rédactologiques*, 53. [71, 319]

Lafore, R. (2014). L'accès aux droits, de quoi parle-t-on ? *Regards N°46*(2), 21. [15, 16]

Langlois, C. (2002). Le COSLA, une occasion rêvée de simplifier le langage administratif. *L'Actualité Terminologique*. [1]

Langrod, G. (1953). Contribution à l'étude de la terminologie administrative : (Essai d'un schéma analytique de notions fondamentales de l'organisation administrative). *Progress in Public Administration 19*(2), 285–324. [206]

Lavoie, B. et H. F. Nielsen (1999). Web Characterization Terminology & Definitions Sheet. [83]

Le Bolzer, J. (2019). Henri Verdier : « La transformation de l'Etat doit surtout être organisationnelle et managériale » | Les Echos. *Les Echos*. [24]

Ledouble, H. (2023). *Médiatisation de la science et diffusion des connaissances*. ISTE Editions. [89]

Lelebina, O. et J.-C. Sardas (2011). L'Expertise et les experts dans les organisations : Une approche multidisciplinaire pour la définition des notions clés. In *Actes Du XXIIe Congrès de l'AGRH*, pp. 18. [5, 53, 59, 327]

Lerat, P. (1995). *Les langues spécialisées*. Presses Universitaires de France. [5, 51, 74, 105]

Levasseur, A. (2021). L'État français dans la révolution numérique Synthèse et réflexions pour de jeunes juristes. *13 en Droit, Université Sorbonne-Paris-Nord 7*. [24]

L'Homme, M.-C. (2009). DiCoInfo : Dictionnaire fondamental de l'informatique et de l'Internet. *« Hyperlien »*. [86]

L'Homme, M.-C. (2011). Y a-t-il une langue de spécialité ? Points de vue pratique et théorique. *Langues et linguistique, numéro spécial Journées de linguistique* (January 2011), 26–33. [51]

Li, J., L. Liu, et Y. Zheng (2017). Application of User Experience Map and Safety Map to Design Healthcare Service. In V. G. Duffy et N. Lightner (Eds.), *Advances in Human Factors and Ergonomics in Healthcare*, Volume 482, pp. 195–204. Cham : Springer International Publishing. [121]

Lochak, D. (2009). Trente Ans de Contentieux à l'Initiative du GISTI. In *Défendre La Cause Des Étrangers*, Pairs. Dalloz ; Gisti. [18]

Lochak, D. (2022). Le sans contact, nouvelle norme du service public :. *Plein droit n° 134*(3), 3–6. [15, 16, 18, 23, 24, 32]

Lombart, E. et C. Fairon (2020). Spécificités de communication des forums de discussion. *Corpus* (21). [108]

Luján-García, C. (2020). The use of English initialisms and abbreviations in the field of pharmaceutical business communication in Spanish. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 26*(1), 60–81. Publisher : John Benjamins. [92]

Madsen, C. Ø., I. Lindgren, et U. Melin (2022). The accidental caseworker – How digital self-service influences citizens' administrative burden. *Government Information Quarterly 39*(1), 101653. [31]

Mahlberg, M. (2006). Lexical cohesion Corpus linguistic theory and its application in English language teaching. In *Lexical Cohesion and Corpus Linguistics*, Volume 17 of *Benjamins Current Topics*, pp. 103–121. John Benjamins Publishing Company. [93]

Maingueneau, D. (2007). Genres de discours et modes de généricité. *Le français aujourd'hui 159*(4), 29–35. [45]

Maingueneau, D. (2016). L'ethos discursif et le défi du Web. *Itinéraires* (2015-3), 2015–2018. [85, 86, 312]

Maingueneau, D. (2021a). Chapitre 2. La notion de discours. In *Discours et analyse du discours*, Volume 2e éd. of *Collection U*, pp. 11–23. Paris : Armand Colin. [45, 65]

Maingueneau, D. (2021b). *Discours et Analyse Du Discours : Une Introduction.* Armand Colin. [44, 45]

Maniez, F. (2007). Prémodification et coordination : quelques problèmes de traduction des groupes nominaux complexes en anglais médical. *ASp. la revue du GERAS* (51-52). Number : 51-52. [90]

Maniez, F. et P. Thoiron (2004). Les groupes nominaux complexes dans le décodage et la traduction en langue de spécialité : quelles ressources lexicales pour l'apprenant en anglais médical ? *Éla. Études De Linguistique Appliquée 135*(3), 327. [90]

Martinet, A. (1969). Fonctions du langage et linguistique appliquée. *Communication & Langages 1*(1), 9–18. [92, 153]

Mayaffre, D. (2005). Rôle et place du corpus en linguistique. Réflexions introductives. In *Actes Du Colloque JETOU'2005*, pp. 5–17. Université de Toulouse-Le Mirail. [102]

Mayeur, I. et M.-A. Paveau (2020). Présentation. Les devenirs du texte numérique natif. *Corela* (HS-33). [85]

Mazet, P. (2019). Vers l'État plateforme La dématérialisation de la relation administrative. *La Vie des idées*, 1–12. [29, 321]

McEnery, T. et A. Wilson (2001). *Corpus Linguistics*. Edinburgh : Edinburgh University Press. [103, 114, 322]

Mericskay, B. (2021). Le crowdsourcing urbain comme nouvelle forme d'engagement citoyen. *Netcom. Réseaux, communication et territoires* (35-1/2). [27]

Mestivier, A. (2015). Productivity and Diachronic Evolution of Adjectival and Participial Compound Pre-modifiers in English for Specific Purposes. *Fachsprache 37*(1-2), 2–23. [176, 215, 217]

Millot, P. (2019). L'évolution terminologique des sciences de l'information : Étude sur corpus diachronique d'un nouveau domaine spécialisé. *Textes et contextes* (13-2). [102]

Ministère de la Santé et des solidarité (2015). L'usager expert : Acteur de l'amélioration des pratiques professionnelles. [56]

Moreno, A. I. et J. M. Swales (2018). Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes 50*, 40–63. [122, 322]

Morley, J. (2006). Lexical cohesion and rhetorical structure. In *Lexical Cohesion and Corpus Linguistics*, Volume 17 of *Benjamins Current Topics*, pp. 5–22. John Benjamins Publishing Company. [93]

Morris, G. (2016). Universal Credit, Ideology and the Politics of Poverty - ProQuest. *Renewal : a Journal of Social Democracy 24*(3), 76. [64]

Mourad, G. (2006). Nouveaux signes de lecture et d'écriture pour les documents électroniques. *Cahiers GUTenberg* (46-47), 19–27. [276]

Mourlhon-Dallies, F., F. Rakotonoelina, et S. Reboul-Touré (2004). Les discours de l'internet : quels enjeux pour la recherche ? *Les Carnets du Cediscor. Publication du Centre de recherches sur la didacticité des discours ordinaires* (8), 9–19. [81]

Moynihan, D., P. Herd, et H. Harvey (2015). Administrative Burden : Learning, Psychological, and Compliance Costs in Citizen-State Interactions. *Journal of Public Administration Research and Theory 25*(1), 43–69. [30, 31, 321]

Muracciole, M. et D. Massé (2018). L'inclusion sociale numérique : Le cas de trois dispositifs numériques améliorant l'accès aux droits. *Terminal* (122). [62, 67]

Nakbi, K. (2002). La rédactologie : Domaine, méthode et compétences. *ASp. la revue du GERAS* (37-38), 15–26. [123]

O'Keeffe, A. (2010). *The Routledge Handbook of Corpus Linguistics*. Routledge. [105]

O'Reilly, T. (2011). Government as a Platform. *Innovations : Technology, Governance, Globalization 6*(1), 13–40. [25]

O'Donnell, M. (2008). The UAM CorpusTool : software for corpus annotation and exploration. [141]

Paltridge, B. (2012). *Discourse Analysis : An Introduction*. London : Bloomsbury Academic. [43, 45]

Panckhurst, R. (2006). Le discours électronique médié : bilan et perspectives. pp. 345. [80]

Parker, J. (2023). *Analysing the History of British Social Welfare : Compassion, Coercion and Beyond*. Policy Press. [177]

Paveau, M.-A. (2009). Mais où est donc le sens ? Pour une linguistique symétrique. *Res per nomen 2*(Maio), 21–31. [44]

Paveau, M.-A. (2013). Genre de discours et technologie discursive. Tweet, twittécriture et twittérature. *Pratiques. Linguistique, littérature, didactique* (157-158), 7–30. [83, 84, 108]

Paveau, M.-A. (2015). Ce qui s'écrit dans les univers numériques. *Itinéraires* (2014-1). [3, 82, 85, 322]

Paveau, M.-A. (2016). Des discours et des liens. Hypertextualité, technodiscursivité, écrilecture. *Semen. Revue de sémio-linguistique des textes et discours* (42). [87]

Paveau, M.-A. (2017). *L'analyse du discours numérique. Dictionnaire Des formes et des pratiques*. Hermann. [5, 46, 81, 82, 84, 105]

Paveau, M.-A. (2019). Introduction. Écrire, parler, communiquer en ligne : nos vies sociolangagières connectées. *Langage et société 167*(2), 9–28. [82]

Pecman, M. (2004). L'enjeu de la classification en phraséologie. In *Actes du congrès EUROPHRAS 2004*, Université de Bâle. [91]

Pecman, M. (2007). Approche onomasiologique de la langue scientifique générale. *Revue française de linguistique appliquée XII*(2), 79–96. [63]

Pecman, M. (2014). Variation as a cognitive device. *Terminology. International Journal of Theoretical and Applied Issues in Specialized CommunicationTerminology / International Journal of Theoretical and Applied Issues in Specialized CommunicationTerminology 20*(1), 1–24. [6, 89, 90, 93, 104]

Pecman, M. (2018). Langue et construction de connaisSENSes : énergie lexico-discursive et potentiel sémiotique des sciences. pp. 1–265. [107, 128, 266]

Percebois, J. (2001). Fonctions et vie des sigles et acronymes en contextes de langues anglaise et française de spécialité. *Meta : journal des traducteurs / Meta : Translators' Journal 46*(4), 627–645. [92]

Perea, F. (2012). Les sites pornographiques par le menu : pornotypes linguistiques et procédés médiatiques. *Genre, sexualité & société* (7). [107, 316]

Petit, M. (2010). Le discours spécialisé et le spécialisé du discours : Repères pour l'analyse du discours en anglais de spécialité. *E-rea* (8.1). [50, 51]

Pic, E. et G. Furmaniak (2012). Du discours spécialisé au discours vulgarisé : approche grammaticale. *ASp. la revue du GERAS* (61), 39–54. [248, 252]

Picton, A. (2018). Terminologie outillée et diachronie : éléments de réflexion autour d'une réconciliation. *ASp. la revue du GERAS* (74), 27–52. [102]

Pinède, N., D. Reymond, B. Le Blanc, et V. Lespinet-Najib (2011). Une étude terminologique de la communication hypertexte web. Caractéristiques du domaine universitaire en perspective diachronique. In *Proceedings of the 9th International Conference on Terminology and Artificial Intelligence*, Paris, France, pp. 139–142. [107]

Piven, F. F. et R. Cloward (2012). Regulating the Poor : The Functions of Public Welfare. [31]

Pointurier, S. (2016). *Théories et pratiques de l'interprétation de service public*. Presses de la Sorbonne Nouvelle. [7]

Pomikálek, J. (2011). Removing boilerplate and duplicate content from web corpora. *Disertacnı práce, Masarykova univerzita, Fakulta informatiky*. [107]

Quilan, O. How we're making it easier to access government forms online – Government Digital Service. `https://gds.blog.gov.uk/2024/01/12/how-were-making-it-easier-to-access-government-forms-online/`, Consulté le 3 août 2024. [2]

Rastier, F. (1989). Linguistique et recherche cognitive. *Histoire Épistémologie Langage 11*(1), 5–31. [48]

Rastier, F. (1995). Le terme : entre ontologie et linguistique. *La banque des mots* (7), 35–65. [91]

Reboul-Touré, S. (2004). Sciences, Médias et Société : 11. Écrire la vulgarisation scientifique aujourd'hui. In *Sciences, Médias et Société*, Lyon. Sciences, Médias et Société. [66, 67, 75]

Rey, G. (2008). Le rôle de la phraséologie dans la mise en discours de la langue juridique—F. In *Aspectos formales y discursivos de las expresiones fijas*, pp. 121–140. Peter Lang publishing group Alemania. [6, 93, 230, 264]

Richard, I. (2015). Tautologies en anglais juridique : Termes atypiques au service du droit. pp. 101–124. Classiques Garnier. [64]

Roche, C. (2008). Le terme et le concept : fondements d'une ontoterminologie. arXiv :0801.1275 [cs]. [92]

Roche, C. (2012). Ontologie : entre terminologie et connaissance de spécialité. In *GLAT - Terminologies : textes, discours et accès aux savoirs spécialisés*, Terminologies : textes, discours et accès aux savoirs spécialisés, Genova, Italy. Aula Meridiana Facoltà di Giurisprudenza de l'Università di Genova. [91]

Romain, C., É. Tortochot, et V. Rey (2022). Performativité des choix de simplification par un rédacteur professionnel lors de la révision de documents administratifs numériques. *Communication écrite État-citoyens : Défis numériques, perspectives rédactologiques*, 97. [68]

Rossetti, A., P. Cadwell, et S. O'Brien (2020). "The Terms and Conditions Came Back to Bite" : Plain Language and Online Financial Content for Older Adults. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 12426 LNCS*(April 2021), 699–711. [6, 69]

Rouleau, M. (1995). La langue médicale : une langue de spécialité à emprunter le temps d'une traduction. *TTR : traduction, terminologie, rédaction 8*(2), 29–49. [51, 74]

Rouski, M. (2018). Marqueurs référentiels déictiques et anaphoriques dans le discours juridique. *Romanica Cracoviensia 18*(2), 89–96. [247]

Roy, S. (2013). Le langage clair en droit : Pour une profession plus humaine, efficace, crédible et prospère ! *Les Cahiers de droit 54*(4), 975. [63, 69]

Rychlý, P. (2008). A Lexicographer-Friendly Association Score. *RASLAN*, 6–9. [127]

Sabri, O., N. El Hana, et Z. Yasmine (2021). CCI Paris-Île-De-France : Créer ses buyers personas à partir de focus groups pour un parcours utilisateur efficace. [121]

Sadick, G. (2022). Utiliser des données associatives pour la recherche sur l'asile : L'exemple de la Cimade. In *Semaine Data-SHS*, Paris, France. Plateforme universitaire de données de Nanterre and Plateforme universitaire de données des Grands Moulins and MSH Mondes and UP Cité. [18]

Saemmer, A. (2010). Approche semio-rhétorique des couplages texte-mouvement dans le discours numérique. *Document numérique*, 25. [81]

Sager, J. C. (2002). *A Practical Course in Terminology Processing*. John Benjamins Publishing Company. [6]

Samraj, B. (2014). Move structure. In *Pragmatics of Discourse*, Volume 3, pp. 385–405. Walter de Gruyter GmbH & Co KG. [122]

Schäfer, R., A. Barbaresi, et F. Bildhauer (2013). The Good, the Bad, and the Hazy : Design Decisions in Web Corpus Construction. In *8th Web as Corpus Workshop*, Proceedings of WAC-8, Lancaster, United Kingdom, pp. 7–15. ACL SIGWAC. [104]

Schiffrin, D. (1994). *Approaches to Discourse*. Oxford : Blackwell Publishing Ltd. [46]

Schiffrin, D., H. E. Hamilton, et D. Tannen (2015). *The Handbook of Discourse Analysis*. John Wiley & Sons. [4, 45, 74]

Service, G. D. On Clear Writing. [70]

Service, G. D. (2024). Content design : planning, writing and managing content - Writing for GOV.UK - Guidance - GOV.UK. [318]

Service d'Information du gouvernement (2024a). En-tête - Header - Système de Design de l'État. [281]

Service d'Information du gouvernement (2024b). Fil d'Ariane - Breadcrumb - Système de Design de l'État. [279]

Service d'Information du gouvernement (2024c). Pied de page - Footer - Système de Design de l'État. [282]

Service d'information du Gouvernement (SIG) et Direction de l'information légale et administrative (DILA) (2022). Qu'est-ce que l'accès au droit ? [15]

Silletti, A. M. (2016). Le discours hypertextualisé des magazines de vulgarisation scientifique français en ligne. *Semen. Revue de sémio-linguistique des textes et discours* (42). [83, 85, 87]

Simon, J. (2015). Le discours hypertextualisé. *H2PTM'15 : Le numérique à l'ère de l'Internet des objets, de l'hypertexte à l'hyper-objet*, 3. [86, 87]

Simon, J. (2018). Introduction : Analyse du discours hypertextualisé : théories, méthodes, interdisciplinarité. In *Le discours hypertextualisé : Espaces énonciatifs mosaïques*, Annales littéraires, pp. 7–17. Besançon : Presses universitaires de Franche-Comté. Code : Le discours hypertextualisé : Espaces énonciatifs mosaïques. [86]

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press. [95, 100, 322]

Sinclair, J. (2005a). Corpus and Text. Basic principles. *Developing Linguistic Corpora : a Guide to Good Practice*. [106]

Sinclair, J. (2005b). The phrase, the whole phrase, and nothing but the phrase. *Plenary lecture presented at Phraseology*. [206]

Sintomer, Y. (2008). Du savoir d'usage au métier de citoyen ? *Raisons politiques 31*(3), 115–133. [55]

Sitri, F. et C. Barats (2017). Chapitre II. Constituer un corpus en analyse du discours, un moment crucial. In *Méthodes et outils informatiques pour l'analyse des discours*, Didact Méthodes, pp. 41–62. Rennes : Presses universitaires de Rennes. [44]

Souchier, E. (1996). L'écrit d'écran, pratiques d'écriture & informatique. *Communication & Langages 107*(1), 105–119. [80, 81]

Souchier, E., Y. Jeanneret, et J. Le Marec (Eds.) (2003). *Lire, écrire, récrire : Objets, signes et pratiques des médias informatisés*. Éditions de la Bibliothèque publique d'information. [86]

Stanfill, M. (2015). The interface as discourse : The production of norms through web design. *New Media and Society 17*(7), 1059–1074. Publisher : SAGE Publications Ltd. [317]

Storrer, A. (2000). Was ist" Hyper" am Hypertext. *Sprache und neue Medien 222249*. [88, 312]

Studer, R., V. R. Benjamins, et D. Fensel (1998). Knowledge engineering : Principles and methods. *Data & Knowledge Engineering 25*(1), 161–197. [91]

Swales, J. M. (1990). *Genre Analysis : English in Academic and Research Settings*. Cambridge : Cambridge University Press. [45, 47, 52, 122, 322]

Swales, J. M. (2016). Reflections on the concept of discourse community. *ASp. la revue du GERAS* (69), 7–19. [52, 53, 74]

Teubert, W. (2009). La linguistique de corpus : une alternative . *Semen. Revue de sémio-linguistique des textes et discours* (27), 185–211. [99]

Tiersma, P. (2008). *The Nature of Legal Language*, Volume 5. John Benjamins Publishing Amsterdam/Philadelphia. [64]

Topper, A. (1997). Comparing Face-to-Face and Electronic Discourse : Issues and Questions Raised in a Research Study. Technical report. [80]

Trombert, C. (2013). Expertise professionnelle et contre-expertise militante dans l'accès aux droits sociaux : Tension à front renversé autour du général et du singulier. *SociologieS*. [54, 55, 56, 327]

UK Parliament (2022). The Cost of War. [20]

Vaupot, S. (2020). La traduction des sites web touristiques, vers la localisation ou la culturalisation ? *DELTA : Documentação de Estudos em Lingüística Teórica e Aplicada 36*, 2020360208. [83]

Vecchiato, S. (2019). Parcours de simplification. Éduquer à la clarté textuelle dans la communication sur la santé. *Éla. Études de linguistique appliquée 195*(3), 343–362. [49]

Vial, S. (2014). Critique du virtuel : en finir avec le dualisme numérique. *Psychologie Clinique 37*(1), 38–51. [82]

Vicari, S. (2021). Rando, trek, promenade… des mots pour quel(s) sport(s) ? De la construction du discours d'autorité dans les blogs de tourisme d'aventure. *Revue italienne d'études françaises. Littérature, langue, culture* (11). [83]

Vicentini, A. (2003). The economy principle in language. *Notes and Observations from early modern english grammars. Mots, Palabras, Words 3*, 37–57. Publisher : Citeseer. [92]

Vitali-Rosati, M. (2014). *Égarements. Amour, mort et identités numériques*. Hermann. [85, 121]

Weiss, A. E. (2019a). Automating, 1980s : Arthur Andersen and the Operational Strategy. In *Management Consultancy and the British State*, pp. 149–198. Cham : Springer International Publishing. [20]

Weiss, J. (2019b). The rise and fall of UK digital government – learning from the past. [20]

Weller, J.-M. (2018). Les figures de l'usager dans les réformes de modernisation des services publics. *Informations sociales 198*(3), 48–56. [55]

Wellman, C. P. (2015). Welfare Rights. In *Wiley Encyclopedia of Management*, pp. 1–2. Chichester, UK : John Wiley & Sons, Ltd. [15, 16]

Wigham, C. R. et C. Poudat (2020). Corpus complexes et standards : Un retour sur le projet CoMeRe. *Corpus* (20). [108]

Willmann, C. (2012). Politiques de l'emploi, prestations sociales : existe-t-il un modèle vertueux ? À propos de la « conditionnalité » des revenus de remplacement et autres minima sociaux. *Droit Social 1*, 79–86. [27, 321]

Wilson, J. (2005). Political Discourse. In *The Handbook of Discourse Analysis*, Chapter 20, pp. 398–415. John Wiley & Sons, Ltd. [64]

Wolton, D. (1997). De la vulgarisation à la communication. *Hermès* (1), 9–14. [66]

Annexes

.1 Exemples de l'utilisation des termes autour du concept de *unemployment benefits*

Termes	Occ.	Contexte	Source
New Style Jobseeker's Allowance (JSA)	4	If you have worked as an employee and paid enough National Insurance contributions you may also be able to get either **New Style Jobseeker's Allowance (JSA)** or New Style Employment and Support Allowance (ESA)	understanding universal credit.gov.uk
'new style' JSA	112	Apply for **'new style' JSA**	gov.uk
		Make a claim for **'new style' JSA** or contribution or income-based JSA	gov.uk
		To be entitled to **'New Style' JSA** you must have paid national insurance contributions over a certain number of years, normally 2 full tax years before the year you're claiming in.	falkirk.gov.uk
new style JSA	33	Apply for **new style JSA**	gov.uk
		If you're claiming **new style JSA** you'll need to check if you can challenge your sanction under the Universal Credit system	londoncouncils .gov.uk
income-based Jobseeker's Allowance (JSA)	77	To get a Budgeting Loan you must have been getting one or more of these benefits for the past 6 months : — Income Support — **Income-based Jobseeker's Allowance** — Income-related Employment and Support Allowance — Pension Credit	gov.uk

Termes	Occ.	Contexte	Source
		There's a limit to how much they can take if you get : — Pension Credit — **Income-based Jobseeker's Allowance** — Income-related Emplyment Support — Allowance — Income Support	understanding universal credit.gov.uk
		Universal Credit has replaced **income-based Jobseeker's Allowance** (JSA) for most people	londoncouncils .gov.uk
income-based JSA	151	If your **income-based JSA** claim is ending because you're making a new claim for Universal Credit, you'll automatically continue to get the amount of JSA you currently receive, as long as you're still eligible.	gov.uk
		Make a claim for 'new style' JSA or contribution or **income-based JSA**	gov.uk
		Income-Based JSA will require you to meet the following criteria :	jobcentreguide .co.uk
contribution-based JSA	24	If your partner is claiming **contribution-based JSA**, your training will not affect their claim or reduce the amount paid to them.	gov.uk
		If you also receive **contribution-based JSA** /ESA, this will be included in your additional payment.	understanding universal credit.gov.uk
Jobseeker's Allowance	326	Explore the topic **Jobseeker's Allowance** and low income benefits	gov.uk

Termes	Occ.	Contexte	Source
		You can apply for **Jobseeker's Allowance** (JSA) to help you when you look for work.	gov.uk
		Where you can claim benefits **Jobseeker's Allowance** Maternity and childcare benefits	gov.uk
Unemployment benefits	15	Ideally, a Jobcentre gives out advice on the available jobs and also facilitates **unemployment benefits** to British citizens.	jobcentreguide .co.uk
		It's voluntary - unless you've been out of work and claiming **unemployment benefits** for 24 months.	gov.uk
Income Support	439	Disability premium payments can be added to your : — **Income Support** income-based Jobseeker's Allowance (JSA) — income-related Employment and Support Allowance (ESA) — housing benefit	gov.uk
		We do not make a deduction if the non-dependant is : — under 18 years old getting pension credit (either kind) (for Housing Benefit), — under 25 years old and getting **Income Support**, — income-based Jobseeker's Allowance, is in the assessment phase of income related Employment and Support Allowance	bristol.gov.uk
		If you or your partner get **Income Support**, income-related ESA or income based JSA, your claim can be backdated for up to a month.	londoncouncils .gov.uk
Employment and Support Allowance	256	You can apply for **Employment and Support Allowance** (ESA) if you have a disability or health condition that affects how much you can work.	gov.uk

Termes	Occ.	Contexte	Source
		Related content Disability Living Allowance (DLA) for adults **Employment and Support Allowance** (ESA) Financial help if you're disabled	gov.uk
new style Employment and Support Allowance	4	check if you can claim **new style Employment and Support Allowance** (ESA)	londoncouncils.gov.uk
Employment Support allowance	33	Most people on a low income will claim for Universal Credit, but you might also be able to claim for **Employment Support Allowance** (ESA) or Statutory Sick Pay (SSP).	bristol.gov.uk
Work allowance	23	If you earn more than your **Work Allowance**, your Universal Credit payment will be reduced.	gov.uk
JSA [2]	371	You do not need to go to **JSA** interviews and appointments at the moment because of coronavirus (COVID-19).	gov.uk
		Use a benefits calculator to check how much **JSA** you can get, and how your other benefits will be affected.	gov.uk

2. (sans parenthèses ni compléments)

Termes	**Occ.**	**Contexte**	**Source**
		Common forms of income are : — money from a private pension — money you get from the State Pension — most earnings from an employer or from being self-employed - your earnings will be worked out as an average if they go up and down over the year — benefits such as **JSA** or ESA — You'll also need to consider what savings and investments you have.	londoncouncils .gov.uk
income-related JSA	3	You can get both contribution-based JSA and **income-related JSA** at the same time.	londoncouncils .gov.uk

TABLE 1 – Exemples de l'utilisation des termes autour du concept de *unemployment benefits*

.2 Exemples de l'utilisation des termes autour du concept d'allocation chômage

Terme	Occ.	Contexte	Source
allocation chômage	571	Le salaire journalier de référence (SJR) , qui sert à calculer vos **allocations chômage** , est déterminé en fonction d'un salaire annuel de référence.	service-public.fr
		Vous pouvez estimer le montant de vos **allocations chômage** en utilisant ce simulateur	pole-emploi.fr
		Prime exceptionnelle et **allocation chômage** : quelle articulation ?	paris.fr
allocation d'aide au retour à l'emploi	43	La durée de versement de l' **allocation d'aide au retour à l'emploi** (ARE) dépend de votre durée d'affiliation lors des 36 derniers mois.	service-public.fr
		Si vous exercez une activité professionnelle tout en déclarant être toujours à la recherche d'un emploi, vous pouvez cumuler votre **allocation d'aide au retour à l'emploi** (ARE) avec des revenus d'activité professionnelle.	service-public.fr
ARE	474	Pour toucher l'**ARE** , vous devez être involontairement privé d'emploi.	service-public.fr
		Si votre **ARE** journalière est supérieure à un salaire antérieur journalier brut en moyenne de 159,68 € (soit 4 857,81 € mensuel brut en moyenne), une réduction de 30 % est appliquée à partir du 7e mois de versement.	service-public.fr
Allocation de retour à l'emploi	19	Le montant de l'ASP est alors au moins équivalent au montant de l' **allocation de retour à l'emploi** (ARE) que la salarié aurait perçu s'il n'avait pas bénéficié du CSP.	service-public.fr

Terme	Occ.	Contexte	Source
		En principe, vous n'avez pas droit à l'allocation chômage (**allocation de retour à l'emploi** - ARE) lorsque vous démissionnez.	service-public.fr
Assurance chômage	85	Vous bénéficiez d'une allocation chômage avant votre expatriation. Vous conservez vos droits à l'**assurance chômage** pendant 3 ans.	service-public.fr
		La rémunération mensuelle est équivalente au montant de l'allocation d'**assurance chômage** (ARE).	service-public.fr
		Si vous reprenez une activité professionnelle en France et la perdez ultérieurement, vous pourrez percevoir l'allocation chômage en France et vous pourrez faire valoir vos périodes à l' **assurance chômage** déjà acquis dans un autre pays européen (EEE + Suisse).	service-public.fr

TABLE 2 – Exemples de l'utilisation des termes autour du concept d'allocation chômage

LISTE DES ÉLÉMENTS SOUS DROITS

Liste de tous les éléments retirés de la version complète de la thèse faute d'en détenir les droits

Illustrations, figures, images...

Légende de l'image	**N° de l'image**	**Page(s) dans la thèse**
Capture d'écran d'une rubrique du site France Travail (anciennement Pôle emploi)	5.16	261