



UCLouvain
Faculté de philosophie et lettres
Département d'études romanes

Les unités phraséologiques : un phénomène linguistique complexe ?

Séquences (semi-) figées construites avec les verbes
prendre et *donner* en français écrit L1 et L2
Approche descriptive et acquisitionnelle

Thèse de doctorat présentée
en vue de l'obtention du grade
de docteur en langues et lettres

par Catherine BOLLY

Directeur de thèse :	Jean René KLEIN, Professeur émérite (UCLouvain)
Président du jury :	Paul-Augustin DEPROOST, Professeur (UCLouvain)
Membres du jury :	Inge BARTNING, Professeur (Université de Stockholm)
	Cédric FAIRON, Professeur (UCLouvain)
	Sylviane GRANGER, Professeur (UCLouvain)
	Béatrice LAMIROY, Professeur (KULeuven)

Louvain-la-Neuve
2008

Ce travail s'inscrit dans le cadre d'une *Action de recherches concertées* financée par la Communauté française de Belgique. Le champ d'étude de ce projet collectif, intitulé *Apprentissage des langues étrangères: niveaux phraséologique et discursif* (ARC 03/08-301), est l'appropriation des langues étrangères (français, néerlandais et anglais) d'un point de vue phraséologique et discursif, essentiellement en situation de production écrite.

Remerciements

Il émanera certainement de cette page quelque pigment de ma personnalité, qui n'avait d'ailleurs pas sa place ailleurs qu'ici-même. Ce pigment, qui s'est effacé au fur et à mesure que les pages blanches se sont noircies, s'exprime enfin en remerciant les nombreuses personnes qui m'ont aidée, de près ou de loin, à chercher, à trouver, à perdre, à rechercher et à retrouver, pour permettre finalement à cette thèse de voir le jour.

Mes remerciements vont tout d'abord aux membres de l'équipe du projet ARC qui m'ont encadrée pendant ces quatre années de travail, et en particulier à mon promoteur de thèse, Jean Klein, pour ses lectures assidues et la patience dont il a fait preuve durant ces derniers mois, ainsi qu'à Liesbeth Degand, pour la confiance qu'elle me témoigne depuis plus de cinq ans et qui me pousse à croire que la linguistique se conjugue parfaitement à la première personne du féminin maternel.

Je remercie ensuite les personnes qui ont contribué à l'élaboration de ma recherche, à savoir le Pr. Fr. Thyryon (UCLouvain) pour son aide à la constitution du corpus CAFLaM, le Pr. M. Goyens, L. Melis et Y. Sturbeaux (KULeuven) pour l'extension du corpus FRIDA, le CENTAL et L. Degand (UCLouvain) pour m'avoir transmis les corpus de presse. Je remercie également le Pr. M. Francard et le Pr. S. Granger (UCLouvain) d'avoir accepté de mettre à ma disposition leurs corpus respectifs, Valibel et FRIDA. Merci aussi à mon papa et à mon frère, pour le risque qu'ils ont pris de se frotter à la langue opaque des linguistes. Merci à Nicole, Papa, Frédéric, Olivier, Antoin', Hélène, Cécile, Monique et David, qui ont passé de longues heures à s'écorcher les nerfs sur des phrases fumeuses truffées des verbes *prendre* et *donner*.

Merci enfin à mes proches, famille et amis, qui ont égayé mes journées de labeur en me rappelant de temps à temps à la vie. Merci à mes collègues Susanne, Magali et Fanny F., pour leurs conseils et leur amitié. Merci à mes parents et à mes beaux-parents pour leur soutien sans faille et leur disponibilité sans fin. Merci surtout à Christophe, mon compagnon de toujours et de tous les jours, pour sa présence (souvent silencieuse) et sa patience dans mes absences (souvent capricieuses). Il aura incarné son personnage de papa moderne à merveille... Merci à mes deux filles d'amour, Jeannou et Lili, mi-anges mi-petits-monstres, qui ont rythmé ces années de doctorat de leur premier cri, de leurs premiers pas, de leurs petits et grands bobos, de leurs pleurs nocturnes et de leurs rires diurnes : leur force de vie m'a aidée à porter ce projet de thèse jusqu'au bout.

Table des matières

Table des matières	p.1
Liste des tableaux	p.5
Table des figures	p.9
Table des graphiques	p.9
Liste des abréviations	p.11
Introduction	p.13
Chapitre I: Qu'est-ce qu'une unité phraséologique ?	p.17
I.1. Phraséologie au sens strict (approche fonctionnelle)	p.18
I.1.1. Critères de figement	p.23
I.1.1.1. Mémorisation et institutionnalisation	p.25
I.1.1.2. Polylexicalité	p.25
I.1.1.3. Fixité : figement syntaxique	p.28
I.1.1.4. Non compositionnalité : figement sémantique	p.29
I.1.1.5. Restriction paradigmatique : figement lexical	p.30
I.1.2. Variations : variante, modification et défigement	p.30
I.1.3. Classification des unités phraséologiques (UP)	p.33
I.1.3.1. Phraséologie idiomatique et parémiologique	p.36
I.1.3.2. Phraséologie collocationnelle (définition fonctionnelle)	p.37
I.1.4. Degré de figement des unités phraséologiques : étude pilote en L1	p.45
I.1.4.1. Méthode et corpus	p.45
I.1.4.2. Postulats et hypothèses de travail	p.46
I.1.4.3. Analyse paramétrique : opérationnalisation des données et résultats	p.47
1) Variables extralinguistiques et sociolinguistiques	p.48
2) Variables formelles et syntaxiques : degré de fixité	p.48
3) Variables sémantiques : degré d'opacité	p.49
4) Variables lexicales : degré de restriction lexicale	p.51
I.1.4.4. Traitement des données et interprétation des résultats	p.52
I.1.4.5. Conclusion	p.56
I.2. Phraséologie au sens large (approche statistique)	p.57
I.2.1. Combinaisons syntagmatiques récurrentes	p.59
I.2.2. Cooccurrences et collocations (définition statistique)	p.60
I.2.3. Séquences récurrentes	p.62
I.2.4. Prosodie sémantique	p.63
I.2.5. Colligations et collocations grammaticales	p.63
I.2.6. Patrons, cadres collocationnels et collostructions	p.65
I.3. Conclusion intermédiaire	p.67

Chapitre II : Quelle compétence phraséologique pour les apprenants

du français L2 ?	p.69
II.1. Acquisition d'une L2	p.70
II.1.1. Appropriation, acquisition et apprentissage	p.70
II.1.2. Langue maternelle/ première (L1) vs langue étrangère/ seconde (L2)	p.72
II.1.3. Interlangue (IL)	p.74
II.1.4. Dynamisme et instabilité du processus acquisitionnel (L2)	p.75
II.1.4.1. Axe développemental	p.78
1) Ordre et séquences d'acquisition	p.78
2) Itinéraires acquisitionnels et profils d'apprenants avancés en français L2	p.79
a. Stades d'acquisition	p.79
b. Variété avancée	p.82
II.1.4.2. Axe variationnel	p.85
1) Facteurs extralinguistiques vs intralinguistiques	p.85
2) Âge, aptitude, motivation et attitude	p.87
II.2. Analyse de déviations	p.90
II.2.1. Norme vs usage	p.90
II.2.1.1. Norme subjective vs norme objective (L1)	p.90
II.2.1.2. Normes pédagogiques (NP) (L2)	p.92
1) Définition	p.93
2) Rôle des erreurs dans l'élaboration des normes pédagogiques	p.94
3) Pluralité des normes natives et multi-target model	p.95
II.2.1.3. Conclusion	p.95
II.2.2. Analyse d'erreurs	p.98
II.2.2.1. Repérage/ identification des erreurs	p.99
II.2.2.2. Description des erreurs	p.100
II.2.2.3. Explication des erreurs	p.102
II.2.2.4. Evaluation des erreurs	p.104
II.2.3. Analyse contrastive interlangagière	p.106
II.2.4. De l'erreur à la déviance : agrammaticalité, non-acceptabilité, inappropriété et étrangeté	p.108
II.3. Précision, complexité et fluidité	p.112
II.3.1. Triple processus acquisitionnel en production d'une L2	p.112
II.3.2. Précision/ Exactitude	p.114
II.3.3. Complexité	p.115
II.3.4. Fluidité	p.117
II.4. Acquisition des unités phraséologiques en L2	p.118
II.4.1. Acquisition du lexique en L2	p.118
II.4.2. Compétence phraséologique en L2	p.120
II.4.2.1. Itinéraire acquisitionnel des unités phraséologiques en L2	p.120
II.4.2.2. Etudes acquisitionnelles sur le français L2	p.123
1) Rôle des unités phraséologiques dans la structuration et la fluidité verbale	p.123
2) Etude acquisitionnelle et comparative sur le français parlé L2	p.124
II.4.2.3. Etudes acquisitionnelles sur l'anglais L2	p.126
1) Medium langagier et types de texte	p.126
2) Aptitude, attitude, motivation et intégration socioculturelle	p.128
3) Constructions verbo-nominales	p.129
a. Unités phraséologiques construites avec des verbes à haute fréquence	p.130
b. Déviations collocationnelles dans des textes académiques d'apprenants adultes	p.132
c. Déviations collocationnelles dans des textes argumentatifs/ descriptifs d'apprenants universitaires	p.138
II.5. Conclusion intermédiaire	p.140

Chapitre III : Analyse sur corpus d'apprenants – Méthode et corpus	p.141
III.1. Objectifs de recherche	p.142
III.2. Hypothèses de travail	p.144
III.2.1. Compétence phraséologique	p.144
III.2.2. Déviations phraséologiques	p.145
III.3. Méthode d'analyse paramétrique et statistique de données de corpus	p.148
III.3.1. Opérationnalisation des données	p.149
III.3.2. Manipulation et traitement statistique des données	p.150
III.4. Bases de données et corpus pour le français L1	p.153
III.4.1. Liste des expressions verbales figées du LADL	p.153
III.4.2. Base lexicale distributionnelle <i>Les Voisins De Le Monde</i>	p.154
III.4.3. Corpus de presse <i>Le Soir</i> (1997) et corpus oral <i>Valibel</i>	p.156
III.5. Corpus d'apprenants en français L2	p.158
III.6. Données de corpus	p.164
Chapitre IV : Résultats 1 – Analyse collocationnelle et analyse de déviance	p.169
IV.1. Trois outils d'analyse	p.170
IV.1.1. Liste des expressions du LADL	p.170
IV.1.2. Indice d'information mutuelle (IM)	p.177
IV.1.3. Fréquence dans l'usage L1	p.179
IV.2. Identification des UP	p.183
IV.2.1. Sélection des SVN à haut potentiel phraséologique	p.183
IV.2.2. Exclusion des SVN libres	p.188
IV.3. Identification et description des UP déviantes	p.193
IV.3.1. Reconstructions natives et types de déviations	p.195
IV.3.2. Localisation des déviations	p.208
IV.3.3. Degré de gravité des UP déviantes	p.213
IV.3.3.1. Nombre de déviations par UP	p.214
IV.3.3.2. Hiérarchisation en types de déviations	p.216
IV.3.3.3. Hiérarchisation par catégories grammaticales	p.217
IV.3.3.4. Collocabilité de l'UP dans l'usage natif	p.218
IV.3.3.5. Fréquence de l'UP dans l'usage natif	p.220
IV.3.3.6. Degré global de gravité	p.221
IV.4. Conclusion intermédiaire	p.225
Chapitre V : Résultats 2 – Variables linguistiques et extralinguistiques	p.227
V.1. Caractéristiques formelles	p.228
V.1.1. <i>Prendre vs donner</i>	p.228
V.1.1.1. Suremploi et sous-emploi	p.228
V.1.1.2. Perspective explicative : système flexionnel	p.232
V.1.1.3. Mésusage	p.234
V.1.1.4. Synthèse des résultats	p.240
V.1.2. Profil syntaxique des syntagmes nominaux	p.241
V.2. Caractéristiques sémantiques	p.248
V.2.1. Classification sémantique	p.249
V.2.1.1. Noms concrets, vivants et abstraits	p.252
V.2.1.2. Emplois prédicatifs : états, événements et actions	p.254
V.2.2. Profil sémantique des UP	p.257
V.3. Medium langagier et types de textes	p.264
V.3.1. Code oral	p.264
V.3.2. Types de textes	p.268
V.3.2.1. Distribution des UP	p.269
V.3.2.2. Profil syntaxique et sémantique des UP	p.271
V.3.2.3. UP avec <i>prendre vs donner</i>	p.276
V.4. Conclusion intermédiaire	p.278

Chapitre VI : Résultats 3 – Analyse de précision et de complexité	p.283
VI.1. Analyse de précision phraséologique	p.284
VI.2. Analyse de complexité phraséologique	p.288
VI.2.1. Complexité structurelle	p.289
VI.2.2. Diversité lexicale	p.291
VI.2.3. Structures collocationnelles	p.293
VI.3. Conclusion intermédiaire	p.311
Chapitre VII : Synthèse et perspectives	p.317
VII.1. Synthèse des résultats	p.317
VII.1.1. Etude pilote sur le degré de figement (français L1)	p.318
VII.1.2. Analyse contrastive interlangagière sur corpus d'apprenants informatisés (français académique L2 vs L1)	p.318
VII.1.2.1. Analyse collocationnelle (français académique L2 vs L1)	p.318
VII.1.2.2. Analyse de déviance (français académique L2 vs L1)	p.321
VII.1.2.3. Analyse de précision et de complexité phraséologiques (français argumentatif L2 vs L1)	p.322
VII.1.2.4. Analyse de structures collocationnelles (français argumentatif L2 vs L1)	p.322
VII.2. Perspectives explicatives	p.325
VII.2.1. Transfert des verbes nucléaires	p.325
VII.2.2. Complexité linguistique du phénomène phraséologique dans la LC	p.327
VII.2.3. Facteurs extralinguistiques	p.327
VII.3. Perspectives didactiques	p.328
Conclusion	p.331
Bibliographie	p.333
Description des annexes	p.343

Liste des tableaux

Tableau 1. Critères de figement	p.24
Tableau 2. Critère de polylexicalité – Morphologie et phraséologie	p.27
Tableau 3. Définition fonctionnelle de la collocation, à partir de François & Manguin (2006)	p.41
Tableau 4. Etude pilote sur le degré de figement en français L1 – Description des corpus	p.45
Tableau 5. Degré de figement de quatre SVN représentatives des seuils fixés entre catégories de figement	p.54
Tableau 6. Variabilité du degré de figement de la SVN type construite avec <i>prendre + forme</i>	p.55
Tableau 7. Cooccurents syntaxiques les plus significatifs pour le lemme <i>amoureux</i> (<i>Les Voisins De Le Monde</i>)	p.60
Tableau 8. Séquences récurrentes (trigrammes) construites avec <i>prendre</i> et un constituant nominal postverbal (<i>WordSmith Tools 4.0</i>), par ordre décroissant de fréquence	p.62
Tableau 9. Collexèmes montrant la plus forte attraction collustruotionnelle par rapport à la construction [N <i>waiting to happen</i>] (Extrait de la Table 4 dans Stefanowitsch & Gries, 2003: 219)	p.67
Tableau 10. Itinéraire de développement d'après Bartning & Schlyter (2004: 294) et Granfeldt & Nugues (2007: 4) (Légende : - = pas d'occurrences ; app = apparaît ; prod = productif niveau avancé)	p.81
Tableau 11. Facteurs d'acquisition en L2 – Apprenant, enseignant et tâche	p.86
Tableau 12. Analyse sur corpus d'apprenants – Description des corpus non natifs et natif	p.160
Tableau 13. Distribution des séquences à structure verbo-nominale dans les corpus non natifs et natifs	p.168
Tableau 14. Exemple de correspondance formelle stricte entre l'expression de la liste du LADL et la SVN tirée de nos corpus	p.173
Tableau 15. Exemple de correspondance, désambiguïsée au moyen du cotexte, entre une SVN tirée de notre corpus et deux expressions de la liste du LADL	p.174
Tableau 16. Légende des annotations pour l'examen de la correspondance formelle entre les expressions figées de la liste du LADL et les données de corpus	p.174
Tableau 17. Exemple de correspondance avec variante formelle entre l'expression tirée de la liste du LADL et la SVN tirée de notre corpus	p.175
Tableau 18. Exemple de correspondance partielle entre l'expression tirée de la liste du LADL et la SVN tirée de notre corpus	p.176
Tableau 19. Extrait de la liste paramétrique – Correspondance avec la liste des expressions du LADL	p.176
Tableau 20. Moyenne d'IM des SVN phraséologiques (de la liste des expressions du LADL) avec <i>prendre</i> et <i>donner</i>	p.178
Tableau 21. Couple de cooccurents syntaxiques et couples concurrents pour l'UP <i>donner Dét même possibilité de N à X</i> (occ. 488)	p.179
Tableau 22. Extrait de la liste paramétrique – Fréquences absolues et relatives des SVN dans l'usage L1	p.182
Tableau 23. Fréquence absolue (FqA) et relative (FqR), dans l'usage L1, d'une SVN construite avec le verbe <i>donner</i> et le lemme nominal <i>explication</i>	p.182
Tableau 24. Couple de cooccurents syntaxiques, couples concurrents et fréquence dans l'usage natif pour la séquence construite avec <i>prendre + voie</i> (occ. 191)	p.189
Tableau 25. Couple de cooccurents syntaxiques et couples concurrents, avec leur IM, pour les séquences construites avec <i>prendre + indépendance</i> (occ. 659 et 583)	p.190

Tableau 26. Extrait de la liste paramétrique – Statut phraséologique ou libre des SVN	p.192
Tableau 27. Distribution des SVN phraséologiques <i>vs</i> libres dans les corpus non natifs et natifs	p.193
Tableau 28. Extrait de la liste paramétrique – Types de déviance : formelle, contextuelle, quantitative ou multiple	p.196
Tableau 29. Distribution des UP déviantes <i>vs</i> non déviantes dans les corpus non natifs et natifs	p.206
Tableau 30. Distribution des types de déviations phraséologiques dans les corpus non natifs et natifs	p.207
Tableau 31. Extrait de la liste paramétrique – Localisation de la déviance en termes de catégories grammaticales	p.208
Tableau 32. Distribution des types de déviations phraséologiques dans les corpus non natifs et natifs	p.210
Tableau 33. Tableau détaillé des valeurs attribuées aux types de localisations de la déviance	p.212
Tableau 34. Extrait de la liste paramétrique – Variable recodée pour les déviations portant sur le verbe	p.214
Tableau 35. Extrait de la liste paramétrique – Nombre total de déviations par UP	p.215
Tableau 36. Extrait de la liste paramétrique – Degré de gravité de l'UP déviante en fonction du nombre total de déviations par UP	p.215
Tableau 37. Extrait de la liste paramétrique – Degré de gravité de l'UP déviante en fonction du type de déviance phraséologique	p.216
Tableau 38. Extrait de la liste paramétrique – Degré de gravité de l'UP déviante en fonction des catégories grammaticales	p.218
Tableau 39. Extrait de la liste paramétrique – Degré de gravité de l'UP déviante en fonction de l'IM en L1	p.219
Tableau 40. Extrait de la liste paramétrique – Degré de gravité de l'UP déviante en fonction de sa fréquence relative en L1	p.221
Tableau 41. Extrait de la liste paramétrique – Degré global de gravité de l'UP déviante	p.222
Tableau 42. Distribution des UP déviantes dans les corpus non natifs et natifs, en fonction de leur degré global de gravité	p.223
Tableau 43. Détail de l'analyse d'une UP déviante dont le degré de gravité est élevé (occ. 185)	p.224
Tableau 44. Détail de l'analyse d'une UP déviante dont le degré de gravité est moyen (occ. 923)	p.224
Tableau 45. Détail de l'analyse d'une UP déviante dont le degré de gravité est faible (occ. 916)	p.225
Tableau 46. Extrait de la liste paramétrique – Variables formelles	p.228
Tableau 47. Distribution des UP dans les corpus non natifs et natifs, en fonction du verbe à haute fréquence	p.229
Tableau 48. Distribution des 10 verbes en <i>_ER</i> les plus fréquents en L1, dans les corpus non natifs et natifs	p.233
Tableau 49. Distribution des 10 verbes en <i>_ENDRE</i> les plus fréquents en L1, dans les corpus non natifs et natifs	p.233
Tableau 50. Distribution des catégories flexionnelles en <i>_ER</i> et en <i>_ENDRE</i> dans les corpus non natifs et natifs	p.233
Tableau 51. Distribution des UP construites avec les verbes <i>prendre vs donner</i> , en fonction de leur statut déviant ou non déviant	p.235
Tableau 52. Distribution des UP déviantes construites avec les verbes <i>prendre et donner</i> dans les corpus non natifs et natif	p.236
Tableau 53. Distribution des UP déviantes construites avec <i>prendre vs donner</i> , en fonction du type de déviance	p.237
Tableau 54. Extrait de la liste paramétrique – Variable formelle : structure syntaxique des SN postverbaux	p.241

Tableau 55. Distribution des structures syntaxiques du SN des UP dans les corpus non natifs et natifs	p.243
Tableau 56. Profil syntaxique des syntagmes nominaux dans les UP construites avec <i>prendre vs donner</i> en français L1	p.244
Tableau 57. Distribution des structures syntaxiques du SN des UP dans le corpus anglophone (<i>vs</i> natif) en fonction du verbe utilisé	p.246
Tableau 58. Distribution des structures syntaxiques du SN des UP dans le corpus néerlandophone (<i>vs</i> natif) en fonction du verbe utilisé	p.247
Tableau 59. Extrait de la liste paramétrique – Classification sémantique générale	p.251
Tableau 60. Extrait de la liste paramétrique – Classification sémantique des N_Concrets, N_Vivants et N_Abstraits	p.252
Tableau 61. Extrait de la liste paramétrique – Classification sémantique des N_Prédicatifs d'état affectif/ sentimental et psychologique/ relationnel	p.254
Tableau 62. Extrait de la liste paramétrique – Classification sémantique des N_Prédicatifs d'état de qualité/ propriétés	p.255
Tableau 63. Extrait de la liste paramétrique – Classification sémantique des N_Prédicatifs d'événement	p.256
Tableau 64. Extrait de la liste paramétrique – Classification sémantique des N_Prédicatifs d'action physique	p.256
Tableau 65. Extrait de la liste paramétrique – Classification sémantique des N_Prédicatifs d'activité mentale	p.257
Tableau 66. Distribution des 906 SVN dans les corpus non natifs et natifs, en fonction de leur répartition en catégories sémantiques	p.258
Tableau 67. Fréquence des 703 UP des corpus non natifs et natifs, en fonction de leur répartition en catégories sémantiques	p.259
Tableau 68. Distribution des 646 UP prédicatives (sans les actions physiques) dans les corpus non natifs et natifs	p.260
Tableau 69. Profil syntaxique des syntagmes nominaux dans les UP construites avec <i>prendre vs donner</i> en français L1	p.261
Tableau 70. Distribution des catégories sémantiques prédicatives des UP (sans les actions physiques) dans le corpus anglophone (<i>vs</i> natif) en fonction du verbe utilisé	p.263
Tableau 71. Distribution des catégories sémantiques prédicatives des UP (sans les actions physiques) dans le corpus néerlandophone (<i>vs</i> natif) en fonction du verbe utilisé	p.263
Tableau 72. Extrait de la liste paramétrique – Variables extralinguistiques	p.264
Tableau 73. Extrait de la liste paramétrique – Variables extralinguistiques : types de texte académique	p.269
Tableau 74. Distribution des structures du SN des UP dans les textes argumentatifs anglophones (fréquence absolue), par comparaison avec l'ensemble du corpus anglophone et le corpus natif	p.272
Tableau 75. Distribution des UP prédicatives (sans les actions physiques) dans les textes argumentatifs anglophones (fréquence absolue), par comparaison avec l'ensemble du corpus anglophone et le corpus natif	p.275
Tableau 76. Distribution des UP prédicatives (sans les actions physiques) dans les textes argumentatifs néerlandophones (fréquence absolue), par comparaison avec l'ensemble du corpus néerlandophone et le corpus natif	p.276
Tableau 77. Distribution des UP dans les corpus non natifs et natifs, en fonction du verbe à haute fréquence	p.276
Tableau 78. Distribution des UP avec <i>prendre vs donner</i> dans le corpus anglophone, en fonction du type de texte	p.277
Tableau 79. Distribution des UP avec <i>prendre vs donner</i> dans le corpus néerlandophone, en fonction du type de texte	p.277
Tableau 80. Analyse de complexité et de précision – Description des corpus de textes argumentatifs	p.283
Tableau 81. Analyse de précision phraséologique dans les textes argumentatifs non natifs et natifs	p.286

Tableau 82. Comparaison de la mesure de précision pour 1000 mots et de la mesure de précision « améliorée » pour 1000 mots, pour les textes argumentatifs non natifs et natifs	p.286
Tableau 83. Analyse de complexité structurelle phraséologique – Pourcentage d’UP à structure potentiellement contrainte (<i>vs</i> flexible) dans les textes argumentatifs non natifs et natifs	p.290
Tableau 84. Analyse de complexité structurelle pour les UP avec <i>prendre vs donner</i> – Pourcentage d’UP à structure potentiellement contrainte (<i>vs</i> flexible) dans les textes argumentatifs non natifs et natifs	p.291
Tableau 85. Analyse de diversité polylexicale – Indice de Guiraud (IG) pour les UP construites avec <i>prendre et donner</i> dans les textes argumentatifs des corpus non natifs et natifs	p.292
Tableau 86. Structures collocationnelles à contenu prédicatif d’activité mentale, avec déterminant « zéro », dans les textes argumentatifs des corpus non natifs et natifs	p.299
Tableau 87. Structures collocationnelles à contenu prédicatif d’événement/achèvement, avec déterminant « zéro », dans les textes argumentatifs des corpus non natifs et natifs	p.300
Tableau 88. Structures collocationnelles à contenu prédicatif d’activité mentale, en <i>N de N</i> , dans les textes argumentatifs des corpus non natifs et natifs	p.301
Tableau 89. Structures collocationnelles à contenu prédicatif d’activité mentale, avec proposition subordonnée complément du nom, dans les textes argumentatifs des corpus non natifs et natifs	p.303
Tableau 90. Structures collocationnelles avec un modifieur quelconque, construites avec les noms <i>décision, chance, choix et occasion</i> , dans les textes argumentatifs des corpus non natifs et natifs	p.305
Tableau 91. Structures collocationnelles à contenu prédicatif d’activité mentale d’opportunité, avec ou sans proposition subordonnée complément du nom, dans les textes argumentatifs des corpus non natifs et natifs	p.306
Tableau 92. Structures collocationnelles prépositionnelles à contenu prédicatif d’activité mentale, dans les textes argumentatifs des corpus non natifs et natifs	p.308
Tableau 93. Structures collocationnelles dans les corpus argumentatifs non natifs, présentant un écart de fréquence par rapport à leur fréquence d’emploi dans le corpus natif	p.314

Table des figures

Figure 1. Typologie phraséologique et classification fonctionnelle des unités phraséologiques (UP)	p.34
Figure 2. Tableau des données dans le logiciel <i>SPSS (Data View)</i>	p.150
Figure 3. Aperçu des variables dans le logiciel <i>SPSS (Variable View)</i>	p.151
Figure 4. Liste des 11 cooccurrents syntaxiques dont l'indice de collocabilité est le plus significatif pour le nom <i>opportunité</i>	p.155
Figure 5. Liste des 14 voisins distributionnels les plus significatifs pour le nom <i>opportunité</i>	p.155
Figure 6. Extrait de la concordance effectuée pour la SVN prendre place dans les corpus de français L1 (<i>Le Soir</i> et <i>Valibel</i>)	p.157
Figure 7. Extrait de la concordance effectuée pour le verbe <i>prendre</i> dans les corpus non natifs et natifs	p.166
Figure 8. Extrait de la concordance obtenue pour le verbe <i>donner</i> dans le corpus <i>Valibel</i>	p. 180
Figure 9. Extrait du fichier brut ayant servi de texte-source à la concordance obtenue pour les SVN incluant le verbe <i>donner</i> dans le corpus <i>Valibel</i>	p.180
Figure 10. Sélection des « corpus » de concordances <i>Le Soir</i> et <i>Valibel</i> comme textes-sources de la concordance des SVN construites avec le verbe <i>donner (WordSmith Tools 4.0)</i>	p.181
Figure 11. Extrait de la concordance obtenue pour les SVN construites avec le verbe <i>donner</i> et le lemme nominal <i>explication</i> dans les corpus <i>Le Soir</i> et <i>Valibel</i>	p.181
Figure 12. Fonction <i>Transform – Recode into Different Variables</i> du logiciel <i>SPSS</i>	p.214
Figure 13. Fonction <i>Transform – Compute Variable</i> du logiciel <i>SPSS</i> , utilisée pour calculer le degré de gravité global des UP déviantes	p.222
Figure 14. Classification sémantique « à la Vendler » des SVN construites avec <i>prendre</i> et <i>donner</i>	p.250

Table des graphiques

Graphique 1. Répartition des 200 SVN construites avec <i>prendre</i> en fonction de leur degré de figement	p.55
Graphique 2. Distribution des types de textes académiques dans les corpus non natifs, anglophone et néerlandophone, et dans le corpus natif	p.163
Graphique 3. Distribution des différentes catégories sémantiques dans l'ensemble des corpus	p.251

Liste des abréviations

Phraséologie

CSR	Combinaison syntagmatique récurrente
CVS	Construction à verbe support
DF	Degré de figement
SF	Séquence figée
SL	Séquence libre
SQF	Séquence quasi-figée
SSF	Séquence semi-figée
SVN	Séquence verbo-nominale
UP	Unité phraséologique

Acquisition

IL	Interlangue
L1	Langue première
L2	Langue seconde ou étrangère
LC	Langue cible
LE	Langue étrangère
LM	Langue maternelle
LN	Locuteur natif
LNN	Locuteur non natif
LS	Langue source
NP	Norme pédagogique

Corpus

FL1-pilote	Corpus en français langue première utilisé pour l'étude pilote sur le degré de figement
FL1	Corpus en français langue première constituant le corpus de contrôle de l'analyse sur corpus d'apprenants du français L2
FL2-AN	Corpus d'apprenants anglophones du français L2
FL2-NL	Corpus d'apprenants néerlandophones du français L2

Statistique

FqA	Fréquence absolue
FqR	Fréquence relative
IM	Information mutuelle
M	Million

Introduction

Il est actuellement admis que les *unités phraséologiques* ou *séquences (semi-) figées*, en tant qu'unités polylexicales conventionnelles dont les constituants co-apparaissent de manière contrainte dans l'usage, constituent des obstacles à l'appropriation d'une langue étrangère ou seconde jusqu'à un niveau avancé d'acquisition. Partant de ce constat, nous tenterons de déterminer, dans cette étude, en quoi les unités phraséologiques construites avec des verbes à haute fréquence peuvent être comprises comme des phénomènes linguistiques *complexes*, en français écrit langue première (L1) et langue étrangère ou seconde (L2).

Les unités phraséologiques : un phénomène linguistique complexe à décrire et à acquérir?

Afin de répondre à cette question, nous avons tout d'abord abordé le phénomène phraséologique sous l'angle de la description en français L1, afin d'en dégager les caractéristiques définitoires et de fournir à nos analyses empiriques de données de corpus (en L1 et L2) un ancrage théorique solide, héritier de la littérature tant francophone qu'anglo-saxonne. C'est dans cette perspective de description du phénomène phraséologique en L1 que nous avons effectué une étude pilote, menée sur 200 séquences verbo-nominales construites avec le verbe *prendre* dans des corpus journalistiques et littéraires en français contemporain (400.000 mots).

Dans une perspective descriptive en français L2, nous avons ensuite procédé à une analyse contrastive interlangagière de 906 séquences verbo-nominales, dont 703 unités phraséologiques, construites avec les verbes à haute fréquence *prendre* et *donner* dans des textes de français académique produits par des apprenants avancés anglophones (185.000 mots), néerlandophones (90.000 mots) et par des étudiants francophones natifs (185.000 mots). L'objectif principal de cette analyse contrastive interlangagière est descriptif : il s'agit de déterminer, d'une part, les traits phraséologiques qui caractérisent l'interlangue avancée du français académique L2 et, d'autre part, les difficultés spécifiques aux apprenants avancés anglophones *vs* néerlandophones dans leur production d'unités phraséologiques. Nous avons ainsi étudié l'influence de variables externes (langue source, type de textes, medium oral/écrit) et internes (nature du verbe à haute fréquence, structure formelle du syntagme nominal, catégorie sémantique de la séquence) sur la production d'unités phraséologiques (déviantes ou non déviantes) par les apprenants avancés du français L2.

Si nous avons restreint notre objet d'étude aux unités phraséologiques verbo-nominales construites avec les deux verbes *prendre* et *donner*, c'est que ces verbes de base se caractérisent par : (i) leur potentiel phraséologique important et leur haute fréquence dans l'usage qui, en compensant la rareté du phénomène phraséologique sous sa forme très figée (par opposition au semi-figement omniprésent), nous ont permis d'effectuer des analyses sur corpus intégraux de données phraséologiques authentiques en contexte ; (ii) l'existence de formes équivalentes pour ces verbes dits *nucléaires* dans de nombreuses langues (Viberg 2002b), qui permet d'étendre l'étude de ces verbes à d'autres langues et/ou à d'autres verbes de ce type. D'un point de vue méthodologique, cette limitation de l'objet d'étude nous a par ailleurs permis : (i) de procéder à une analyse fine et exhaustive de données de corpus authentiques en L2 et en L1, en nous appuyant sur des bases de données lexicales et phraséologiques préexistantes disponibles pour le français L1 ; (ii) de développer des outils de travail reposant sur une approche multidimensionnelle du langage (syntaxe, sémantique et lexicale), incluant des critères phraséologiques qualitatifs (approche fonctionnelle de la phraséologie), mais aussi quantitatifs (approche statistique de la phraséologie) ; (iii) d'adopter une méthodologie mixte s'inspirant, entre autres, de la méthode d'analyse paramétrique et statistique de données de corpus (décrite dans Degand & Bestgen, 2004), de la sémantique contextuelle (Rastier *et al.*, 1994), de la linguistique de corpus (Habert *et al.*, 1997 ; McEnery & Wilson, 1996 ; Kennedy, 1998) et de l'analyse contrastive sur corpus d'apprenants informatisés (Granger, 2004), de l'analyse d'erreurs (James, 1998) et des analyses de précision et de complexité en acquisition d'une L2 (Ellis & Barkhuizen, 2005).

Comme nous le verrons, cette étude des unités phraséologiques s'inscrit dans une conception large du figement et porte une attention toute particulière aux séquences *semi-figées*, puisque celles-ci représentent la quasi-totalité des données que nous avons extraites de nos corpus (en L2 et L1). L'étude de ce type d'unités phraséologiques met en avant un phénomène crucial en acquisition d'une L2 : les séquences semi-figées (*vs* figées), qui sont les plus fréquentes dans l'usage, sont aussi les moins évidentes à appréhender, tant du point de vue de leur description linguistique que des phénomènes acquisitionnels qui les sous-tendent. Notre démarche, basée sur la méthode d'analyse paramétrique et statistique de données de corpus décrite dans Degand & Bestgen (2004), suit un schéma général qui présente l'avantage de pouvoir s'appliquer à de nombreux domaines linguistiques :

- 1° Préalable théorique: examen des littératures existantes sur le phénomène linguistique étudié
- 2° Constitution/choix du corpus d'étude et extraction semi-automatique des données
 - Extraction automatique *via* un concordancier (ici : *WordSmith Tools 4.0*)
 - Tri manuel avec désambiguïsation syntaxique et sémantique
- 3° Formulation des hypothèses de travail
- 4° Opérationnalisation des notions clés et analyse paramétrique :
 - Elaboration d'une liste paramétrique à partir des notions théoriques pertinentes pour : (i) déterminer le statut (non) phraséologique et (non) déviant des SVN étudiées ; (ii) procéder à une analyse combinatoire (sémantique et syntaxique) des deux verbes *prendre* et *donner*
 - Encodage des données chiffrées obtenues par la confrontation du corpus avec les critères paramétriques, sous forme de tableau (ici : *SPSS*)
 - Manipulation des données chiffrées et calculs statistiques *via* un programme statistique (ici : *SPSS*)
- 5° Interprétation des résultats : vérification/falsification des hypothèses de travail.

La structure de la présente étude suit *grosso modo* les grandes lignes de ce schéma. Les deux premiers chapitres seront ainsi consacrés à l'examen des littératures préexistantes en phraséologie (cf. Chapitre I) et en acquisition d'une L2 (cf. Chapitre II). Le Chapitre III, dont le propos est méthodologique, sera quant à lui consacré à la formulation des objectifs et des hypothèses de travail, à la méthode d'analyse que nous avons adoptée et à la description des corpus qui ont été explorés pour effectuer l'analyse contrastive interlangagière sur corpus d'apprenants en français L2. Les chapitres suivants présentent les résultats de cette analyse interlangagière : le Chapitre IV, portant sur l'analyse collocationnelle et de déviance des séquences étudiées, a pour objectif de déterminer le statut phraséologique ou non, ainsi que le caractère déviant ou non de ces séquences dans les corpus académiques L2 et L1 ; le Chapitre V consiste en l'étude de quelques variables linguistiques et extralinguistiques susceptibles d'avoir un impact sur la production d'unités phraséologiques par les locuteurs non natifs et natifs ; le Chapitre VI envisage les unités phraséologiques dans les textes uniquement argumentatifs, d'une part, sous l'angle d'une compétence phraséologique plus générale formulée en termes de *précision* et de *complexité* et, d'autre part, dans une perspective lexico-grammaticale centrée sur la mise au jour de *structures collocationnelles* particulières aux interlangues non natives. Enfin, dans le Chapitre VII de ce travail, une synthèse des résultats sera effectuée, à partir de laquelle nous formulerons quelques propositions explicatives et didactiques applicables au phénomène phraséologique étudié.

Chapitre I: Qu'est-ce qu'une unité phraséologique ?

Le champ d'étude de la phraséologie, au-delà de la complexité terminologique qu'il sous-tend, peut être envisagé de manière restreinte (approche fonctionnelle, cf. I.1.) ou élargie (approche statistique, cf. I.2.): ces deux conceptions de la phraséologie sont tributaires à la fois de la méthode d'investigation choisie par les chercheurs et de leur objet d'étude. Plutôt que d'envisager celles-ci de manière dichotomique, nous les aborderons de manière complémentaire dans notre analyse sur corpus d'apprenants (cf. Chapitres IV, V et VI). Notre propos dans le présent chapitre est de déterminer quels sont les critères définitoires des unités phraséologiques (cf. I.1.1.) et d'en examiner les principales catégories existantes (cf. I.1.3.). Nous étudierons plus particulièrement les dimensions scalaire et polyfactorielle du figement par le biais d'une étude pilote sur corpus des séquences à structure verbo-nominale construites avec le verbe *prendre* en français langue première (cf. I.1.4.). Nous nous attarderons en outre sur la notion de collocation (cf. I.1.3.2. et I.2.2.), puisque les séquences verbo-nominales qui font l'objet de notre analyse de corpus relèvent presque exclusivement de cette catégorie d'unités phraséologiques semi-figées.

- I.1. Phraséologie au sens strict (approche fonctionnelle)
 - I.1.1. Critères de figement
 - I.1.1.1. Mémorisation et institutionnalisation
 - I.1.1.2. Polylexicalité
 - I.1.1.3. Fixité : figement syntaxique
 - I.1.1.4. Non compositionnalité : figement sémantique
 - I.1.1.5. Restriction paradigmatique : figement lexical
 - I.1.2. Variations : variante, modification et défigement
 - I.1.3. Classification des unités phraséologiques (UP)
 - I.1.3.1. Phraséologie idiomatique et parémiologique
 - I.1.3.2. Phraséologie collocationnelle (définition fonctionnelle)
 - I.1.4. Degré de figement des unités phraséologiques : étude pilote en L1
 - I.1.4.1. Méthode et corpus
 - I.1.4.2. Postulats et hypothèses de travail
 - I.1.4.3. Analyse paramétrique : opérationnalisation des données et résultats
 - I.1.4.4. Traitement des données et interprétation des résultats
 - I.1.4.5. Conclusion
- I.2. Phraséologie au sens large (approche statistique)
 - I.2.1. Combinaisons syntagmatiques récurrentes (CSR)
 - I.2.2. Cooccurrences et collocations (définition statistique)
 - I.2.3. Séquences récurrentes
 - I.2.4. Prosodie sémantique
 - I.2.5. Colligations et collocations grammaticales
 - I.2.6. Patrons, cadres collocationnels et collostructions
- I.3. Conclusion intermédiaire

Collostructional analysis is an extension of collocational analysis specifically geared to investigating the interaction of lexemes and the grammatical structures associated with them. (Gries & Stefanowitsch, 2004: 209)

Les lexèmes qui sont sélectionnés par des structures particulières y sont appelés les *collexèmes* (*collexemes*) et, inversement, les structures qui sont sélectionnées par un lexème particulier sont appelées les *collostructs*.

Collostructional analysis always starts with a particular construction and investigates which lexemes are strongly attracted or repelled by a particular slot in the construction (i.e. occur more frequently or less frequently than expected). (Stefanowitsch & Gries, 2003: 214)

Dans le Tableau (9) (ci-dessous), nous présentons un exemple des résultats de l'analyse collostructionnelle menée par Stefanowitsch & Gries (2003: 219). Il s'agit en l'occurrence d'étudier quels sont les collexèmes les plus significatifs (test statistique de Fisher) en termes d'attraction collostructionnelle vis-à-vis des constructions [N *waiting to happen*].

Collexème (n)	p: Test exact de Fisher (force collostructionnelle)
accident (14)	2.12E-34
disaster (12)	1.36E-33
welkom (1)	4.46E-05

Tableau 9. Collexèmes montrant la plus forte attraction collostructionnelle par rapport à la construction [N *waiting to happen*] (Extrait de la Table 4 dans Stefanowitsch & Gries, 2003: 219)

Il ressort de ce tableau que l'attraction collostructionnelle la plus forte s'exerce entre la construction [N *waiting to happen*] et le collexème *accident*.

I.3. Conclusion intermédiaire

Sur la base des littératures préexistantes en phraséologie, nous avons examiné dans ce chapitre des notions théoriques centrales qui ont déterminé à leur tour les méthodes d'investigation des corpus. Ce travail dépasse le cadre de la phraséologie *au sens strict*, puisqu'il prend en compte le processus de figement lexical dans sa dimension polyfactorielle et scalaire, et prend pour acquise la fréquence de cooccurrence comme indice crucial de l'attraction mutuelle entre unités lexicales. Au-delà de la division notionnelle du champ phraséologique en phraséologie *fonctionnelle* et *statistique*, nous considérerons donc les deux acceptions de manière complémentaire. Cette complémentarité méthodologique se traduit notamment par la définition plurielle que nous avons donnée de la *collocation*. Nous prendrons ainsi le parti de définir la

collocation en tenant compte à la fois du critère de fréquence et des critères linguistiques de non-compositionnalité et de restriction lexicale. La collocation, en tant que semi-figement, est donc bien elle aussi un phénomène linguistique polyfactoriel et scalaire.

Comme c'est le cas de nombreuses études descriptives, les analyses inductives de données de corpus ne sont qu'une étape préliminaire pour des recherches plus approfondies, telles que l'étude de *profils collocatifs* (François & Manguin, 2006) ou *combinatoires* (Blumenthal, 2005, 2006) qui visent à déterminer la distribution cotextuelle des sens et des usages d'une forme lexicale. Dans cette thèse, nous nous attacherons ainsi à décrire et à comparer le profil combinatoire de deux verbes à haute fréquence, *prendre* et *donner*, dans leurs emplois natifs et non natifs en situation de production écrite. Nous étudierons en particulier les UP collocationnelles, qui représentent la quasi-totalité de nos données de corpus, en nous centrant sur la problématique de la production de telles unités par les apprenants du français langue étrangère ou seconde (L2).

Chapitre II : Quelle compétence phraséologique pour les apprenants du français L2?

Dans notre analyse sur corpus, nous étudierons la question de l'acquisition des unités phraséologiques par des apprenants universitaires, en situation guidée d'apprentissage du français langue étrangère ou seconde (L2). Il est donc essentiel d'introduire ici quelques notions issues de la recherche en didactique et en acquisition des langues secondes (cf. II.1.). Nous mettrons également en exergue deux axes de recherche exploités en acquisition d'une langue seconde, à savoir l'axe développemental (cf. II.1.4.1.) et l'axe variationnel (cf. II.1.4.2.). Nous aborderons ensuite la notion controversée de la norme en linguistique (cf. II.2.1.) mise en rapport avec la notion de déviance et d'erreur (cf. II.2.2. et II.2.4.). Enfin, après avoir examiné le triple processus acquisitionnel de complexité, précision et fluidité (cf. II.3.), nous décrirons la place qu'occupent spécifiquement les unités phraséologiques dans les études en acquisition d'une L2 (cf. II.4.), en mettant l'accent sur quelques recherches menées sur le français L2 (cf. II.4.2.2.) et sur les constructions verbo-nominales (cf. II.4.2.3.).

- II.1. Acquisition d'une L2
 - II.1.1. Appropriation, acquisition et apprentissage
 - II.1.2. Langue maternelle/ première (L1) vs langue étrangère/ seconde (L2)
 - II.1.3. Interlangue (IL)
 - II.1.4. Dynamisme et instabilité du processus acquisitionnel (L2)
 - II.1.4.1. Axe développemental
 - II.1.4.2. Axe variationnel
- II.2. Analyse de déviances
 - II.2.1. Norme vs usage
 - II.2.1.1. Norme subjective vs norme objective (L1)
 - II.2.1.2. Normes pédagogiques (NP) (L2)
 - II.2.1.3. Conclusion
 - II.2.2. Analyse d'erreurs
 - II.2.2.1. Repérage/ identification des erreurs
 - II.2.2.2. Description des erreurs
 - II.2.2.3. Explication des erreurs
 - II.2.2.4. Evaluation des erreurs
 - II.2.3. Analyse contrastive interlangagière
 - II.2.4. De l'erreur à la déviance : agrammaticalité, non-acceptabilité, inappropriété et étrangeté
- II.3. Précision, complexité et fluidité
 - II.3.1. Triple processus acquisitionnel en production d'une L2
 - II.3.2. Précision/ Exactitude
 - II.3.3. Complexité
 - II.3.4. Fluidité

- II.4. Acquisition des unités phraséologiques en L2
 - II.4.1. Acquisition du lexique en L2
 - II.4.2. Compétence phraséologique en L2
 - II.4.2.1. Itinéraire acquisitionnel des unités phraséologiques en L2
 - II.4.2.2. Etudes acquisitionnelles sur le français L2
 - II.4.2.3. Etudes acquisitionnelles sur l'anglais L2
- II.5. Conclusion intermédiaire

II.1. Acquisition d'une L2 : notions générales

II.1.1. Appropriation, acquisition et apprentissage

La recherche acquisitionnelle opère traditionnellement une distinction entre l'*acquisition*, l'*apprentissage* et l'*appropriation* d'une langue étrangère ou seconde (L2). L'*appropriation* est le terme générique qui désigne le processus d'intégration d'une langue, par assimilation et accommodation progressive des divers phénomènes langagiers qui la constituent. Issu de la psychologie piagétienne du développement de l'enfant, le terme d'*assimilation* représente « l'action qu'un sujet opère sur un objet pour absorber celui-ci » (Cuq, 2003: 27). Il correspond, dans les recherches en acquisition d'une L2, à l'intégration des données langagières par l'apprenant. De manière complémentaire, l'*accommodation* « désigne la capacité d'un sujet à s'adapter à une réalité qui lui est extérieure » (Cuq, 2003: 12) et correspond, en acquisition d'une L2, à la capacité qu'a l'apprenant d'adapter et de réactualiser des données langagières en fonction de la situation de communication verbale effective.

Le terme d'appropriation est employé comme hyperonyme par certains didacticiens qui souhaitent neutraliser la dichotomie acquisition/apprentissage. L'appropriation désigne l'ensemble des conduites [les stratégies d'apprentissage et les processus d'acquisition] de l'apprenant, des plus conscientes et volontaires (ce que d'autres appelleraient apprentissage) aux moins conscientes (ce que d'autres appelleraient acquisition). (Cuq, 2003: 25-26)

L'*acquisition* est généralement associée à la dimension innée et inconsciente de l'appropriation d'une langue. Ce processus est typiquement représenté par les situations d'immersion de l'apprenant dans un contexte natif, par *exposition* (en anglais *input*) de l'apprenant à des données langagières en situation naturelle d'appropriation. Rappelons que l'*exposition* se réfère à l'environnement langagier dans lequel l'apprenant évolue et qu'elle se distingue de la *saisie* (en anglais *intake*) qui se réfère à la matière langagière

underuse *and* (see Granger *et al.*, 1994). Qualitative research can then be carried out to analyse the reasons for this. (S. Granger, 1996: 45)

S. Granger souligne l'importance de mener des analyses contrastives fines à partir desquelles des hypothèses sur les mécanismes de transfert pourraient être testées : « CIA [Contrastive Interlanguage Analysis] results can only be reliably interpreted as being evidence of transfer if supported by clear CA [Contrastive Analysis] descriptions » (S. Granger, 1996: 46). L'auteure prend l'exemple du passif qui est deux fois plus fréquent en anglais L1 qu'en français L1, ce qui pourrait mener à formuler l'hypothèse d'un sous-emploi du passif en IL par les apprenants francophones de l'anglais L2. Mais inversement, pour pouvoir expliquer des cas de suremploi ou de sous-emploi en IL, il est nécessaire de recourir à des études contrastives fines qui fournissent des observations comparables dans la LS et la LC sur le comportement du phénomène linguistique étudié.

Au-delà de la notion d'*erreur*, les deux notions de *sous-emploi* et de *suremploi* deviennent centrales en analyse contrastive interlangagière. Le *sous-emploi* peut refléter une stratégie d'*évitement* de l'apprenant soit quand il veut contourner un obstacle qu'il considère comme tel par rapport à sa production langagière, soit quand il ne connaît pas la forme ou la structure (*ignorance*) à utiliser pour arriver à transmettre un contenu informationnel. Le *suremploi* peut quant à lui refléter notamment une surgénéralisation d'un phénomène linguistique déjà acquis ou un mécanisme de transfert par *divergence* (cf. *supra*). Comme nous l'avons déjà souligné à plusieurs reprises, le processus d'acquisition et la production d'erreurs dépendent d'une multitude de variables, tant linguistiques qu'extralinguistiques. De même, comme nous le verrons plus loin, les mécanismes de suremploi et de sous-emploi dépendent de facteurs divers qui sont susceptibles d'interagir simultanément.

II.2.4. De l'erreur à la déviance: agrammaticalité, non-acceptabilité, inappropriété et étrangeté

Traditionnellement, une forme est *erronée* quand elle ne correspond pas à une forme attestée dans la LC, elle est donc dite *agrammaticale* ou *non attestée* (cf. la notion d'*erreur absolue*, en II.2.2.1., p. 100). Nous considérerons dans notre étude les deux notions d'*erreurs* et de *déviances* comme étant équivalentes. La perspective sera néanmoins élargie et nous regrouperons sous ces deux termes des types d'erreurs dépassant le simple constat d'*agrammaticalité*. Dans une perspective plus étendue, une forme sera dite *erronée* ou *déviant*e dans une IL quand elle présente un certain degré de

non-acceptabilité (i.e. quand cette forme n'est pas attestée dans l'usage L1), d'*inappropriété* (cf. la notion d'*erreur relative*, en II.2.2.1., p. 100) ou d'*étrangeté* par rapport à l'usage natif d'une variété de la LC : (i) la déviance *formelle* concerne les formes de l'IL qui ne sont pas attestées dans l'usage de la LC (constat de *non-acceptabilité* marqué par un *), par ex. dans *l'accueil*³⁸ **donné aux enfants immigrés non francophones* (FL2-AN, <TAN231>, occ. 474), l'UP **donner Dét accueil* n'est pas attestée en L1, alors que l'UP *réserver Dét accueil* est attestée avec un haut degré de collocabilité (IM = 6,321) et est présente dans la liste des expressions verbales figées du LADL ; (ii) la déviance *contextuelle* concerne les formes de l'IL qui sont inappropriées en contexte (constat d'*inappropriété* marqué par un ?), par ex. dans *chaque homme peut ?prendre un rôle différent dans la société* (pour *jouer un rôle*) (FL2-AN, <TNL163>, occ. 130), l'UP *prendre un rôle*, dans son sens d' 'accepter un rôle proposé pour un spectacle', est utilisée en dehors de son contexte habituel ; (iii) la déviance *quantitative* concerne les formes de l'IL qui sont suremployées ou sous-employées par rapport à leur fréquence d'usage dans la LC ou par rapport au médium langagier usuel (constat d'*étrangeté* ou de *non-conventionalité* marqué par un !), par ex. dans *Les cours généraux !donnent une espèce d'éventail de tous les terrains d'intérêt possible* (FL2-NL, <TNL120>, occ. 1099), l'UP *donner un éventail de N* montre un certain degré de collocabilité (IM = 2,608) mais est peu fréquente dans l'usage natif (FqR³⁹ = 0,08), alors que l'UP *offrir un éventail de N* montre un degré de collocabilité nettement supérieur (IM = 6,08) qui traduit un usage préférentiel, plus conventionnel, de l'UP dans l'usage natif. Comme nous l'expliquons dans les pages qui suivent, nous avons déterminé l'acceptabilité d'une séquence principalement en fonction de son caractère attesté ou non dans l'usage en L1. Notre objectif est en effet d'étudier la compétence des non-natifs en la comparant à une compétence native *objective*, i.e. reflétant au maximum l'usage langagier en L1. C'est pourquoi, plutôt que d'utiliser le terme de *grammaticalité*, qui renvoie à la conformité vis-à-vis de règles de syntaxe abstraites, nous lui préférons celui d'*acceptabilité*, qui renvoie au caractère attesté d'une forme dans l'usage.

James (1998: 65-76) détermine quant à lui quatre types de mesures de l'erreur: (i) la *grammaticalité*, (ii) l'*acceptabilité*, (iii) le statut *correct* (*correctness*) et (iv)

³⁸ Rappelons que les exemples sont reproduits dans leur forme originale, tels qu'ils apparaissent dans nos corpus. Cela implique entre autres le maintien des erreurs orthographiques et des coquilles éventuelles.

³⁹ La notation *FqR* correspond à la *fréquence relative* de cooccurrence par million de mots, qui a été calculée pour le corpus de français L1 constitué d'une année du quotidien *Le Soir* et du corpus *Valibel*. Quand aucune indication supplémentaire n'est donnée, la *FqR* est par défaut celle relative au corpus du *Soir*.

l'*étrangeté* (*strangeness*). Nous laisserons de côté le concept de *correctness*, qui relève d'une étape que nous n'aborderons pas dans ce travail, à savoir l'étape de correction des erreurs qui ressortit au champ des applications pédagogiques en analyse d'erreurs. Nous retiendrons par contre de cette catégorisation les mesures en termes de *grammaticalité/acceptabilité*, qui concernent les déviations formelles et contextuelles, et les mesures en termes d'*étrangeté*, qui concernent les déviations quantitatives et stylistiques.

Tout d'abord, alors que la *grammaticalité* d'une forme est déterminée indépendamment de l'usage, par rapport à une norme grammaticale *subjective* (cf. *supra*), l'*acceptabilité* est définie par rapport à son caractère attesté ou non dans l'usage, en fonction d'une norme *objective* (cf. *supra*):

This is not a theoretical but a practical notion, being determined by the use or usability of the form in question (...). While it is the knower (NS) of a language who decides whether an utterance is grammatical, it is the user who decides what is acceptable. (James, 1998: 66)

ou encore

To decide on the acceptability of a piece of language we refer not to rules, but to contexts trying to contextualize the utterance in question. (James, 1998: 67)

Certains affirment que la *grammaticalité* et l'*acceptabilité* d'une forme sont indissociables. Cette affirmation a été remise en cause notamment par le constat qu'une forme agrammaticale à l'écrit pouvait être acceptable à l'oral. C'est ce que Blanche-Benveniste *et al.* (1990) ont constaté pour des formes du français relevant de ce qu'ils nomment la *grammaire seconde*, notamment pour la structure grammaticale des relatives avec *dont* qui est très peu productive à l'oral. Inversement, l'omission de la particule *ne* de la négation à l'oral ferait partie en quelque sorte d'une grammaire de l'oral, puisque très fréquente (voire généralisée) et donc acceptable en français oral, alors qu'elle correspond à un phénomène linguistique *agrammatical*. Dans notre analyse sur le caractère *attesté* ou non des séquences étudiées, nous utiliserons principalement la mesure d'*acceptabilité* des séquences par rapport à leur emploi dans l'usage des francophones natifs (*norme objective*), en investiguant des corpus intégraux de données authentiques (corpus *Le Soir* et corpus *Valibel*), ainsi qu'une base distributionnelle lexicale en *cooccurrents syntaxiques* élaborée à partir de dix années du corpus *Le Monde* (*Les Voisins De Le Monde*). Nous nous appuierons également sur une base de données contenant plus de 45.000 expressions verbales figées en français L1, élaborée par le LADL dans le cadre du lexique-grammaire de M. Gross.

Nous maintenons donc ici la distinction, bien que la *grammaticalité* et *l'acceptabilité* soient finalement, dans notre travail, traitées de manière équivalente puisque renvoyant au caractère *attesté* ou non des séquences par rapport aux ressources linguistiques dont nous disposons pour le français L1. Notre travail s'ancre en effet dans une approche de traitement de données de corpus intégraux et authentiques visant à étudier l'usage langagier en situation effective de communication langagière. La conception courante de la *grammaticalité*, qui distingue les énoncés bien formés ou *grammaticaux* (par ex. *Jean mange la pomme*) par opposition à un énoncé mal formé ou *agrammatical* (par ex. **Manger pomme le Jean*) sur la base du respect ou non de règles de syntaxe abstraites, ne sera donc pas (ou peu) prise en considération ici. Dans notre perspective d'analyse sur corpus d'apprenants, *l'acceptabilité* d'une séquence sera principalement déterminée par son caractère *attesté* ou non dans l'usage réel.

Nous définissons ensuite *l'inappropriété* en nous basant sur la remarque de Lennon concernant le statut des erreurs: « Most “erroneous forms” are, in fact, in themselves not erroneous at all, but become erroneous only in the context of the larger linguistic units in which they occur » (1991, cité dans James, 1998: 68). Autrement dit, alors qu'une forme *non acceptable* se caractérise par son caractère non attesté dans l'usage, une forme *inappropriée* se caractérise par son caractère *acceptable* dans l'usage, mais *inapproprié* par rapport au contexte dans lequel elle apparaît de manière effective.

Enfin, *l'étrangeté* dénote, ne fût-ce que par l'imprécision du terme employé, le flou qui entoure encore cette notion. Alors que *l'inappropriété* d'une forme se définit par rapport au contexte, une forme est dite *étrange* quand elle est appropriée au contexte, mais n'est pas le reflet d'un usage conventionnel (cf. la notion de *conventionalité* en I.2.2., p. 61) préféré par les natifs, soit en termes de fréquence d'emploi, soit en termes d'écart stylistique. D'une part, *l'étrangeté* se traduit par rapport à la fréquence d'usage dans la LC, notamment par l'emploi (voire un suremploi) dans l'IL de formes attestées en L1, mais qui ne sont pas les plus courantes ou les plus conventionnelles par rapport à l'usage des natifs (*déviances quantitatives*). D'autre part, *l'étrangeté* est aussi souvent associée à des *déviances stylistiques* liées au décalage entre la forme de l'IL et le type de texte ou le medium langagier (écrit vs oral) visés.

II.3. Précision, complexité et fluidité

Un champ de recherche assez récent en L2 se préoccupe d'étudier un triple phénomène observable lors de la production immédiate (*on-line*) de textes d'apprenants. Selon les analyses interlangagières basées sur ces trois concepts, il est admis que les apprenants peuvent, quand ils s'expriment dans une L2, se concentrer prioritairement sur la nécessité de s'exprimer (i) avec *précision/ exactitude (accuracy)*, (ii) de manière *complexe (complexity)* ou (iii) avec *fluidité (fluency)*. Il y aurait une tentative par l'apprenant de compenser le déséquilibre entre sa connaissance acquise et les besoins liés à l'impératif de communication. Premièrement, la *précision* correspond au « degré de coïncidence du système de l'IL avec le système langagier du locuteur natif (la norme) » (Van Daele *et al.*, 2007: diapositive 4). Deuxièmement, la *complexité* correspond à « la diversité interne et [au] degré d'élaboration (lexicale, morphologique, syntaxique) du système de l'IL » (*Ibidem*). Troisièmement, la *fluidité* correspond aux « parties automatisées du système de l'interlangue (facilité avec laquelle des unités de la connaissance sont accessibles, activées et sorties de la mémoire) » (*Ibidem*). Cette triple perspective repose sur le principe de l'indépendance théorique et empirique de chacune des notions, et trouve sa justification dans les applications didactiques qu'elle permet en s'appuyant sur des mesures précises et les plus reproductibles pour chacun des trois aspects.

II.3.1. Triple processus acquisitionnel en production d'une L2

If L2 forms have not yet been automatized, they require controlled processing, which places a heavy demand on learners' information-processing systems. The result is that they resort to the use of non-standard forms that have been acquired earlier and are automatized. (Ellis & Barkhuizen, 2005: 62)

Ellis & Barkhuizen (2005: 141) soulignent la difficulté pour l'apprenant de combiner en situation de production les deux attitudes *ségréative*, visant à donner priorité au contenu communicationnel, et *intégrative*, donnant priorité à la forme du message linguistique (cf. Pienemann *et al.*, 1988). Skehan (1998b, cité dans Ellis & Barkhuizen, 2005) opère quant à lui une distinction entre les systèmes linguistiques *exemplar-based* et *rule-based*. Le premier est constitué de l'ensemble des séquences (semi-) figées ou des combinaisons lexico-grammaticales d'un système linguistique. Les unités phraséologiques (UP) sont en effet intéressantes d'un point de vue du processus acquisitionnel en situation de production, puisqu'elles permettent d'optimiser la visée communicationnelle en minimisant l'effort lié à la nécessité de générer et de structurer des informations

linguistiques avant de les produire : « they contribute to increased production speed and fluency and they can also be employed to “buy time” while a speech act is planned » (Wray, 2000, citée dans Ellis & Barkhuizen, 2005: 142). Notons que le système *exemplar-based* concerne principalement les UP hautement figées, qui sont mémorisées comme un tout et appartiennent au stock lexical enregistré en mémoire à long terme. Outre la fonction optimisante des UP liée à la nécessité de transmettre de l'information en situation de communication, nous pensons pour notre part que la complexité du phénomène phraséologique peut aussi freiner la production langagière. Dans ce cas, la complexité peut être due soit à une non-correspondance entre les formes de la LS et celles de la LC donnant lieu à des transferts négatifs ou positifs, soit à une méconnaissance de la variabilité sémantique, lexicale ou syntaxique de certaines UP qui peut se traduire par la simplification, voire l'évitement des UP. En outre, la majorité des UP sont des SSF et se situent donc entre une combinatoire libre régie par des règles syntaxiques et sémantiques (*rule-based system*), et le figement à strictement parler (*exemplar-based system*). Selon ce point de vue intermédiaire, les UP collocationnelles partageraient certaines caractéristiques propres au système *ruled-based* défini par Skehan (1998, cité dans Ellis & Barkhuizen, 2005).

Language users also have access to a rule-based system. That is, they store knowledge of abstract rules that can be used to compute an infinite variety of well-formed utterances/sentences. The advantage of such a system is that it allows complex propositions to be expressed clearly, concisely and (as in the case of literature) in novel and creative ways. The disadvantage is that it is costly in processing effort, difficult to operate in online communication, especially where planning time is limited. (Ellis & Barkhuizen, 2005: 142)

Cette observation concernant la production des UP confirme l'idée selon laquelle les deux systèmes sont en fait complémentaires et sont activés successivement par les apprenants au cours d'une même situation de communication.

(...) they move between them naturally in accordance with the processing demands of the particular task they are performing and individual differences factors such as their language aptitude and learning style. (Ellis & Barkhuizen, 2005: 142)

Ceci nous amène à la nécessité de définir un peu plus avant les trois notions indissociables de *précision/ exactitude*, de *complexité* et de *fluidité*.

II.3.2. Précision/ exactitude

Le concept de précision est le plus souvent intuitivement compris comme l'absence d'erreurs d'ordre linguistique que fait un apprenant au moment où il tente d'utiliser la L2, tant en situation décontextualisée (dans le cadre d'un exercice scolaire artificiel) qu'en situation authentique de communication et, souvent, elle est aussi définie comme la connaissance adéquate que possède un individu des unités et des règles de la langue. (Germain & Netten, 2004: 1)

Les apprenants qui donnent priorité à la *précision* dans leurs productions langagières prennent généralement peu de risques et sont peu créatifs, puisqu'ils visent en priorité à utiliser de la manière la plus *exacte* qui soit les données langagières de la LC acquises antérieurement. Dans ce sens, la *précision* « refers to “how well the target language is produced in relation to the rule system of the target language” » (Skehan, 1996, cité dans Ellis & Barkhuizen, 2005: 139). Différents moyens permettent de calculer le taux de *précision* des productions d'apprenants, tant au niveau de leur compétence grammaticale que lexicale : (i) le pourcentage de propositions verbales non déviantes (*error-free clauses*), i.e. « the number of error-free clauses divided by the total number of independent clauses, sub-clausal units and subordinate clauses multiplied by 100 » (Foster & Skehan, 1996, cités dans Ellis & Barkhuizen, 2005: 150) ; (ii) le nombre de déviances pour 100 mots, i.e. « the number of errors divided by the total number of words produced [divided] (*sic*) multiplied by 100 » (Mehnert, 1998, cité dans Ellis & Barkhuizen, 2005: 150) ; (iii) l'analyse en termes d'occasions obligatoires (*obligatory occasions analysis*) visant à obtenir le pourcentage d'un phénomène linguistique spécifique employé correctement comparé au nombre total d'occasions que ce phénomène linguistique a de se produire: par exemple, pour obtenir le pourcentage d'emplois corrects de formes du pluriel par un apprenant, il faut calculer « the number of correctly used plurals divided by the number of obligatory occasions for plurals multiplied by 100 » (Crookes, 1989, cité dans Ellis & Barkhuizen, 2005: 150). Les deux premières mesures sont les plus répandues, mais leur application se heurte à un obstacle méthodologique de taille: comment définir précisément et de manière généralisable une *proposition verbale (clause)* et quels critères permettent de définir une *erreur* ? Alors que nous avons esquissé une réponse à la question de l'erreur (*vs* norme) dans les pages qui précèdent, la question de la délimitation de l'unité discursive, qui est étudiée plus en détail par Foster *et al.* (2000), devra faire l'objet d'une recherche ultérieure plus spécifique. La troisième mesure en *occasions obligatoires*, que les auteurs appellent *target-like uses* (Ellis & Barkhuizen, 2005: 150-151), est surtout pertinente pour l'examen

de phénomènes linguistiques particuliers, par exemple la morphologie verbale ou l'emploi de formes au pluriel. Pour étudier un trait langagier spécifique, l'analyse en *occasions obligatoires* compare ainsi le nombre observé d'occurrences de ce trait par rapport au nombre de fois où il aurait dû apparaître, autrement dit, en fonction des *contextes* ou des *occasions obligatoires* qui nécessitent virtuellement l'emploi de cette forme (par ex., le trait du subjonctif a entre autres contextes obligatoires les subordonnées introduites par *bien que*). Les recherches utilisant les mesures précitées se sont focalisées principalement sur des phénomènes syntaxiques, et rarement sur des phénomènes lexicaux. La mesure de *précision* spécifiquement lexicale la plus répandue consiste à diviser le nombre de déviations lexicales par le nombre total de mots contenus dans un texte (cf. Ellis & Barkhuizen, 2005: 152). En résumé, on peut répartir les différentes mesures de *précision* en deux groupes : les *mesures locales* qui se centrent sur des phénomènes linguistiques particuliers, et les *mesures globales*, qui visent à déterminer le niveau général de compétence. Nous aurons l'occasion de revenir sur certaines de ces mesures lors de l'analyse de déviance menée sur nos corpus d'apprenants (cf. Chapitre VI), en particulier sur la mesure élaborée par Foster & Wigglesworth (2007).

II.3.3. Complexité

La *complexité* – définie comme «the extent to which learners produce elaborated language» (Ellis & Barkhuizen, 2005: 139) –, peut être comprise de deux manières. Premièrement, elle correspond à l'usage par l'apprenant de données de la LC qu'il considère comme les plus difficiles à acquérir et qui lui demandent le plus d'effort à produire dans son interlangue, tenant compte du niveau d'acquisition auquel il se trouve. Deuxièmement, la *complexité* peut correspondre au niveau de risque et de créativité dont fait preuve un apprenant quand il s'exprime dans une L2, au-delà de l'utilisation des connaissances acquises antérieurement. Nous retiendrons quant à nous la définition globale donnée par Ellis & Barkhuizen (cf. *supra*) de la *complexité* comme désignant le caractère plus ou moins élaboré des textes de l'apprenant. Diverses études mesurent ce niveau de *complexité* en se centrant sur un plan langagier en particulier : syntaxe, interaction, lexique, etc. La majorité des études menées sur la *complexité* des productions d'apprenants d'une L2 sont syntaxiques par nature. Les chercheurs ont alors traditionnellement recours à l'analyse en termes de subordination, l'idée sous-jacente étant que la subordination reflète un niveau supérieur de complexité linguistique. Mais

certaines chercheurs ont aussi étudié la complexité syntaxique en dénombrant par exemple simplement les différentes formes verbales utilisées par l'apprenant. Sur le plan interactionnel, les chercheurs choisissent de compter le nombre de tours de parole par minute par interactant dans une même interaction verbale ou de calculer la moyenne du nombre de mots par interactant, voire la moyenne de la longueur des tours de parole par interactant. Sur le plan lexical, on parle de *richesse lexicale*. « It is also important to consider learners' use of vocabulary. This is because some learners may employ relatively simple grammatical structures but a wide range of different words. » (Ellis & Barkhuizen, 2005: 155). La mesure généralement utilisée est le *type-token ratio* (TTR), qui étudie la *diversité lexicale*, sachant que plus l'indice est proche de 1 plus la richesse lexicale est grande : le TTR est le résultat du nombre de lexèmes *types* divisé par le nombre d'occurrences lexicales (*tokens*) (cf. Meunier, 1998: 32). Cet indice est cependant controversé puisqu'il varie en fonction de la taille des textes étudiés: plus un texte est long, plus le *type-token ratio* aura tendance à être bas. On lui préfère donc parfois d'autres mesures, telles que l'indice de Guiraud (IG) qui n'est pas influencé par la taille des textes: « The *Indice de Guiraud* tries to compensate for this effect by dividing the number of types by the square root of the number of tokens » (Bulté, 2007: 35). Par ailleurs, Bulté (2007) propose pour analyser la *richesse lexicale* une mesure basée sur la qualité du lexique employé, plutôt que sur la seule quantité de formes lexicales utilisées. En effet, selon lui, « it is not difficult to imagine a text that is not very complex or rich in terms of vocabulary, but still would yield a high score on a TTR-based measure » (Bulté, 2007: 35), surtout à un niveau avancé d'appropriation. Cette dernière mesure renvoie au *profil de fréquence lexicale* (en anglais *lexical frequency profile*) de Laufer & Nation (1995) qui comparent les unités lexicales d'un texte à des listes de fréquence préexistantes en L1.

The reasoning behind this is that less frequent words are more difficult to acquire and consequently more complex. (...) If frequency has a qualitative dimension in terms of complexity (be it as a cause or as an effect), it is taken into account by this measure. Purely TTR-based measures ignore the 'quality' of words. (Bulté, 2007: 35)

Dépassant quant à nous le niveau des productions individuelles, nous pensons que la *complexité* serait aussi interprétable au niveau des groupes d'apprenants.

II.3.4. Fluidité

La *fluidité* concerne en particulier les productions orales d'apprenants : « (...) *fluency* is the production of language in real time without undue pausing or hesitation » (Ellis & Barkhuizen, 2005: 139). L'attention de l'apprenant se porte alors avant tout sur le sens de l'information à transmettre, celui-ci visant d'abord à atteindre son objectif communicationnel, quitte à employer des stratégies lui permettant d'éviter ou de contourner les obstacles linguistiques. Par définition, la *fluidité* concerne principalement les productions orales. Il serait néanmoins possible de développer des mesures qui soient applicables aux productions écrites, tenant compte par exemple du cadre temporel dans lequel est produit le texte (durée du temps de rédaction), de la situation génératrice ou non de stress (examen ou production libre), et du nombre de mots contenus dans le texte.

The accuracy, complexity, and fluency of learner language of both oral and written production can be measured. By and large the specific measures of accuracy and complexity that have been developed can be applied to both media. Obviously, however, fluency needs to be operationalized differently for oral and written language. (Ellis & Barkhuizen, 2005: 145)

Les mesures pour calculer le niveau de *fluidité* d'un texte sont entre autres le nombre de syllabes par seconde ou par minute (*speech/ writing rate*), le nombre de pauses ou la longueur de celles-ci par apprenant, ou encore la longueur moyenne des tours de paroles par apprenant.

Although the measures of accuracy, complexity, and fluency can be applied to both oral and written language, the bulk of the research that has employed these measures has examined oral language only. (Ellis & Barkhuizen, 2005: 164)

Dans notre travail, nous nous centrerons sur les notions de *précision* et de *complexité*, qui sont mesurables dans la perspective d'une analyse de corpus écrits d'apprenants, et laisserons donc de côté la notion de *fluidité*.

Après avoir abordé des notions qui relevaient du domaine général de l'acquisition en L2, nous adoptons dans les pages qui suivent une perspective plus spécifique axée sur l'acquisition des UP proprement dites.

II.4. Acquisition des unités phraséologiques en L2

Il est actuellement admis que les unités phraséologiques (UP) constituent des obstacles à l'acquisition d'une langue seconde ou étrangère (Mel'čuk, 1993 ; Colson, 2000 ; González Rey, 2002 ; Hunston & Francis, 2000 ; Schmitt & Carter, 2004), jusqu'à un niveau avancé d'apprentissage (S. Granger, 2004 ; Nesselhauf, 2004, 2005 ; Schmitt *et al.*, 2004).

The knowledge of and the ability to use prefabricated units are (...) essential for the language learner; unfortunately, however, they also pose considerable difficulties, even for the advanced learner. (Nesselhauf, 2005: 2)

Ce constat nous amène à examiner la place qu'occupe aujourd'hui la phraséologie en acquisition d'une L2, notamment par rapport à l'acquisition du lexique. Les UP étudiées dans ce travail se caractérisent en effet par leur fonction lexicale à visée référentielle et font donc partie, comme nous le verrons plus loin, de notre répertoire lexical.

II.4.1. Acquisition du lexique en L2

Depuis quelques années, l'importance a été reconnue de développer des méthodes et des outils didactiques pour l'apprentissage des unités phraséologiques (Howarth, 1998b ; Partington, 1998 ; Lewis, 2000 ; Willis, 2003 ; Nesselhauf, 2004), en prônant notamment l'introduction de « certaines collocations et expressions très courantes » dès le début de l'apprentissage du vocabulaire (Binon *et al.*, 2001 : 47). Howarth (1998b) souligne également la nécessité d'aborder en classe ces unités qui, même chez l'apprenant avancé universitaire, représentent des freins à la communication, et cela quel que soit le degré de gravité de la déviance. Selon l'auteur, les déviances phraséologiques peuvent avoir un impact plus ou moins grand sur l'*intelligibilité* des textes et engendrent, par leur effet *cumulatif*, un effet d'*imprécision*⁴⁰.

Many difficulties correspond to traditional fields of language study (including grammar and vocabulary), which teachers, lecturers and the writers themselves recognize and understand, and which are extensively treated in grammars and dictionaries. However, there is one type of error, resulting from a lack of phraseological competence, that has received relatively little close attention from linguists. These are errors that most learners are only dimly aware of, that many teachers can identify as a problem though cannot describe, and which linguists do not fully understand. Although they are not all major errors in themselves, and while the degree to which intelligibility is affected varies, they can have an appreciable impact on the effectiveness of a piece of writing, and their cumulative effect can be a serious loss of precision. (Howarth, 1998b: 161-162)

⁴⁰ Ce terme d'*imprécision* (*loss of precision*) est assimilable à l'effet d'*étrangeté* tel que nous l'avons défini plus haut (cf. II.2.4., p. 109) et renvoie aussi à la triple notion en acquisition de *précision/ complexité/ fluidité* (cf. II.3.).

Cette prise de conscience en didactique des langues, essentiellement répandue du côté anglo-saxon, va de pair avec la remise en cause de la distinction dichotomique héritée de la tradition linguistique opposant la grammaire et le lexique⁴¹. Les unités phraséologiques (UP), qui étaient naguère reléguées à la périphérie de la langue (González Rey, 2002) et considérées comme des irrégularités linguistiques, sont désormais au cœur du système langagier (Hausmann, 1997). Erman & Warren (2000) ont, comme d'autres chercheurs (Mel'čuk, 1993 ; Mel'čuk *et al.*, 1995 ; Howarth, 1998b), montré l'importance de la part de phraséologie dans les productions langagières par opposition aux séquences libres, en examinant la proportion d'UP dans des textes écrits et oraux en anglais L1 (par rapport au nombre de mots graphiques). Ces auteures ont ainsi constaté une proportion de 58% d'UP dans les corpus oraux et de 52% d'UP dans les corpus écrits. Dépassant la définition traditionnelle du lexique, il s'agit dès lors de prendre en compte, tant en didactique des langues qu'en linguistique théorique, les unités lexico-grammaticales issues de la phraséologie, considérées comme des unités complexes relevant à la fois de la syntaxe et du lexique.

Pattern grammar, lexical phrases and collocation have one thing in common: they all show us that words exist in company with other words and related concepts. We cannot learn to use a word unless we learn the patterns and phrases in which it occurs and the words with which it is associated. (Willis, 2003: 46)

C'est dans cette perspective que Wray (2002) intègre les unités polylexicales dans le répertoire lexical, en élaborant une conception nouvelle du lexique qu'elle appelle *hétéromorphique* (Wray, 2002). Le répertoire lexical d'un individu comprendrait ainsi trois modules, que nous mettons ici en rapport avec les notions développées plus haut concernant le domaine de la morphologie (cf. Tableau 2, p. 27): (i) les *morphèmes* (*morphemes*) concernent les traits morphologiques et morphosyntaxiques (désinences et affixes) (par ex. en anglais : *-ly*), ainsi que les *mots simples* (par ex. en anglais : *and, dog*) ; (ii) les *unités polymorphémiques* (*formulaic words*), qui sont mémorisées et activées de manière holistique, concernent les mots *dérivés* et *fléchis* (par ex. en anglais : *unhappy, kindness, established*), ainsi que les *mots complexes continus* (par ex. en anglais : *into, alright*) ; (iii) les *unités phraséologiques* (*formulaic words strings*) concernent les suites de mots mémorisées et activées de manière holistique, autrement dit les *mots complexes discontinus* (par ex. en anglais : *in order to, pull NP+GEN leg, Goodness gracious !*). L'idée de base de ce système hétéromorphique est que les différents modules

⁴¹ Nous rappelons que cette conception non dichotomique de la langue a été essentiellement développée en linguistique francophone par M. Gross (1975, 1981, 1988) dans le cadre du *lexique-grammaire* (cf. I.2, pp.57-58).

interviennent tous les trois en production langagière. Ces trois types d'unités lexicales se répartissent dans cinq types de lexiques que Wray distingue selon la fonction communicationnelle qu'ils remplissent: (i) lexique grammatical (Ex.: *in order to, on account of*) ; (ii) lexique référentiel (Ex.: *to take a chance*) ; (iii) lexique interactionnel (Ex.: *Great to see you*) ; (iv) lexique mémoriel (*memorized*) (par ex., les comptines enfantines) ; (v) lexique réflexif (*reflexive*) (Ex.: *Bloody hell !*). Notons que cette division va dans le sens d'une catégorisation des UP en unités *référentielles, organisationnelles* et *interactionnelles* (cf. Figure 1, p. 34).

D'après Wray (2002), les trois modules lexicaux seraient présents chez le locuteur natif (LN) et le locuteur non natif (LNN), mais dans des proportions différentes. Alors que le troisième module phraséologique est très important chez le LN, ce sont les deux premiers qui occuperaient une place privilégiée dans l'espace mental du LNN (Wray, 2002: 208). En situation de communication, le LNN manquerait ainsi, par rapport au LN, de moyens linguistiques lui permettant de s'exprimer de façon à la fois précise et fluide. Cette idée appuie notre choix d'aborder l'acquisition des UP en termes de *précision, complexité* et *fluidité* (cf. *supra*).

Après avoir souligné l'importance grandissante des UP en didactique des langues et abordé la question de leur place dans le répertoire lexical, nous examinons dans le point qui suit les UP du point de vue de leur rôle en acquisition d'une L2.

II.4.2. Compétence phraséologique en L2

Nous mettrons tout d'abord en évidence deux hypothèses qui ont été formulées sur l'acquisition des UP en L2, tout en appuyant la nécessité de nuancer ces perspectives. Ensuite, nous examinerons les résultats de quelques recherches menées en acquisition d'une L2 qui se révèlent pertinentes pour notre étude, en nous référant essentiellement à la synthèse théorique opérée par Forsberg (2006).

II.4.2.1. Itinéraire acquisitionnel des unités phraséologiques

Deux hypothèses principales ont été développées dans le domaine de l'acquisition d'une L2 dans une perspective phraséologique, autour de l'idée que les UP « play an essential role in language learning, as they seem to be the basis for the development of creative language in first language and childhood second language acquisition » (Nesselhauf, 2005 : 2).

La première hypothèse *décomposante* repose sur l'idée que les UP (*formulaic sequences*), qui sont acquises de manière *non analysée* aux stades initiaux de l'acquisition, feraient ensuite l'objet d'une décomposition analytique progressive. Comme le mentionne Forsberg lorsqu'elle rend compte des recherches effectuées sur la base de l'une ou l'autre hypothèse concurrente (soit *décomposante*, soit *holistique-analytique-holistique*), ces *unités non analysées* seraient, dans les phases initiales, « un outil pour entrer dans la communication et encourageraient ainsi l'apprenant à s'exprimer » (Forsberg, 2006 : 18). D'aucuns pensent que cette évolution d'une approche holistique à une approche analytique irait de pair avec l'émergence d'un système de règles créatives (Wong-Fillmore, 1976, citée dans Forsberg, 2006) : les LNN se serviraient des *unités non analysées* « comme des “cadres” permettant l'insertion de différents éléments » pour ensuite pouvoir construire « des énoncés librement, sans recours aux unités non analysées ni aux cadres » (Forsberg, 2006: 5). En outre, bien que l'évolution de l'acquisition des UP en L2 aille d'un mode holistique vers un mode analytique, les LNN partent généralement d'une perspective analytique dans leur mode d'acquisition du système langagier d'une L2 : « ce n'est que pendant la toute première période d'acquisition que l'apprenant L2 applique un mode holistique, mais il passe rapidement à un mode analytique » (Forsberg, 2006 : 7). Cette perspective ne rend cependant pas compte de l'usage d'unités phraséologiques (UP) en tant que traits plus tardifs des interlangues dans le parcours acquisitionnel.

Dans une visée spécifique étudiant l'influence des UP sur l'acquisition des marques du futur dans le système aspecto-temporel, Bardovi-Harlig (2002) montre quant à elle que les LNN emploient des UP pendant tout le processus acquisitionnel, bien que dans des proportions variables selon les stades, et pas uniquement aux stades initiaux. En outre, elle remarque qu'aux niveaux avancés, les UP sont beaucoup plus difficiles à distinguer qu'aux niveaux débutants. Pour sa part, Forsberg (2006) remarque la présence, aux stades avancés, d'un certain type d'UP dites *idiomatiques* qui, au contraire, sont relativement faciles à repérer et se distinguent des unités *non analysées* par les contraintes sémantiques et/ou formelles qui les caractérisent : selon l'auteure, il est ainsi « beaucoup plus facile de se prononcer sur des combinaisons préfabriquées ayant des contraintes sémantiques ou formelles [les UP *idiomatiques*] que sur celles qui sont parfaitement transparentes et qui ne sont détectables que par leur niveau grammatical plus avancé (...) ou par leur fonction [les UP *non analysées*] » (Forsberg, 2006 : 32). L'auteure souligne

dès lors l'importance de spécifier la nature des UP en cause (*non analysées* ou *idiomatiques*) en fonction des stades d'acquisition concernés.

C'est ainsi que la seconde hypothèse, qui prend en considération les niveaux avancés d'acquisition (Wray, 2002 ; N. Ellis, 2002, 2006 ; Forsberg, 2006), donne lieu à une vision de l'itinéraire acquisitionnel des UP qui n'est plus linéaire mais courbe (holistique-analytique-holistique). Les LNN, qui utilisent une grande quantité d'UP en tout début d'acquisition, passeraient ensuite par une phase intermédiaire plus analytique avant de recourir à nouveau à des emplois phraséologiques aux stades avancés de l'acquisition. Suivant cette hypothèse, les UP relatives aux stades initiaux et celles relatives aux stades avancés ne seraient pas totalement identiques. Il existerait deux types d'UP relevant de deux perspectives distinctes : (i) la perspective psycholinguistique en acquisition d'une L2, qui considère les UP comme des groupes de mots utilisés « en bloc » (*chunks*), mémorisés et activés de manière holistique et ayant une fonction première communicative (non nécessairement motivée ou référentielle) ; (ii) la perspective en description linguistique, qui aborde le phénomène phraséologique sur la base de critères internes (syntaxiques, sémantiques et lexicaux). Suivant cette dualité, Forsberg (2006 : 6) distingue les *unités non analysées* propres aux stades initiaux et les *séquences préfabriquées idiomatiques* repérables aux stades avancés, qui correspondraient à deux types d'apprentissage distincts s'inscrivant dans le processus acquisitionnel. L'exemple ci-dessous, donné par Forsberg (2006 : 8), illustre le cas du processus d'acquisition d'une *unité non analysée*, en tant qu'UP typique des premiers stades acquisitionnels:

[L]es unités non-analysées sont acquises très tôt, grâce à une attention explicite portée à une unité forme-sens (*form-meaning pairing*). Prenons l'exemple de *je m'appelle*. L'apprenant conçoit, bien qu'il ne comprenne pas la structure interne de l'unité, qu'il s'agisse [*sic*] d'un énoncé d'autoprésentation. Pour que cette séquence soit ancrée et consolidée dans le lexique mental, une certaine quantité d'input de la même séquence est nécessaire. On peut supposer, selon ce raisonnement, que moins la séquence est fréquente, plus de temps cela prendra-t-il [*sic*], avant que la séquence soit bien ancrée dans le lexique mental de l'apprenant. Certaines séquences demandent davantage de temps que d'autres pour être bien maîtrisées.

Selon N. Ellis (2002, 2006), c'est ainsi la quantité et la fréquence d'*input* qui seraient décisives dans l'acquisition des unités lexicales et des unités phraséologiques (UP) à un stade avancé d'apprentissage :

Just as we are sensitive to the combinatorial sequential probabilities of sublexical units, so, too, the recognition and production of words is a function of their frequency of occurrence in the language. (...) Kirsner (1994) has shown that there are strong effects of word frequency on the speed and accuracy of lexical recognition processes (speech

perception, reading, object naming, and sign perception) and lexical production processes (speaking, typing, writing, and signing), in children and adults as well as in L1 and L2. (N. Ellis, 2002: 152)

Au-delà de ces effets de fréquence, qui sont applicables tant aux unités lexicales simples qu'aux séquences de mots (*chunks*) (Ellis, 2002 ; Forsberg, 2005), nous examinons brièvement ci-dessous quelques études qui ont investigué l'acquisition des UP en L2 sous des angles divers.

II.4.2.2. Etudes acquisitionnelles sur le français parlé L2

Nous résumons ci-dessous, dans un premier point, deux études qui se sont focalisées sur l'acquisition du français parlé L2 du point de vue de leur rôle dans la structuration du discours et dans la fluidité verbale. Les UP y sont considérées comme des moyens stratégiques répondant aux besoins de la communication en L2. La troisième étude, la plus récente et de loin la plus exhaustive sur le français L2, porte sur l'élaboration d'un itinéraire acquisitionnel des UP en français L2, d'un niveau débutant jusqu'à un niveau très avancé d'apprentissage. Ces trois études allient les perspectives acquisitionnelle et phraséologique, en français L2, dans un contexte guidé d'apprentissage.

1) Rôle des unités phraséologiques dans la structuration et la fluidité verbale

[P]refabricated units are essential for fluency in both spoken and written language. Psycholinguistic evidence indicates that the human brain is much better equipped for memorizing than for processing, and that the availability of large numbers of prefabricated units reduces the processing effort and thus makes fluent language possible. (...) Enhancing fluency through reducing processing effort must be a particular interest for non-native speakers, as they naturally need more processing effort to convey their intended message. Indeed, it has been shown that whether or not L2 production is fluent crucially depends on the learner's control over a large repertoire of prefabricated units (...). (Nesselhauf, 2005: 2)

Raupach (1984, cité dans Forsberg, 2006) considère les UP (*formulae*) comme des outils aidant le L2N à structurer son discours. Les UP, qui font partie des phénomènes automatisés en langue, seraient ainsi repérables dans le flux langagier en fonction de critères intonatifs (hésitations, pauses, syllabes allongées) et du débit de parole, ainsi qu'en fonction des répétitions et des autocorrections (1984 : 117, cité dans Forsberg, 2006 : 32) : « d'après Raupach, les *formulae* se trouveraient entre les phénomènes d'hésitation ou en connexion avec eux » (Forsberg, 2006 : 32). Ces *formulae*, qui peuvent être monolexicales ou polylexicales, seraient de deux types selon leur fonction dans l'organisation du discours : (i) les *remplisseurs* (Ex. : *je ne sais pas, comment dire, oui,*

bon)⁴² « constituent des “points de repos” qui laissent à l’apprenant un peu de temps pour réfléchir » (Forsberg, 2006 : 32-33) ; (ii) les *organiseurs* (Ex. : *d’un côté... de l’autre, c’est pourquoi, c’est*) qui aident l’apprenant à structurer son discours.

Towell *et al.* (1996, cités dans Forsberg, 2006) étudient la fluidité verbale chez des LNN de niveau universitaire en français L2. Le public est donc proche du nôtre, si ce n’est la différence de médium langagier. « Towell *et al.* (1996) postulent qu’une des prémisses de l’obtention de la fluidité verbale des apprenants L2 est la procéduralisation des connaissances linguistiques » (Forsberg, 2006 : 33). Les auteurs visent ainsi à déterminer quels facteurs (e.a. les UP), sont déterminants ou non dans l’augmentation de la fluidité verbale de LNN ayant effectué un stage en pays francophone. Notons que les UP qui sont prises en considération dans cette étude relèvent de la catégorie des UP *interactionnelles* ayant une fonction d’ancrage dans une situation de communication particulière (*situationally bound lexical phrases*). Les résultats sont probants, notamment du point de vue de l’influence des UP sur la fluidité langagière des LNN :

Les auteurs découvrent que le développement de la fluidité verbale se traduit notamment par des énoncés plus longs et plus complexes et non pas par une augmentation du débit. Cette complexité syntaxique qui se manifeste après le stage est caractérisée par des constructions syntaxiques propres à la langue parlée française comme les clivées, les pseudo-clivées, les unités présentatrices comme *il y a un homme qui arrive* et, enfin, des phrases lexicalisées. (...) Towell *et al.* (1996) ont donc pu constater que les phrases lexicales situationnelles jouent un rôle important dans le développement de la fluidité verbale chez les apprenants de français L2. Ces résultats coïncident avec ceux de Raupach qui constate aussi l’importance du langage préfabriqué dans la production orale. (Forsberg, 2006 : 33)

2) Etude acquisitionnelle et comparative sur le français parlé L2

Les recherches de Forsberg (2005, 2006) visent à déterminer quelle fonction principale les UP, dénommées *séquences préfabriquées* (SP), remplissent en français L2 et L1 à l’oral⁴³, ainsi qu’à cerner quelles sont les différences significatives entre ces deux groupes, en termes de fréquence d’emploi (approche *quantitative*), de catégorisation des UP et du rapport types/occurrences des UP (approche *qualitative*). L’auteure distingue ainsi cinq catégories de séquences préfabriquées, dont les trois premières couvrent le champ général de la phraséologie alors que les deux dernières sont spécifiques aux UP repérables dans les IL des apprenants d’une L2 : (i) les séquences préfabriquées *lexicales*

⁴² Les exemples sont extraits de Forsberg (2006 : 32).

⁴³ Les données sur lesquelles travaille Forsberg sont extraites du corpus InterFra (Université de Stockholm, Suède) : 34 interviews de LNN pour l’étude allant des niveaux débutant à avancé ; 6 interviews de LNN pour l’étude menée sur les apprenants très avancés (Corpus de Paris).

concernent les mots *de contenu* et se caractérisent par leur contenu dénotatif et référentiel (par ex. *coup de foudre, faire la fête*)⁴⁴ ; (ii) les séquences préfabriquées *grammaticales* sont des mots *outils* à contenu procédural qui ont une fonction de déterminant, de quantificateur ou de marquage aspecto-temporel au sein du syntagme (par ex. *une sorte de, quelque chose, pas du tout, (être) en train de*) ; (iii) les séquences préfabriquées *discursives* désignent des mots *outils*, i.e. des *remplisseurs* (cf. la catégorisation de Raupach, ci-dessus), ayant une fonction de marqueur textuel (par ex. *Il y a, Par contre, C'est, Etant donné que*), de marqueur autorégulateur (par ex. *Je crois (que), Je veux dire, C'est-à-dire (que)*) ou de marqueur interactif (par ex. *Bien sûr, D'accord, Si vous voyez ce que je veux dire*) ; (iv) les *séquences préfabriquées idiosyncrasiques intérimaires* (SPIDI) sont des unités déviantes, à caractère généralement récurrent chez un même locuteur et propres à celui-ci, qui peuvent être soit déviantes contextuellement (i.e. *inappropriées*), soit déviantes formellement (i.e. *agrammaticales* ou *non attestées*) ; (v) les séquences préfabriquées *situationnelles*, considérées dans une perspective strictement acquisitionnelle et fonctionnelle, désignent des UP qui sont activées de manière holistique, qui sont liées à une fonction illocutoire spécifique (par ex. celle de 'se présenter') et qui peuvent être soit *autobiographiques* (par ex. *je m'appelle, j'habite*), soit *sociales* (par ex. *je voudrais, je ne comprends pas*). Notre étude portera pour sa part uniquement sur les séquences préfabriquées lexicales et les SPIDI qui lui correspondent. Concernant les SPIDI, Forsberg fait remarquer qu'elle n'a retenu pour son étude que les SPIDI *contextuelles* dont l'*inappropriété* était la plus évidente, laissant de côté les cas plus ambigus qu'elle qualifie de *subtilement non-natifs* (Forsberg, 2006 : 61). Dans notre étude focalisée sur les niveaux plus avancés, nous viserons pour notre part à rendre compte de ces écarts plus *subtils*, notamment à travers la notion d'*étrangeté* du point de vue des emplois préférentiels (*déviances quantitatives*).

Les résultats obtenus par Forsberg montrent que le LNN emploie des UP à tous les stades d'acquisition, avec une augmentation significative de la fréquence d'emploi des UP au fur et à mesure du continuum acquisitionnel (Forsberg, 2006 : 78). En termes de catégories de séquences préfabriquées utilisées, il y a une plus grande proportion de séquences préfabriquées discursives à tous les stades d'acquisition, et les séquences préfabriquées idiosyncrasiques et situationnelles sont caractéristiques des stades initiaux. De manière générale, on remarque que les séquences préfabriquées lexicales se

⁴⁴ Les exemples repris dans ce point sont tirés de Forsberg (2005 : 188-190).

développent très peu chez les LNN, ce qui conduit l'auteure « à penser que l'emploi plus important de séquences préfabriquées lexicales est un trait tardif et semble constituer la zone la plus faible pour les apprenants dans la mise en œuvre de SP [Séquences préfabriquées] » (Forsberg, 2006 : 79). Ce constat va dans le sens de l'étude menée par Howarth (1998b : 163) qui considère la catégorie d'UP comprenant les *collocations lexicales* (Ex. : *adopt a policy, aquiline nose*) comme étant la plus problématique en acquisition d'une L2 par opposition aux *collocations grammaticales* (Ex. : *argument about*). Chez les apprenants universitaires, de niveau intermédiaire à avancé, Forsberg constate que le répertoire des séquences préfabriquées lexicales se diversifie avec notamment une plus haute fréquence de constructions verbales prépositives, telles que *essayer de...* et *commencer à...* (Forsberg, 2006 : 83).⁴⁵ Chez les LN, elle constate une grande variété de types de séquences préfabriquées lexicales, avec la production d'emplois idiomatiques (par ex. *avoir envie, avoir l'air*). Nous ne détaillerons pas ici les résultats concernant les autres catégories de séquences préfabriquées et renvoyons le lecteur aux publications de Forsberg pour plus de détails à ce sujet.

II.4.2.3. Etudes acquisitionnelles sur l'anglais L2

Notons, à l'instar de Forsberg (2006), que la tradition acquisitionnelle en phraséologie est avant tout anglo-saxonne et que « jusqu'ici aucune étude générale sur le langage préfabriqué n'a été effectuée en linguistique acquisitionnelle française » (2006: 3). Nous nous devons donc de résumer ici quelques études issues de la tradition acquisitionniste anglo-saxonne, portant pour la plupart sur l'anglais L2, en mettant en évidence plusieurs observations essentielles qui interviendront lors de l'élaboration de nos hypothèses de travail.

1) Medium langagier et types de texte

Dans son étude sur l'anglais L2, Wiktorsson (2003, citée dans Forsberg, 2006) étudie les productions écrites d'apprenants lycéens et universitaires suédophones, en les comparant à des corpus natifs de l'anglais L1. Après avoir procédé à une étude de fréquence des UP

⁴⁵ La contrainte exercée par le verbe (*essayer, commencer*) sur la sélection de la particule complétive (*de, à*) confère toutefois à ces exemples un statut ambigu : en tant que mots de contenu, bien qu'à fonction référentielle faible, ils appartiennent encore à la catégorie des séquences préfabriquées lexicales, mais par leur fonction de marquage aspectuel (imperfectif pour *essayer* et inchoatif pour *commencer*) allant de pair avec leur cooccurrence avec une particule complétive, ils appartiennent déjà à la catégorie des séquences préfabriquées grammaticales.

dans ces deux corpus, l'auteure constate (de manière assez surprenante) que les LNN et les LN ont une compétence phraséologique similaire, au niveau du nombre d'UP produites et de la répartition en catégories fonctionnelles. La différence observée entre les locuteurs concerne le médium langagier, les apprenants utilisant davantage d'UP appartenant à l'oral que les natifs.

[L]es SP [Séquences Préfabriquées] utilisées par les apprenants appartiennent à un registre oral, tandis que les locuteurs natifs emploient des SP plus appropriées à la langue écrite. Les différences sont cependant plus marquées entre apprenants lycéens et locuteurs natifs : les apprenants utilisent les SP à un moindre degré et, de plus, l'écart registre oral/écrit est plus important. (Forsberg, 2006 : 35-36)

L'étude menée par Margareta Lewis (2005, citée dans Forsberg, 2006) porte sur des apprenants suédophones de l'anglais L2 (écrit) en milieu guidé et à un niveau intermédiaire d'acquisition. L'auteur montre des résultats similaires à ceux de Wiktorsson (2003), quoique divergeant en termes de fréquence et de catégorisation des emplois phraséologiques : (i) les LNN emploient moins d'UP que les LN (ce qui est moins surprenant, comparé aux résultats obtenus par Wiktorsson, cf. *supra*) ; (ii) la catégorie des UP la plus lacunaire est celle des UP *référentielles*, contrairement aux UP *organisationnelles* qui sont employées fréquemment (voire suremployées, cf. Forsberg, 2006) ; (iii) les LNN préfèrent les emplois verbaux, plus fréquents à l'oral, aux emplois nominaux, plus fréquents à l'écrit ; (iv) les LNN empruntent fréquemment des tournures issues du code oral ; (v) la fréquence d'UP augmente avec l'amélioration de la compétence langagière.

L'influence des types de textes sur la production d'UP a été mentionnée par Howarth (1998a, 1998b), qui a constaté chez le LN une proportion plus élevée d'emplois phraséologiques dans les textes journalistiques que dans les textes académiques. L'auteur suggère que cette différence serait due à la spécificité des types de textes.

In certain registers, the reader's attention can be deflected from message content to linguistic form by the conscious manipulation of standard collocations: newspaper headlines, feature articles, and advertisements are examples of discourse types where such changes are commonly made for deliberate effects. In academic discourse, however, the native writer feels under pressure to conform to an impersonal style and to keep the individual personality suppressed. Adherence to the conventional allows message content to be clearly perceived; linguistic divergence permits the personality to intrude. (1998b: 186).

De même Fitzpatrick & Wray (2006, citées dans Forsberg, 2006 : 114), qui ont examiné les facteurs extralinguistiques de compétence, d'aptitude, de passé linguistique, d'attitude, de motivation et de type de textes sur l'acquisition et la production d'UP, ont identifié les facteurs d'*intégration socioculturelle* et de *personnalité (risktaker)* comme étant les plus

décisifs dans l'acquisition des UP, à côté du facteur lié au *type de texte* influençant lui aussi la production d'UP.

2) Aptitude, attitude, motivation et intégration socioculturelle

Outre le facteur extralinguistique lié à la nature du médium utilisé et au type de texte concerné, des facteurs plus spécifiquement liés au contexte d'acquisition et à l'apprenant ont été étudiés d'un point de vue phraséologique en acquisition d'une L2. Wray (2002) a ainsi remarqué une différence entre l'acquisition des UP produites par les apprenants *non guidés* et celles produites par les apprenants *guidés*. Elle formule cette différence en termes d'*analyse au-cas-où* (en anglais *needs-only analysis*, *NOA* – cf. Wray, 2005).

The principle of NOA is simply that the individual does not break down input any further than is necessary to extract or create meaning. That is, there is no gratuitous analysis of form beyond the point where form-meaning mapping is sufficient for the present comprehension event, or for the construction of the presently required output. Over a period of time, an accumulation of event-specific comprehension and production requirements will lead to the identification of many small, recombinable items, plus rules for their combination. However, large units that never require such reduction will remain intact, and this will result in a mixed inventory of small and large items, as determined by the patterns in the input. In many cases a large unit will loosen up to permit limited morphological or lexical variation, creating a partially lexicalised frame (e.g. *at the end of the ___*; *NP1 pull + TENSE/AGR NP2+ POSSESSIVE leg*) (Wray, 2005).

Chez les apprenants *non guidés*, il y a une plus grande variation dans l'emploi des UP qui résulterait de la diversité des besoins individuels des apprenants en fonction du type de situation de communication dans laquelle ils se trouvent. Cette variation dépendrait de la motivation *intégrative* ou *instrumentale* dont fait preuve l'apprenant. Quant aux apprenants *guidés*, ils ne sont pas confrontés à des situations de communication authentiques et auraient donc à la fois moins besoin d'utiliser des UP (Wray, 2002) et moins d'*input* leur permettant d'automatiser celles-ci (cf. les effets de fréquences, N. Ellis, 2002 – *supra*). La *motivation* de l'apprenant serait donc liée aux besoins qu'engendrent les situations effectives de communication langagière.

Il ressort par ailleurs des études récentes menées par Schmitt *et al.* (2004) et Dörnyei *et al.* (2004) que « language learners show considerable variation in their acquisition of formulaic sequences [i.e. UP] » (Dörnyei *et al.*, 2004: 87). Ces chercheurs ont ainsi examiné quelles étaient les causes de la variation individuelle dans le processus d'acquisition des UP par des apprenants avancés de l'anglais L2. Dans un premier temps, Schmitt *et al.* (2004) ont étudié l'influence de trois variables individuelles (*aptitude*, *attitude* et *motivation*) sur l'acquisition des UP. Les résultats se sont cependant révélés

non significatifs: il semblerait qu'il n'y ait pas de corrélation entre ces trois facteurs extralinguistiques et l'acquisition d'une compétence phraséologique. Partant de ces premiers résultats non concluants, une étude a été menée ultérieurement par Dörnyei *et al.* (2004) consistant en l'examen des trois facteurs individuels d'*aptitude*, de *motivation* et d'*intégration socioculturelle*. Six personnes ont été interrogées à plusieurs reprises pendant quatre mois pour cette étude. Les auteurs ont ainsi mis au jour la forte corrélation existant entre l'acquisition d'une compétence phraséologique à un stade avancé d'apprentissage et le facteur d'*intégration socioculturelle* de l'apprenant dans une communauté native de la LC: « success in acquiring formulaic sequences is strongly related to the learners' active involvement in some English-speaking social community » (Dörnyei *et al.*, 2004: 104). Les auteurs ont en outre constaté que ce facteur jouait un rôle moins décisif dans le processus acquisitionnel des UP uniquement quand les deux autres facteurs gagnaient considérablement en importance. Allant dans ce sens, les résultats de Forsberg (2006) montrent que seuls les apprenants très avancés ayant séjourné un certain temps en France produisaient la même proportion de séquences préfabriquées lexicales que les natifs. Comme pour Fitzpatrick & Wray (2006, cf. supra), l'*intégration socioculturelle* paraît donc aussi jouer un rôle dans l'acquisition des UP, en particulier concernant les UP référentielles.

3) Constructions verbo-nominales

Nous développerons davantage ce point que les précédents, puisque nous y abordons les travaux qui ont focalisé leur attention sur les constructions à structure verbo-nominale (V + N) qui nous occupent. Nous partirons tout d'abord de la question soulevée par le statut des verbes à haute fréquence en acquisition d'une L2, pour ensuite explorer plus avant les travaux de deux auteurs étudiant en particulier les constructions verbo-nominales, l'un dans des textes académiques en sciences sociales produits par des apprenants matures de l'anglais L2 (Howarth, 1998a, 1998b) et l'autre dans des textes argumentatifs et descriptifs produits par des étudiants germanophones de l'anglais L2 à l'université (Nesselhauf, 2004, 2005). Nesselhauf (2004: 110) souligne ainsi l'importance de ce type de constructions, qui sont fréquentes en L1 et s'avèrent problématiques en L2, pour les apprenants de l'anglais L2.

a. Unités phraséologiques construites avec des verbes à haute fréquence

Howarth (1998b) a souligné le haut potentiel phraséologique des verbes à haute fréquence, en considérant le caractère arbitraire ou conventionnel des cooccurrences verbo-nominales comme une source d'erreurs pesant sur la sélection de l'élément verbal dans les collocations :

It is not surprising that such errors of selection occur, given the somewhat arbitrary way in which one delexical verb rather than another forms a conventional combination with a noun. (Consider **make an assurance* compared with *give an assurance*.) (Howarth, 1998b: 174)

Dans leur article sur le verbe à haute fréquence *make*, Altenberg & Granger (2001) mettent quant à eux en exergue la coexistence de résultats de recherche *a priori* contradictoires concernant l'usage des verbes à haute fréquence par des apprenants de l'anglais L2. Alors que les premiers résultats tendent à démontrer un suremploi des verbes à haute fréquence par l'apprenant (Hasselgren, 1994 ; Källkvist, 1999), les seconds mettent en évidence une tendance au sous-emploi de ces verbes à haute fréquence (Sinclair, 1991).

Premièrement, le suremploi de ces verbes serait lié à divers facteurs (Altenberg & Granger, 2001: 174) : leur apprentissage précoce dans le curriculum acquisitionnel, la richesse de leurs possibilités d'emploi, ainsi que l'influence du transfert positif entre des langues proches (Kellerman, 1979, cité dans Ellis & Barkhuizen, 2005 – cf. *supra*). Hasselgren (1994) propose de dénommer les unités lexicales suremployées par l'apprenant des *lexical teddy bears*, autrement dit, des *ours en peluche* : cette appellation, qui renvoie à l'objet transitionnel de la petite enfance appelé couramment en français le *doudou*, suppose que l'apprenant surgénéraliserait les emplois lexicaux qui le « rassurent » (*the words we feel safe with*) afin d'optimiser la communication⁴⁶ et de compenser les contraintes dues à la production langagière immédiate (*on-line*). Ces contraintes seraient notamment liées à des facteurs situationnels (e.a. le cadre temporel restreint dans lequel s'inscrit la production langagière) ou linguistiques (e.a. l'importance du répertoire lexical auquel l'apprenant a accès). En d'autres termes, le recours aux *doudous lexicaux* serait une stratégie compensatoire permettant à l'apprenant de s'exprimer avec plus de précision (et de fluidité à l'oral), en lui évitant au maximum de commettre des erreurs langagières.

⁴⁶ La visée serait donc ici plus instrumentale qu'intégrative.

Finding the right word in a second language is not easy. The choice may seem vast and time is rarely on our side. And the pressure not to produce a ‘bloomer’⁴⁷ intensifies as we advance. Little wonder then that, stripped of the confidence and ease we take for granted in our first language flow, we regularly clutch for the words we feel safe with: our ‘lexical teddy bears’. (Hasselgren, 1994: 237)

Les unités lexicales les plus susceptibles de devenir des *doudous lexicaux* sont généralement des termes usuels et hautement polysémiques qui, ayant fait l’objet d’un apprentissage précoce, sont suremployés de manière systématique à un stade plus avancé d’apprentissage (Hasselgren, 1994: 237-238). Les verbes se caractérisant par une haute fréquence dans l’usage sont donc de bons candidats pour devenir des *doudous lexicaux*.

[T]hese verbs have equivalents in many other languages, which may cause learners to overuse them in the case of formal or semantic similarity, as well as promote positive and negative transfer. (Bolly, à paraître).

Jaglinska (2006) applique quant à elle la notion de *lexical teddy bears* au domaine de l’acquisition des UP en L2 en instaurant le concept de *phrasal teddy bears* (que nous traduisons par *doudous polylexicaux*). Ses corpus d’étude sont constitués de textes écrits produits par des apprenants universitaires polonophones de l’anglais L2. Du point de vue longitudinal, l’auteure observe qu’il y a une augmentation de la fréquence d’emploi des UP par les apprenants, ce qui tend à les rapprocher d’une compétence phraséologique native. Mais elle note par ailleurs une corrélation régressive entre l’acquisition de cette compétence phraséologique et la diversité lexicale (*type/token ratio*) du répertoire des apprenants. Alors qu’ils emploient progressivement plus d’UP, leur répertoire lexical est de moins en moins riche. L’auteure suggère dès lors que les apprenants auraient recours à plus d’UP au stade le plus avancé, mais en suremployant certains types d’UP, des *doudous polylexicaux*, qu’ils préféreraient à d’autres.

Deuxièmement, le sous-emploi des verbes à haute fréquence serait dû quant à lui principalement à des stratégies d’évitement, ces verbes étant souvent constitutifs de constructions *délexicales*⁴⁸ (Howarth, 1998b), c’est-à-dire de séquences verbo-nominales construites avec des verbes dont le contenu sémantique est faible, comme c’est le cas généralement des constructions à verbe support (CVS). En d’autres termes, le sous-emploi des verbes à haute fréquence par l’apprenant serait notamment dû à une stratégie d’évitement mise en place pour contourner la difficulté des emplois phraséologiques de ces verbes, qui sont hautement polysémiques.

⁴⁷ En anglais familier, *to make a bloomer* signifie ‘faire une gaffe/ bêtise’.

⁴⁸ Nesselhauf (2004: 109) définit les unités *délexicales* comme des « expressions consist[ing] of an eventive noun that carries the bulk of the meaning, and a verb whose contribution to the lexical meaning is comparatively small, and which is therefore often called “delexical” ». Cette définition se rapproche de celle que nous avons donnée des *constructions à verbe support* (CVS) (cf. I.1.3.2., pp. 42-44).

[H]igh-frequency verbs are polysemous lexical items (Decock & Granger, 2004) which “display the whole gammut of phraseological constraints and have a very high rate of restricted and idiomatic uses” (Granger *et al.*, 2002: 13) (...).

Nous émettons l’hypothèse que la difficulté de l’acquisition d’une compétence phraséologique, en ce qui concerne les UP construites avec des verbes à haute fréquence, réside entre autres dans le fait que ces unités sont linguistiquement complexes : (i) elles relèvent à la fois de la syntaxe et du lexique (lexique-grammaire de M. Gross, 1975, 1981, 1988 ; Partington, 1998 ; Willis, 2003) ; (ii) la polysémie des verbes à haute fréquence (richesse sémantique) et la variabilité syntaxique des UP semi-figées (flexibilité) viendraient s’ajouter au caractère arbitraire de la co-sélection des constituants de l’UP.

Rappelons par ailleurs que les UP, quand elles sont abordées dans le curriculum acquisitionnel, le sont tardivement, contrairement aux verbes à haute fréquence qui font l’objet d’un apprentissage précoce.

b. Déviances collocationnelles dans des textes académiques d’apprenants adultes

Dans son étude sur la compétence phraséologique des apprenants de l’anglais L2, Howarth (1998b) a analysé un peu plus de 1.000 constructions verbo-nominales (ici dénommées *collocations*), comprenant des séquences libres (*free collocations*) (Ex. : *publish a book/an article/a prospectus*), semi-figées (*restricted collocations*) (Ex. : *put a question*) et figées (*idioms*) (Ex. : *put a premium on sth* ‘set a special value on sth’), produites par des LNN dans un cadre universitaire. L’auteur reconnaît en outre l’importance du rôle joué par la spécificité du verbe dans la détermination du caractère plus ou moins figé de la séquence. C’est pourquoi il distingue plusieurs niveaux de collocabilité en fonction du degré de restriction paradigmatique pesant sur les éléments verbaux (V) et nominaux (N) (Howarth, 1998b : 169-170):

- Niveau 1 : Liberté paradigmatique du N ; Restrictions paradigmatiques sur le V
Ex.: *adopt/ accept/ agree to a proposal/ suggestion/ recommandation/ convention/ plan, etc.*
- Niveau 2 : Restrictions paradigmatiques sur les deux éléments (V + N)
Ex.: *introduce/ table/ bring forward a bill/ an amendment*
- Niveau 3 : Restrictions paradigmatiques sur le V ; Blocage paradigmatique sur le N
Ex.: *pay/ take heed*
- Niveau 4 : Blocage paradigmatique sur le V ; Restrictions paradigmatiques sur le N
Ex.: *give the appearance/ impression*
- Niveau 5 : Blocage total sur les substitutions paradigmatiques (V + N)
Ex.: *curry favour*

Dans cette étude, le corpus non natif se compose de 10 textes (25.000 mots) produits par des apprenants très avancés (essentiellement par des professeurs d'anglais L2)⁴⁹ et le corpus natif d'environ 240.000 mots. L'objectif est d'identifier et d'analyser les déviations phraséologiques, en comparant les productions interlangagières des apprenants à des productions natives dans un même type de textes, en l'occurrence dans des textes *académiques* en sciences sociales (cf. la description de nos corpus d'apprenants en III.5., pp.161-163). Nous examinons successivement ci-dessous les caractéristiques mises en évidence pour les productions natives et non natives.

- Déviations natives

What deviation there is appears to be the result chiefly of unintended blends between two similar collocations. (Howarth, 1998b: 177)

Dans son corpus natif, Howarth (1998b) relève deux types principaux de déviations : des modifications grammaticales et des substitutions lexicales.

Premièrement, concernant les déviations grammaticales en L1, l'auteur note que leur rareté dénote une certaine prise de conscience par les LN des restrictions syntaxiques propres aux emplois collocationnels :

Overall, there is very little deviation arising from the structural modification of standard collocations; as one would expect, native speakers seem aware of the grammatical restrictions associated with collocations. (...) Most of the modifications are very minor omissions or additions, which cause little disruption to comprehension and may even pass unnoticed by the non-linguist. (Howarth, 1998b: 173)

Howarth (1998b: 173) prend l'exemple **prohibiting a person to **take an advantage** of a mistake* (pour *take advantage of*) pour illustrer le cas d'une déviation grammaticale due à l'insertion d'un article indéfini: la présence de l'article n'est attestée ni dans les ouvrages de référence ni dans l'usage L1, si ce n'est cette occurrence isolée dans le corpus natif de l'étude (1/5 occ.). L'auteur cite aussi des cas où la préposition utilisée n'est pas conforme à la forme collocationnelle standard (Howarth, 1998b : 173), par ex. **have consequences **on*** qui serait une erreur complexe (*blend*) amalgamant les deux emplois prépositionnels de *have consequences for* et *have effects on*.

⁴⁹ Contrairement à notre étude qui dénote une certaine homogénéité concernant le type d'apprenants (apprenants universitaires appartenant à une même communauté socio-linguistique et ayant une même L1), le contexte multilingue et multiculturel des apprenants étudiés par Howarth (1998b) – originaires de sept pays différents (Botswana, Allemagne, Grèce, Hong-Kong, Japon, Taiwan, Thaïlande) – a poussé l'auteur à formuler ses résultats non seulement en termes de groupe, mais aussi en termes de compétence phraséologique individuelle.

Deuxièmement, concernant les déviations lexicales en L1, les plus nombreuses dans son étude, elles concernent en majorité des déviations sur la sélection du verbe, comme dans **we will never reach a justice whose decisions are always conform to substantive truth* où l'auteur propose le verbe *achieve* comme emploi plus standard (1998b : 174). Notons que cette reconstruction native de la collocation déviante par l'auteur ne semble reposer, dans ce cas-ci, que sur l'intuition du chercheur. Nous verrons comment nous avons pour notre part élaboré, dans notre étude, des reconstructions natives à partir des indices de fréquence et de d'association lexicale *dans l'usage*. Howarth propose les notions de *chevauchements (overlaps)* et de *mélanges (blends)* collocationnels pour distinguer deux catégories de déviations lexicales portant sur la sélection erronée de l'élément verbal ou nominal. Les *chevauchements collocationnels* concernent les collocations verbo-nominales du Niveau 2 (cf. la classification en niveaux de collocabilité, p. 132) « in which a set of nouns have partially shared collocability » (Howarth, 1998b: 174). Dans ce type de collocations, plusieurs noms quasi synonymiques, i.e. proches sémantiquement ou appartenant à la même *classe d'objets*⁵⁰, peuvent co-apparaître avec l'élément verbal alors que d'autres quasi-synonymes ne le peuvent pas. Howarth (1998b : 174-175), à partir des données extraites de son corpus, observe ainsi que *draw up a memorandum* et *compile a list* sont acceptables, alors que **compile a memorandum* ne l'est pas. Il y a un chevauchement lexical entre deux constructions verbo-nominales du type [V + N_classe d'objets], en l'occurrence *draw up a memorandum/ list/ table* et *compile a *memorandum/ list/ table*, où « the error can be seen as the result of filling in a collocational gap within a partially overlapping cluster » (1998b: 175). Par contre, les *mélanges collocationnels* concernent des constructions verbo-nominales où il n'y a pas de chevauchement entre les possibilités lexicales collocationnelles. Bien que les noms soient proches sémantiquement, ils ne peuvent pas co-apparaître avec un verbe quasi synonymique. L'auteur donne l'exemple de *pay*

⁵⁰ Pour rappel, la notion de *classe d'objets* a été développée par G. Gross et ses collaborateurs et se définit en ces termes :

Dans leur emploi libre, donc, les prédicats doivent être définis à l'aide de classes et non d'unités lexicales. Certains prédicats ont un large spectre d'arguments, au point de pouvoir être décrits avec une précision suffisante à l'aide des traits syntaxiques (humain, animal, végétal, concret, locatif, temps, abstrait). (...) Mais il existe un très grand nombre de prédicats dont le sens est plus spécifique et pour lesquels l'indication du trait syntaxique est insuffisante. Les différents emplois de *prendre* que nous avons signalés plus haut ont tous comme compléments des substantifs concrets. (...) Il faut [aussi] spécifier à l'intérieur du trait "concret" les différentes classes d'objets qui sélectionnent le sens avec une précision suffisante pour le traitement automatique. (G. Gross, 1996 : 78-79)

*attention/ *pay care* par rapport à **take attention/ take care*.⁵¹ Comme nous le verrons plus loin dans notre analyse descriptive, l'auteur insiste sur le fait qu'il est parfois difficile de décider si la déviance porte sur la sélection du verbe ou sur le nom. Si l'on adhère à la définition fonctionnelle de la collocation (cf. I.1.3.2.) qui part du principe que l'orientation de la collocation se fait de la base (ici, le nom), vers le collocatif (ici, le verbe), il y aurait prédominance des erreurs de sélection du verbe par rapport aux déviances portant sur la sélection du nom. Ainsi, selon Howarth (1998b: 175), l'apprenant sélectionnerait d'abord le nom puis le verbe en fonction de celui-ci, de manière chronologique, ce qui engendrerait un plus grand nombre d'erreurs sur le verbe que sur le nom.

Du point de vue de l'impact de telles déviances sur l'intelligibilité des textes, Howarth souligne qu'il est moindre dans le cas des *mélanges collocationnels* (moins de 1% des collocations auraient un effet perturbateur sur la fluidité langagière), car ceux-ci ne sont pas saillants (*noticeable*) en situation de production langagière, du moins pour des LN.

The significance of blends for native speakers is that intelligibility is generally maintained. (...) That blends can be understood is evidence of shared phraseological competence, and it may be that many pass unnoticed as a result of the way in which semantically similar lexical complexes are stored nearby in the mental lexicon. (Howarth, 1998b: 176)

Se fondant sur ces données natives, Howarth propose ensuite une analyse de la compétence phraséologique des LNN.

- Déviances non natives

L'auteur distingue, comme chez les LN, les déviances collocationnelles qui portent sur les contraintes syntaxiques (*grammatical modification*) et celles qui portent sur les contraintes lexicales (*lexical substitution*) à l'intérieur même de la collocation.

Premièrement, les déviances grammaticales (*grammatical modifications*) ne représentent qu'un peu plus de 1% du nombre total des déviances étudiées par Howarth (1998b). Elles concernent par exemple les formes déviantes par omission de la préposition (Ex. : **respond students' need* pour *respond to students' need*) ou dues à une complémentation verbale erronée (Ex. : **achieve to gain* pour *achieve a gain*).⁵² Howarth

⁵¹ La frontière entre ces deux catégories est cependant fragile et dépend, comme le fait remarquer Howarth (1998b : 175), du jugement d'acceptabilité que le chercheur porte sur la collocation en contexte.

⁵² Alors que Howarth (1998b) choisit de ne pas tenir compte de ce type de déviances grammaticales dans son analyse, nous les prendrons pour notre part en considération dans notre analyse de déviance (comme

(1998b: 180) inclut dans cette catégorie les déviances portant sur un mauvais emploi des possibilités d'extension ou d'ajout d'un modifieur par rapport à une collocation donnée. Pour illustrer son propos, l'auteur prend deux exemples : (i) dans **draw a conclusive comment*, la collocabilité des termes *draw* et *conclusion* a été étendue erronément au modifieur *conclusive* ; (ii) dans **reach a high achievement*, nous retrouvons l'association *achieve + high* présente dans les séquences collocationnelles *a high achiever* et *a high level of achievement*.

Deuxièmement, concernant les déviances lexicales (*lexical substitutions*), elles représentent 6% des formes non natives analysées. Howarth (1998b) distingue les déviances qui ne semblent reposer sur aucune collocation lexicale préexistante (*nonce forms*), des collocations déviantes du type *mélanges* ou *chevauchements collocationnels*. La majorité des déviances observées ressortissent aux déviances dues à la spécificité sémantique des verbes et à leur potentiel combinatoire. L'auteur répertorie trois types de spécialisation sémantique des verbes en fonction du sens technique (Ex. : *acquire a language*), délexical (Ex. : **get contact with* pour *make contact with* ou *get in contact with*) ou figuratif⁵³ (Ex. : *pay care*) qu'ils véhiculent. Il constate tout d'abord que les verbes techniques sont les moins fréquents, mais sont aussi ceux qui posent le moins de problèmes. Ensuite, les emplois délexicaux, qui sont plus fréquents, sont néanmoins sous-employés par les LNN (13% des collocations) par rapport au corpus natif (21% des collocations).

This may be the result of the avoidance by some learners of such combinations due to uncertainty over appropriate collocability. This is not surprising given that even native speakers are capable of confusing delexical verbs (for example, **make an assurance* for *give an assurance*). (Howarth, 1998b: 181)

Enfin, la catégorie principale concerne les emplois figuratifs, qui ont été analysés dans le corpus natif en termes de *mélanges* et de *chevauchements collocationnels*. Le critère principal prôné par Howarth (1998b : 182) pour déterminer le degré de déviance de la forme non native par rapport à l'usage natif repose sur le degré de proximité sémantique entre les noms de la collocation déviante et celui de la collocation native qui était visée. Comme le souligne l'auteur, les sources de la déviance sont parfois difficiles à déterminer et il n'est pas toujours aisé de faire le lien avec la forme collocationnelle native visée par

Nesselhauf, 2004, 2005), considérant qu'elles font partie intégrante de la structure de la séquence étudiée, et influencent dès lors son statut phraséologique.

⁵³ La dimension figurative conférée aux collocations semble être ici très ténue et se rapprocherait plutôt du mécanisme global de l'*aréférenciation* (cf. la définition de cette notion dans notre analyse sémantique du verbe *prendre*, en I.1.4.3., p. 49), par un éloignement du sens premier concret de l'élément en question. Il ne s'agit donc pas dans ce cas à proprement parler d'un procédé topique par métaphore ou métonymie.

le LNN. Au-delà de cet obstacle méthodologique, on constate que les LNN ont moins recours aux *mélanges collocationnels* que les LN. Partant de ce dernier constat, l'auteur suggère que les emplois déviants qui relèvent de *mélanges collocationnels* seraient un trait propre à la variété native.

It is proposed here that blends are a normal part of native-speaker performance (rather than examples of aphasia, as Dechert and Lennon hint) and a sign of competence rather than incompetence. One can only produce blends of collocations that are known. If this view is accepted, blending in learners' production is not so mysterious, and, as in the native-speaker data, blends do not necessarily impair comprehension. (Howarth, 1998b: 185)

Du point de vue de la *normalité* sous-entendue par Howarth (cf. la citation ci-dessus), les *mélanges collocationnels* seraient *normaux* (et non plus *déviants*) puisque couramment utilisés par le LN. Nous considérons quant à nous ce trait comme restant une déviance du point de vue *normatif* (*vs normal*), tant chez le LN que le LNN, puisqu'il traduit un écart par rapport à une norme dite *objective*.⁵⁴ Notre perspective renvoie ainsi à la double définition de la norme native donnée précédemment (cf. II.2.1.1.) : à côté de la norme abstraite et prescriptive (*norme subjective*), les variationnistes mettent en évidence l'importance de la norme relative (*norme objective*) qui reflète la diversité des usages. Cette conception d'une norme *plurielle* et *objective* n'est donc en rien en contradiction avec les résultats obtenus par Howarth qui a mis au jour des déviations chez les LN : les deux processus de *mélanges collocationnels* chez les LN et d'*interférences* chez les LNN donneraient lieu à un même type de formes à la fois non conventionnelles (*non normatives*) mais représentatives de l'usage natif (*normales*).

Nous retenons en outre de cette étude l'importance accordée par l'auteur aux déviations phraséologiques, tant grammaticales que lexicales, portant sur la catégorie intermédiaire des UP représentée par les séquences semi-figées :

From the evidence of NNS collocation deviation, [...] it would seem that many learners fail to understand the existence of the central area of the phraseological spectrum between free combinations and idioms. It is in handling restricted collocations that errors of both lexical and grammatical nature constantly occur. (Howarth, 1998b: 186)

Dans le point suivant, nous mettrons en exergue les résultats obtenus par Nesselhauf (2004, 2005) qui portent sur le même type de constructions phraséologiques étudiées par Howarth (1998b), en l'occurrence les constructions verbo-nominales dites à *verbe support*.

⁵⁴ Notons que Howarth a lui-même comparé les *mélanges collocationnels*, entendus comme des déviations natives de nature intralinguale, aux déviations interlinguales des LNN procédant par transfert (1998b : 182).

c. Déviations collocationnelles dans des textes argumentatifs/ descriptifs d'apprenants universitaires

Les recherches de Nesselhauf (2004, 2005) sur les collocations, dénommées aussi *constructions à verbe support*⁵⁵ (CVS), sont menées sur des corpus argumentatifs et descriptifs d'apprenants germanophones avancés de l'anglais L2⁵⁶.

Collocations are (...) defined as combinations in which at least one element has a non-literal meaning (and at least one a literal one) and in which commutability is arbitrarily restricted, but some commutability is possible.

(...)

It seems (...) that commutability is generally seen as the more relevant of the two criteria; in addition, it is also easier to measure than opacity (cf. Hudson 1998:35). In my definition, I will therefore consider only commutability a defining criterion for collocations. (Nesselhauf, 2005 : 25, 27)

Nesselhauf vise principalement à : (i) identifier les traits et les déviations collocationnel(le)s propres à un groupe particulier d'apprenants avancés ; (ii) identifier les facteurs qui contribuent à la difficulté que présentent certaines collocations ; (iii) mettre au jour les procédés et stratégies mis en œuvre par les apprenants pour produire des collocations en L2 ; (iv) formuler des suggestions didactiques basées sur les résultats obtenus. Il s'agit donc de développer « a definition of collocations and a classification of collocation mistakes that may help material designers and teachers deal with the phenomenon more adequately and more systematically than has been the case hitherto » (Nesselhauf, 2005: 9). Dans son ouvrage sur les collocations dans des corpus d'apprenants (2005), l'auteure a ainsi extrait 2.000 occurrences de collocations verbo-nominales. Nous nous centrerons ici principalement sur les résultats portant sur les 251 occurrences des quatre verbes à haute fréquence *make*, *take*, *give* et *have*, qui ont fait l'objet d'un article par Nesselhauf (2004), puisque nous étudions dans ce travail deux correspondants de ces verbes en français, *prendre (take)* et *donner (give)*.

Les résultats de l'étude portant sur ces quatre verbes ont montré que 23% des collocations étaient déviantes, i.e. contenaient une ou plusieurs déviations. La plus grande proportion de collocations déviantes porte sur le verbe *make* (32% des collocations avec *make* sont déviantes), puis sur les deux verbes *give* (28% des collocations avec *give* sont déviantes) et *take* (26% des collocations avec *take* sont déviantes), et enfin, une proportion plus faible sur le verbe *have* (14% des collocations avec *have* sont déviantes).

⁵⁵ La conception des *constructions à verbe support* de Nesselhauf est une conception élargie de la définition que nous en avons donnée dans notre chapitre théorique. Elle correspond *grosso modo* à la définition fonctionnelle que nous avons donnée des collocations.

⁵⁶ Le corpus sur lequel Nesselhauf travaille est un sous-corpus élaboré à partir du corpus ICLE constitué par le Centre for English Corpus Linguistics à l'Université catholique de Louvain (S. Granger, 1998).

Les déviations répertoriées par l'auteure sont les suivantes : verbe erroné (Ex. : **take changes* pour *make changes*), nom erroné (Ex. : **make a cut* pour *make a distinction*), verbe et nom erronés (Ex. : **give action to* pour *put effort in*), CVS au lieu d'un verbe simple (**give solutions* pour *solve*), complémentation du nom erronée (Ex. : **take care for* pour *take care of*), déterminant erroné/ omis/ superflu (Ex. : **have sb under a strict control* pour *have sb under strict control*), structure syntaxique erronée (Ex. : **give a rest* pour *give sb a rest*), erreurs multiples, i.e. autres que portant sur le verbe ou le nom (Ex. : **make a shopping tour* pour *go shopping*). Du point de vue des catégories grammaticales sur lesquelles portent majoritairement les déviations, il s'agit principalement de mauvais emplois du verbe, puis d'emplois à verbe support au lieu de verbes simples, de complémentation du nom erronée, et enfin de mauvais emplois du nom (Nesselhauf, 2004 : 116). On observe en particulier pour le verbe *take* une majorité de déviations portant sur la complémentation du nom (6 occ.), puis sur le nom (2 occ.), et enfin sur des CVS au lieu d'un verbe simple (1 occ.) et sur le verbe (1 occ.). Comme le fait remarquer l'auteure, ces résultats sont néanmoins à relativiser étant donné le nombre restreint d'occurrences pris en compte par l'étude. Pour le verbe *give*, on observe des déviations dans toutes les catégories, avec une majorité de déviations sur le choix du verbe (4 occ.) et sur l'emploi de CVS (4 occ.).

En conclusion, la problématique majeure pour les apprenants de l'anglais L2 réside dans la sélection de l'élément verbal au sein des collocations. Contrairement à la plupart des études menées antérieurement, deux autres types d'erreurs sont toutefois pris en compte par Nesselhauf (2004 : 116).

Most of the existing studies have stressed the difficulties learners have in choosing the right verb. This is confirmed by the present data, and unsurprisingly so, as the verb in a support verb construction is by definition delexical and therefore to some degree arbitrary. (...) The present results, however, clearly show that there are other frequent types of mistakes besides the wrong choice of verb. (Nesselhauf, 2004: 116)

Ces deux types d'erreurs non formelles concernent des déviations *contextuelles* : d'une part, les déviations qui sont dues à une méconnaissance du sens plus ou moins spécifique des verbes (cf. Howarth, 1998b – *supra*) et, d'autre part, les déviations dont la forme est attestée dans l'usage en L1, mais qui ne sont pas appropriées en contexte. Comme nous aurons l'occasion de le voir plus loin, ces déviations obligent généralement le chercheur à avoir recours à un cotexte d'énonciation élargi pour pouvoir les repérer : elles sont en effet plus difficiles à discerner et sont donc plus *subtilement* (cf. Forsberg, 2006 : 61) déviantes que les erreurs formelles.

II.5. Conclusion intermédiaire

Nous tenterons de réunir dans notre étude les deux perspectives de recherche en acquisition des langues citées dans De Bot *et al.*, qui traduisent en quelque sorte la distinction entre les deux axes *développemental* et *variationnel* (cf. *supra*): (i) *the view from above*, qui propose d’aborder le langage de l’apprenant d’un point de vue global en mettant en évidence les similitudes et les régularités entre apprenants ; (ii) *the view from below*, qui se réfère à une approche plus locale de la L2, which « observe the messy details: behaviour that is variable, fluid, and highly context dependent » (De Bot *et al.*, 2005: 19). Bien que l’étude de l’impact des variables individuelles sur l’acquisition des UP soit indéniablement d’un grand intérêt tant pour les recherches en L2 qu’en phraséologie, nous n’adopterons pas dans ce travail une approche individualisante. Bien plus, notre objectif est de dépasser les limites intrinsèques aux méthodes d’élicitation en psycho- et sociolinguistique, qui portent généralement sur un nombre restreint d’apprenants. Nous déterminerons ainsi dans quelle mesure la variété avancée du français L2 se distingue de la variété francophone native, et en quoi des groupes d’apprenants de L1 différentes (anglais *vs* néerlandais) se distinguent ou se rapprochent l’un de l’autre quand ils ont le français L2 pour même objet d’étude. Il s’agira en quelque sorte de discerner en quoi consiste l’écart qui les éloigne de la LC, en termes de compétence phraséologique. Réunissant des notions issues des deux mondes de la recherche acquisitionnelle francophone et anglo-saxonne, les concepts de *déviance* et de *norme objective*, de *suremploi* et de *sous-emploi*, ainsi que les approches sur la *précision* et la *complexité*, seront centraux dans notre analyse de corpus interlangagière.

Dans le chapitre méthodologique qui suit, nous présenterons les objectifs visés par la présente étude et proposerons un récapitulatif des notions centrales abordées dans notre dernier point sur l’acquisition des UP en L2. Ce sont ces notions qui serviront de base à l’élaboration de nos hypothèses de travail. Nous procéderons ensuite à la description des bases de données et des corpus qui ont servi à notre étude.

Chapitre III: Analyse sur corpus d'apprenants – Méthode et corpus

La méthodologie adoptée dans ce travail, de type hypothético-déductif, nous a conduite à travailler en deux temps : élaboration d'hypothèses de travail (cf. III.2. de ce Chapitre III) et vérification/ falsification de ces hypothèses par confrontation à des données de corpus authentiques (cf. Chapitres IV, V et VI). Après avoir défini nos objectifs de recherche (cf. III.1.), nous abordons ici le premier temps de cette double démarche, i.e. la formulation des hypothèses de travail à partir d'une synthèse des notions développées dans les littératures préexistantes en acquisition d'une L2 (cf. III.2.). Nous décrirons ensuite brièvement la méthode d'analyse adoptée (cf. III.3.), ainsi que les corpus et les bases de données auxquels nous avons eu recours dans ce travail (cf. III.4. et III.5.). Enfin, nous expliquerons comment nous avons procédé à l'extraction de nos données d'étude à partir de ces ressources pour le français L1 et L2 (cf. III.6.).

III.1. Objectifs de recherche

III.2. Hypothèses de travail

III.2.1. Compétence phraséologique

III.2.2. Déviations phraséologiques

III.3. Méthode d'analyse paramétrique et statistique de données de corpus

III.3.1. Opérationnalisation des données

III.3.2. Manipulation et traitement statistique des données

III.4. Bases de données et corpus pour le français L1

III.4.1. Liste des expressions verbales figées du LADL

III.4.2. Base lexicale distributionnelle *Les Voisins De Le Monde*

III.4.3. Corpus de presse *Le Soir* (1997) et corpus oral *Valibel*

III.5. Corpus d'apprenants en français L2

III.6. Données de corpus

III.1. Objectifs de recherche

Le versant phraséologique de la compétence linguistique est encore peu développé en didactique des langues (du moins pour le français), ce qui se traduit par des lacunes ou niveau des moyens pédagogiques mis en œuvre dans ce domaine. Notre approche constitue dans ce contexte une étape *descriptive* préliminaire à d'éventuelles études à visée explicative ou didactique: elle tend à mettre à la disposition du linguiste, de l'apprenant et de l'enseignant, des outils permettant de prendre conscience, d'identifier, et/ou de mieux comprendre le phénomène phraséologique. A l'instar de Howarth (1998b) qui reconnaît la place centrale qu'occupent les unités phraséologiques (UP) dans l'activité langagière, nous pensons donc qu'il est nécessaire de décrire préalablement le phénomène phraséologique pour permettre ensuite à l'apprenant d'en user à bon escient.

Moreover, learners need to understand that restricted collocations make up a significant part of a typical native speaker's production in both speech and writing. Conforming to the native stylistic norms for a particular register entails not only making appropriate grammatical and lexical choices but also selecting conventional collocations to an appropriate extent. Such combinations are not optional stylistic adornments on the surface of text; they are essential for effective communication, and their use by non-native writers is a clear sign that these learners have made an essential adjustment to the academic culture they are entering. (Howarth 1998b: 186)

La description des unités phraséologiques (UP) en L2 repose en outre sur leur description préalable en L1. C'est pourquoi nous avons travaillé en deux temps.

Dans un premier temps, nous avons étudié le phénomène phraséologique en français L1 en élaborant une mesure fine du degré de figement des séquences verbo-nominales (SVN) construites avec le verbe *prendre* (cf. notre étude pilote sur le degré de figement, en I.1.4., pp. 45-56). Cette étude scalaire et polyfactorielle du phénomène collocationnel nous a permis de mettre en évidence la variabilité (syntaxique, sémantique et lexicale) des occurrences des UP étudiées (*tokens*) pour une même séquence type (*type*). À partir des résultats obtenus, nous avons alors formulé, dans la conclusion de cette étude pilote (cf. I.1.4.5., p. 56), la question suivante : la méconnaissance du comportement phraséologique des séquences étudiées (presque exclusivement des semi-figées) et de leur degré de variabilité irait-elle de pair avec une tendance des apprenants au mésusage, au sous-emploi ou au suremploi de celles-ci ?

Dans un second temps, l'étude collocationnelle et l'analyse de déviance des UP, produites par des natifs et des non-natifs, visent à mieux cerner quels sont les emplois phraséologiques (non) déviants les plus fréquents et les plus spécifiques aux apprenants universitaires avancés du français L2, notamment en fonction de leur L1 (cf. Chapitres

IV, V et VI de ce travail). Notons que le cadre de cette étude étant descriptif, nous n'avons pas exploré de manière approfondie les causes potentielles de cet écart. Par contre, nous nous intéresserons à ce qui en résulte, à savoir les UP (tant non déviantes que déviantes) repérables dans les corpus d'apprenants. Nous avons conscience de n'amorcer ici qu'une étape s'insérant dans une étude plus large des productions d'apprenants : aux trois questions de Nesselhauf (2005: 8, *infra*), exprimant la nécessité de déterminer (i) quelles collocations ou quels types de collocations sont les plus problématiques pour certains groupes d'apprenants, (ii) quels types de déviances sont produites et (iii) pourquoi ces dernières sont produites, nous ne répondrons qu'aux deux premières, laissant aux spécialistes en acquisition le soin d'investiguer le *pourquoi* de ce phénomène en L2.

[T]he questions that seem most important for the design of pedagogical material, i.e. which collocations or types of collocations are most difficult for certain groups of learners, what kinds of mistakes occur and why, have received little attention so far. (Nesselhauf, 2005: 8)

Les notions héritées de la recherche en acquisition des langues secondes, qui serviront à l'élaboration de nos hypothèses de travail, ont donc été retenues en fonction de l'intérêt qu'elles présentaient pour la perspective proprement descriptive que nous avons adoptée.

Deux questions ont guidé notre recherche : (i) la compétence phraséologique est-elle à considérer comme une zone de *fragilité* en acquisition du français L2 (Bartning, 1997: 22-23), au même titre que les compétences syntaxique, lexicale, morphologique, rhétorique ou discursive (Bartning, 1997 ; Bartning & Kirchmeyer, 2003 ; Bartning & Schlyter, 2004) ? (ii) et si tel est le cas, quelles seraient les erreurs spécifiques à ce domaine de compétence langagière dans le cas particulier d'apprenants universitaires, néerlandophones et anglophones, du français écrit L2 ? Nous avons déjà partiellement répondu à la première question par l'affirmative, en nous référant aux études antérieures menées sur l'acquisition des UP en L2 (cf. II.4.). Tentant de répondre à la seconde question, notre objectif principal de recherche sera dès lors de déterminer quelles sont les spécificités d'une telle compétence à un stade avancé d'acquisition, via l'étude du comportement phraséologique des deux verbes à haute fréquence (VHF) *prendre* et *donner* en français écrit L2.

III.2. Hypothèses de travail

Pour pouvoir formuler nos hypothèses, nous nous appuyons sur les littératures antérieures en phraséologie, en particulier en phraséologie collocationnelle (cf. I.1.3.2. et I.2.2. de ce travail) et en acquisition des langues secondes (cf. Chapitre II). Nous proposons dans ce point une synthèse des notions que nous avons retenues comme étant pertinentes pour notre analyse sur corpus d'apprenants du français L2 à l'écrit, concernant les UP construites avec les deux verbes à haute fréquence *prendre* et *donner*. Nous élaborerons nos hypothèses de travail en regard de ces notions.

III.2.1. Compétence phraséologique

Nous formulons ci-dessous nos premières hypothèses concernant la compétence phraséologique d'un point de vue plus général et notamment du point de vue de la fréquence d'emploi des UP produites par des apprenants avancés d'une L2.

- Les UP représentent un obstacle à l'acquisition d'une L2 jusqu'à un niveau avancé d'apprentissage (Bardovi-Harlig, 2002 ; S. Granger, 2004 ; Schmitt *et al.*, 2004 ; Nesselhauf, 2004, 2005 ; Margareta Lewis, 2005). On observe ainsi une fréquence des UP chez les locuteurs non natifs (LNN) moins importante que chez les locuteurs natifs (LN) (Margareta Lewis, 2005 ; Forsberg, 2006). De même, Wray (2002) constate que le module phraséologique du lexique *hétéromorphique* (cf. II.4.1., pp. 119-120) est moins important dans le répertoire lexical des LNN que dans celui des LN.
- Deux types d'UP coexisteraient aux stades avancés d'acquisition d'une L2 : les UP dites *non analysées* (plus transparentes), qui sont plus difficiles à repérer dans les productions des LNN (Bardovi-Harlig, 2002) ; les UP dites *idiomatiques*, qui sont plus facilement repérables dans les productions des LNN par le biais des contraintes sémantiques (non-compositionnalité) et syntaxiques (fixité) qui pèsent sur celles-ci (Forsberg, 2006).

Hypothèse 1

Le phénomène phraséologique engendrerait des difficultés pour les apprenants d'une L2 jusqu'à un niveau avancé d'acquisition. Ces difficultés seraient liées entre autres à la complexité linguistique qui caractérise les UP semi-figées, puisque celles-ci se rapprochent de la combinatoire libre tout en possédant déjà des caractéristiques phraséologiques (contraintes syntaxiques, sémantiques et/ou lexicales). L'écart entre la compétence phraséologique interlangagière des apprenants avancés et la compétence phraséologique native serait ainsi mesurable en termes de complexité structurelle et lexicale.

- Les UP construites avec des verbes à haute fréquence (Altenberg & Granger, 2001) peuvent faire l'objet d'un suremploi (Hasselgren, 1994 ; Jaglinska, 2006) mais aussi d'un sous-emploi par les apprenants avancés. Les constructions délexicales verbo-nominales seraient ainsi sous-employées par les apprenants avancés (Howarth, 1998b). Le suremploi des UP en tant que *doudous polylexicaux* (cf. les *phrasal teddy bears* de Jaglinska, 2006) irait de pair avec une richesse lexicale moindre par rapport aux stades antérieurs d'acquisition.

Hypothèse 2

La compétence phraséologique des apprenants d'une L2, concernant les UP construites avec des verbes à haute fréquence, se caractériserait paradoxalement par un écart en termes de (a) suremploi et/ou de (b) sous-emploi par rapport à l'usage natif.

- a) Les apprenants avancés non natifs auraient tendance à recourir à des *doudous polylexicaux*, ce qui se traduirait à la fois par un suremploi de certaines UP *types* et par une diminution de la richesse lexicale des productions phraséologiques des apprenants en L2.
- b) Les constructions délexicales verbo-nominales seraient sous-employées par les apprenants avancés.

- Il y aurait une influence du type de texte (Howarth, 1998a et 1998b) sur la fréquence et le type d'UP produites, par exemple en fonction du degré d'implication subjective du locuteur que requiert ou non le type de texte en question.
- A l'écrit, on observe que le LNN emploie fréquemment des UP typiques du code oral (Wiktorsson, 2003 ; Margareta Lewis, 2005). Les LNN utilisent également plus de constructions verbales que de constructions nominales à l'écrit, alors que les premières sont un trait propre au code oral et les secondes au code écrit.

Hypothèse 3

Il y aurait une influence (a) du medium langagier et (b) du type de textes sur la fréquence et le type d'UP utilisées par l'apprenant avancé d'une L2.

- a) On observerait un emploi fréquent, par les apprenants, d'UP provenant du code oral, ce qui donnerait lieu à un écart stylistique par rapport à l'usage conventionnel en L1.
- b) Les types de textes nécessitant un style plus impersonnel (par ex. les textes réflexifs) comprendraient moins d'UP que les textes plus subjectifs (par ex. les textes argumentatifs).

III.2.2. Déviances phraséologiques

Après avoir formulé des hypothèses d'ordre plus général, nous abordons ci-dessous les notions qui concernent plus directement les déviances phraséologiques, du point de vue de leur fréquence et de leur catégorisation.

- Chez les apprenants d'une L2, on observe une augmentation du nombre d'UP employées correctement au fur et à mesure du cursus acquisitionnel (Wray, 2002 ; Forsberg, 2006).

Hypothèse 4

La compétence phraséologique des apprenants avancés d'une L2 se caractériserait par un écart en terme de mésusage par rapport à l'usage natif, avec une proportion plus importante d'UP déviantes que chez les locuteurs natifs. Cet écart serait mesurable via un calcul de précision : les productions non natives se caractériseraient par un degré de précision moins élevé que les productions natives.

- Des déviances phraséologiques, bien que peu nombreuses, sont aussi observées dans les productions natives (Howarth, 1998b). Howarth observe ainsi quelques rares déviances grammaticales (par ex. l'omission/ ajout d'un article ou la sélection d'une préposition erronée) et une majorité de déviances lexicales portant sur la sélection du verbe.

Hypothèse 5

Comme pour l'anglais L1, des déviances portant sur les UP verbo-nominales seraient repérables dans les textes académiques en français L1. La plupart des erreurs seraient lexicales et porteraient sur l'élément verbal. Ces déviances n'altéreraient pas ou peu l'intelligibilité des textes. Leur degré de gravité serait donc moindre.

- A côté des déviances formelles, certaines déviances plus *subtilement non natives* (Forsberg, 2006) ont été mises en évidence chez les apprenants avancés d'une L2 (Howarth, 1998b ; Nesselhauf, 2004, 2005) : les déviances dues à une méconnaissance des emplois spécifiques du verbe (dans le cas des UP verbo-nominales) et les déviances contextuelles.
- Quel que soit le degré de gravité des déviances et leur impact sur l'*intelligibilité* du texte, les déviances à un stade avancé d'acquisition provoquent au minimum un effet d'*imprécision (loss of precision)* par leur effet cumulatif (Howarth, 1998b : 161-162)

Hypothèse 6

Aux stades avancés d'apprentissage en L2, on observerait, à côté des déviances formelles, une part importante de déviances plus *subtiles* dues à un usage inapproprié en contexte (déviances contextuelles) ou à un usage non conventionnel dénotant une certaine étrangeté (déviances quantitatives). Les déviances phraséologiques provoqueraient donc au minimum un effet d'*imprécision*, mesurable en termes de degré de gravité.

- La catégorie d'UP la plus lacunaire regroupe les UP à fonction référentielle (Margareta Lewis, 2005), appelées aussi *séquences préfabriquées lexicales* (Forsberg, 2006) ou *collocations lexicales* (Howarth, 1998b). Au sein des déviances collocationnelles repérables dans les UP référentielles à structure verbo-nominale, on distingue une majorité de déviances lexicales (6%) et quelques déviances syntaxiques (1%) (Howarth, 1998b). Les déviances lexicales portent principalement sur la sélection du verbe (Howarth, 1998b ; Nesselhauf, 2004, 2005). Une part importante des déviances serait due à la spécificité sémantique des verbes et à leur potentiel combinatoire (Howarth, 1998b). Après les emplois erronés du verbe,

viennent les emplois erronés dus à l'emploi d'une construction à verbe support au lieu d'un verbe simple, à la complémentation du nom erronée et à un mauvais emploi du nom (Nesselhauf, 2004).

Hypothèse 7

Comme pour l'anglais L2, les déviations portant sur les UP verbo-nominales, à un stade avancé d'apprentissage du français L2, seraient majoritairement des déviations lexicales et secondairement syntaxiques. La plupart des erreurs lexicales porteraient sur l'élément verbal. Viendraient ensuite les déviations portant sur les prépositions/ conjonctions (e.a. la complémentation du nom) et les déviations portant sur le nom.

- Pour les UP verbo-nominales construites avec les deux verbes à haute fréquence *take* et *give*, 28% des collocations avec *give* sont déviantes et 26% des collocations avec *take* sont déviantes. *Give* montre des déviations dans toutes les catégories, avec une majorité de déviations sur le verbe (4 occ.) et l'emploi de CVS (4occ.) ; *take* montre une majorité de déviations sur la complémentation du N (6 occ.), puis sur le N (2 occ.), enfin sur la CVS (1 occ.) et le V (1 occ.) (Nesselhauf, 2004).

Hypothèse 8

Les UP verbo-nominales construites avec les verbes *prendre* et *donner* auraient un comportement en partie comparable aux UP construites avec les deux verbes *take* et *give* en anglais. Selon ce point de vue, ¼ des UP construites avec *prendre* et *donner* seraient déviantes.

Nous tenterons par le biais de notre analyse sur corpus d'apprenants du français L2 de vérifier ou d'infirmer les hypothèses émises ci-dessus. De manière globale, notre objectif sera donc de déterminer la nature de l'écart entre des textes produits par des locuteurs non natifs (LNN) à un niveau avancé d'apprentissage et des locuteurs natifs (LN), dans des textes académiques de type expositif. Nous pensons que cet écart se traduirait en termes de *suremploi/ sous-emploi* et de *mésusage*, mais aussi en termes de *précision* et de *complexité*. En effet, les LNN tendraient à un niveau avancé d'apprentissage à se rapprocher le plus possible de l'usage natif (*native-like proficiency*) en s'exprimant avec *précision* et de manière de plus en plus *complexe*. En résumé, la question sera de déterminer quelle est la compétence phraséologique de deux groupes d'apprenants du français L2 ayant une L1 différente, par rapport à la compétence native.

Ces corpus nous ont ensuite permis de calculer la fréquence de cooccurrence des séquences verbo-nominales étudiées, en les désambiguïsant en contexte. En effet, les bases de données décrites précédemment ont été traitées de manière entièrement automatique et ne sont pas contextualisées, ce qui ne permet pas d'étudier l'appropriété sémantique des données en contexte. Par contre, les textes de nos corpus en français de Belgique sont disponibles et nous pouvons à tout moment nous référer à la source cotextuelle, de manière plus ou moins locale ou globale, pour désambiguïser sémantiquement les données quand cela s'avérait nécessaire. De plus, ces corpus sont *intégraux*, contrairement aux bases de données qui, par définition, ne le sont pas :

[U]n corpus n'est pas un corpus de mots (...); ni un corpus d'attestations ou d'exemples (comme Frantext, dès lors qu'on n'a pas accès aux textes-sources); ni un corpus de fragments (comme le British National Corpus, qui ne contient aucun texte complet, mais un échantillonnage). (Rastier, 2002: 2)

C'est ainsi que nous avons pu calculer la fréquence absolue (notée FqA) de cooccurrence pour le corpus journalistique (26,8 M de mots) et le corpus oral (3,8 M de mots), à partir de laquelle une fréquence relative (notée FqR) a ensuite été calculée par million de mots. Cette distinction entre fréquence dans l'usage écrit et dans l'usage oral nous permettra entre autres d'étudier l'influence du médium sur la production d'UP par les apprenants du français L2.

III.5. Corpus d'apprenants en français L2

Les corpus sur lesquels nous travaillons répondent à la définition qu'en a donnée Rastier (2002: 2):

Un corpus est un regroupement structuré de textes intégraux, documentés, éventuellement enrichis par des étiquetages, et rassemblés : (i) de manière théorique réflexive en tenant compte des discours et des genres, et (ii) de manière pratique en vue d'une gamme d'applications.

Les textes ont ainsi été sélectionnés en fonction de la perspective visant la description linguistique d'unités phraséologiques (UP) construites avec les verbes à haute fréquence *prendre* et *donner* dans des productions écrites académiques d'apprenants universitaires du français L2. Nous nous référons pour la perspective interlangagière à la définition que S. Granger (2004: 124) donne des *corpus informatisés d'apprenants*:

Electronic collections of spoken or written texts produced by foreign or second language learners (S. Granger, 2004: 124)

La nouveauté des études sur *corpus informatisés d'apprenants* (en anglais *computer learner corpus* – CLC) (S. Granger, 2004), qui s'imposent comme discipline depuis les années 80' en linguistique anglo-saxonne, réside dans leur mode de compilation qui est informatisé et permet de traiter un grand nombre de données. La grande taille des corpus garantit ainsi une certaine représentativité de la langue en usage et le caractère généralisable des résultats.

Whether collected electronically over a very short period of time or after years of painstaking work, current learner corpora tend to be rather large, which is a major asset in term of representativeness of the data and generalizability of the results. Of course, a very large data sample is not necessary for all types of SLA research. A detailed longitudinal study of one single learner is of great value if the focus is on individual interlanguage development. (...) In addition, as we will see in the following section, size is only really useful if the corpus has been collected on the basis of strict design criteria. (S. Granger, 2004: 125)

Toutefois, comme le souligne S. Granger (2004: 125, *supra*), le recours à de larges corpus n'est véritablement utile que lorsqu'ils ont été constitués en répondant à des critères stricts de sélection, même s'il est utopique de vouloir prendre en considération tous les facteurs susceptibles d'influencer la production des apprenants.

It should be wrong, however, to paint too rosy a picture of current CLC. In all fairness, one must admit that (a) there are not many tightly-designed learner corpora in the public domain, and (b) there are so many variables that influence learner output that one cannot realistically expect ready-made learner corpora to contain all the variables for which one may want to control. Ideally, as stated by Biber (1993: 256), "theoretical research should always precede the initial design and general compilation of texts." This preliminary theoretical analysis is the only way to ensure that the corpus will contain all the relevant design parameters." (S. Granger, 2004: 126)

Dans cet ordre d'idée, une analyse théorique préliminaire est indispensable à l'élaboration et à la compilation des corpus, pour permettre au chercheur de réunir des informations sur les corpus en fonction des variables prédominantes, i.e. celles qui sont les plus susceptibles d'influencer la production de l'apprenant. Nous ajouterons, à l'instar de Rastier (2002), que ces critères doivent également être déterminés en fonction des objectifs visés par l'étude en cours, i.e. « en vue d'une gamme d'applications » particulières (Rastier, 2002: 2 – cf. citation *supra*).

Les corpus d'apprenants du français L2 anglophones (185.000 mots) et néerlandophones (90.000 mots) sont majoritairement issus du corpus informatisé d'apprenants FRIDA (*FRENch Interlanguage DATabase*, Cassart *et al.*, 2002) dont les textes ont été rassemblés, sous la direction de S. Granger, à l'Université catholique de Louvain au sein du CECL (*Centre for English Corpus Linguistics*). Le corpus de contrôle en français L1 (185.000 mots) se compose quant à lui majoritairement de textes produits

par des étudiants universitaires francophones de Belgique, rassemblés à l'occasion de notre mémoire de DEA (CaFLAM – *Corpus d'apprenants du Français Langue Maternelle*), et secondairement du corpus CODIS (*CORpus de DISsertations françaises*, CECL) constitué également de dissertations produites par des étudiants francophones à l'université. Bien que le corpus anglophone soit comparable en taille au corpus de contrôle, ce n'est pas le cas du corpus néerlandophone qui est deux fois plus petit. Nous avons donc effectué nos calculs et nos tests statistiques sur la base du plus petit dénominateur (90.000 mots) pour obtenir des fréquences relatives comparables entre elles.

	Corpus d'étude (FL2)		Corpus de contrôle (FL1)
	LNN anglophones 237 textes	LNN néerlandophones 221 textes	LN francophones 330 textes
FRIDA (CECL)		FRIDA (CECL) 60.000 mots	CaFLAM 132.000 mots
		Extension FRIDA 30.000 mots	CODIS (CECL) 53.000 mots
Total :	185.000 mots	90.000 mots	185.000 mots

Tableau 12. Analyse sur corpus d'apprenants – Description des corpus non natifs et natif

Étant donné la taille de notre corpus et le temps qu'une analyse individuelle demanderait pour déterminer le niveau de compétence linguistique représenté par chaque texte⁶¹, nous n'avons pas utilisé de critères internes pour déterminer le choix de nos corpus. Dans cette étude, les critères externes utilisés pour déterminer le niveau des textes sont ceux qui caractérisent traditionnellement la *variété avancée* en recherche acquisitionnelle (Bartning, 1997 – cf. II.1.4.1., pp. 82-83): l'acquisition guidée (*input* par l'enseignement en situation scolaire), l'apprentissage de la langue cible comme langue étrangère et le niveau d'instruction (cadre institutionnel universitaire).

Dans le corpus FRIDA, une distinction est faite entre les apprenants *spécialistes* de la langue française, i.e. « the learners [who] are university under-graduates of French », et les apprenants *non spécialistes*, i.e. « the learners [who] specialize in a discipline other than French » (Cassart *et al.*, 2002 : 4). Dans notre sous-corpus extrait de FRIDA, cette distinction n'a cependant pas été maintenue. En effet, la variable concernant le statut de spécialiste ou non des apprenants est impossible à contrôler ici pour deux raisons : (i) certains groupes montrent une hétérogénéité puisqu'ils regroupent des non-spécialistes et des spécialistes de la langue française dans une même classe (par ex. les

⁶¹ Nous avons néanmoins procédé à une analyse morphosyntaxique de 75 textes argumentatifs de notre corpus (25 textes en français L1, 25 textes en FL2-NL et 25 textes en FL2-AN), grâce au logiciel *Direkt Profil* élaboré à l'université de Lund (Suède). Les résultats ont montré que les corpus non natifs étudiés se caractérisaient au minimum par un niveau de compétence morphosyntaxique *avancé moyen*.

textes <TAN093> à <TAN112> sont produits par des étudiants en *French Language Studies*, en *Mathematics and French* et en *International Managment with French* au sein d'une même classe) ; (ii) les informations que nous possédons concernent généralement les variables du groupe et rarement les variables individuelles. Les conclusions que nous tirons de notre analyse entrent donc dans un cadre non individualisant, étant donné la documentation limitée et l'hétérogénéité du corpus dont nous disposons pour cette recherche.

Le corpus FRIDA est donc hétérogène puisqu'il regroupe d'une part des textes de niveau intermédiaire à avancé et, d'autre part, des textes écrits par des spécialistes et des non-spécialistes de la langue française. Les textes qui en ont été extraits pour notre étude seront néanmoins définis *a priori* comme étant représentatifs de la variété avancée du français L2. En effet, en tant qu'apprenants *universitaires* ayant opté pour le français comme langue de spécialisation (*apprenants spécialistes*) ou comme langue de communication au sein d'une spécialisation autre que la langue française (*apprenants non spécialistes*), ces apprenants appartiennent tous à la catégorie des *apprenants avancés* ayant un *niveau d'instruction élevé* (AI) (cf. Bartning, 1997 : 12-13 – cf. II.1.4.1., p. 82).

Une analyse de compétence morphosyntaxique, effectuée par le biais du logiciel *Direkt Profil* (Granfeldt *et al.*, 2005), sur 75 textes argumentatifs tirés des corpus L2 et L1 (25 textes sélectionnés de manière aléatoire dans chacun des trois corpus), a en outre révélé que l'hétérogénéité des niveaux de compétence individuelle au sein des groupes d'apprenants ne semblait pas affecter le niveau de compétence morphosyntaxique global de ces groupes. La moyenne des profils obtenus pour chaque corpus correspond en effet au minimum au stade *avancé moyen* (Stade V), tel que défini par Bartning et ses collaborateurs (Bartning, 1997 ; Bartning & Schlyter, 2004).

Par ailleurs, les textes retenus du corpus FRIDA sont des textes *académiques* qui se caractérisent pour la plupart par leur structure *expositive*. Les textes académiques, définis comme un genre de textes circulant au sein d'une communauté de locuteurs (en l'occurrence la communauté académique ou scientifique) (cf. *infra*, la notion de *genre* en relation avec celle de *discourse community* – Swales, 1990 : 58), sont des textes écrits généralement dans un style formel, destinés à un public critique et instruit.

A genre comprises a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the expert members of the parent discourse community, and thereby constitute the rationale for the genre. (Swales, 1990: 58).

La fonction première des textes académiques est le plus souvent de transmettre des informations (comme c'est le cas des textes *expositifs*) ou d'argumenter (comme c'est le cas des textes *argumentatifs* et *persuasifs*). Les textes *expositifs* se caractérisent pour leur part par leur structure organisationnelle *hiérarchisée* autour de la notion de topique, ce qui les distingue des textes *narratifs* dont la structure organisationnelle est *chronologique*.

Les textes *expositifs* sont ainsi définis structurellement comme des textes dont le mode organisationnel se construit autour de la notion de topique. Ils mettent en jeu des concepts à propos desquels ils apportent des informations, à la différence des textes *narratifs* généralement organisés par des relations entre participants et événements. (Hò-Đác, 2007: 25).

Les étudiants de l'enseignement secondaire et universitaire ont très souvent recours aux textes de type *expositif*. Il est généralement recommandé dans ces textes d'user d'un style impersonnel et neutre, à la troisième personne, puisque l'accent est mis sur la transmission des informations. La structure-type d'un texte *expositif* suit les étapes prônées en argumentation de texte, à savoir l'organisation en trois parties: introduction, développement argumentatif et conclusion.

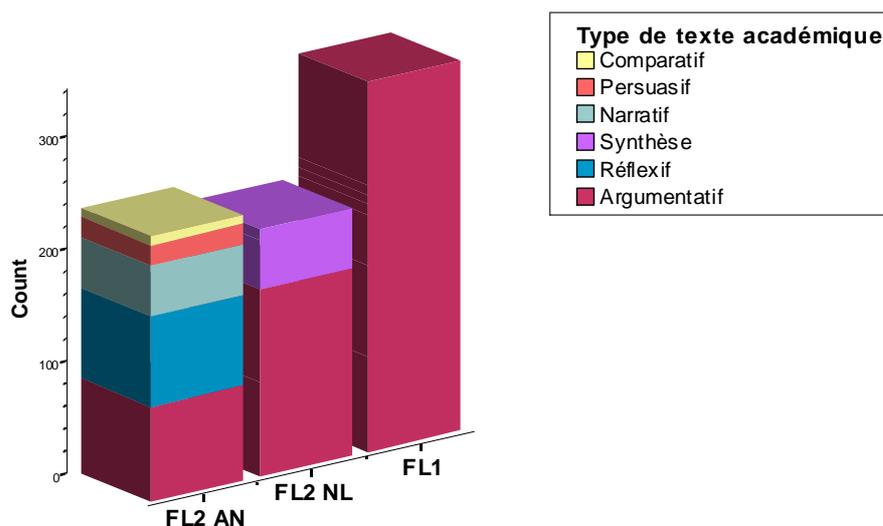
La majorité des textes étudiés dans cette étude sont *expositifs* (94% des textes), hormis une partie de textes *narratifs* (19% des textes du corpus FL2-AN). Nous avons néanmoins conservé ces derniers car ils rentrent, au même titre que les textes *expositifs*, dans la catégorie plus large des textes académiques correspondant à ce que nous appelons en français des *rédactions* ou des *compositions* produites en contexte d'évaluation scolaire (en anglais *academic essays*).

Nous distinguerons ainsi différents types de *compositions académiques*, dont la catégorisation repose sur la combinaison de deux paramètres distincts, en l'occurrence l'*enjeu pragmatique* des *discours* (par ex. dire, convaincre, expliquer, prescrire, questionner, résumer, etc.) et l'*organisation* des *textes* (narration, description, exposition) qui se caractérisent par « une forme, une logique interne, principe de progression et d'organisation, et un mode de clôture » (Mas & Turco, 1991: 91) : (i) les textes à visée *comparative*, dont le but est généralement de montrer en quoi résident les différences et/ou les similitudes apparentes entre des points de vue, des sujets ou des textes donnés ; (ii) les textes à visée *persuasive*, où le but de l'auteur est de convaincre le lecteur de la pertinence de son point de vue ; (iii) les textes à visée *réflexive*, qui traitent de sujets abstraits, relevant notamment du domaine social, politique, philosophique ou religieux ; (iv) les textes à visée *narrative*, qui visent à décrire une situation ou une suite d'événements selon un point de vue subjectif, souvent à la première personne ; (v) les

textes de *synthèse*, qui ont pour objectif de résumer le propos d'un texte préalablement soumis à l'auteur ; (vi) les textes à visée *argumentative*, où il s'agit de débattre d'une problématique sujette à controverse en développant une batterie d'arguments pour asseoir son point de vue. Les textes d'argumentation stricte (en rouge dans le Graphique 2, ci-dessous), i.e. les dissertations, représentent 74% du total des textes analysés. Nesselhauf (2005) a formulé les avantages que représente l'étude de ce type de textes:

Argumentative essays (...) have been chosen because they are fairly neutral in register and style (i.e. they contain non-specialized vocabulary and tend to display a medium level of formality) and will therefore reveal difficulties with those collocations that most learners often need. (Nesselhauf, 2005: 10)

Notons que le corpus de contrôle en français L1 est composé exclusivement de dissertations et que ces dernières représentent 76% du corpus néerlandophone, contre 24% de textes de synthèse relatant des sujets de presse écrite. Au-delà de cette relative homogénéité, le corpus anglophone dénote une hétérogénéité de types textuels: les textes sont majoritairement des dissertations (35,9%) ou des textes réflexifs portant sur un sujet économique ou sur l'histoire de la francophonie (34%) ; viennent ensuite les textes narratifs (19%) ; enfin, les textes restants sont, d'une part, des textes persuasifs ayant la forme d'un discours de campagne électorale (8%) et, d'autre part, des textes comparatifs en analyse de textes littéraires (3%). Cette diversité au sein des textes académiques dans le corpus d'apprenants anglophones sera prise en compte lors de nos analyses de corpus.



Graphique 2. Distribution des types de textes académiques dans les corpus non natifs, anglophone et néerlandophone, et dans le corpus natif

Les corpus sont utilisés dans leur forme brute⁶² et mettent à la disposition du chercheur les informations extralinguistiques concernant: (i) la langue maternelle de l'apprenant ; (ii) la source textuelle (code du texte, nombre de mots et type de texte) ; (iii) le contexte d'apprentissage (nom et pays de l'institution scolaire). Nous proposons en annexe (cf. Annexe 4, cd-rom) un tableau contenant les informations extralinguistiques relatives à chacun des textes (788 textes) des corpus en français académique (L2 et L1), que nous ne présentons pas ici pour des raisons évidentes de manque de place. Nous nous référerons à ces informations, et en particulier à la langue maternelle de l'apprenant, au code du texte et au type de texte, en fonction des besoins de l'analyse.

III.6. Données de corpus

Diverses raisons nous ont poussée à étudier des unités phraséologiques (UP) construites avec des verbes à haute fréquence. La première raison, qui est d'ordre méthodologique, tient au fait que l'analyse de corpus informatisés d'apprenants requiert l'étude d'un phénomène linguistique suffisamment fréquent pour que les résultats de l'analyse soient généralisables.

The issue is how many tokens of a linguistic item are necessary for descriptive adequacy. In a typical text (or a corpus of one million words), 40 to 50 % of the word types occur only once. In the case of polysemous words, at least half of attested meanings of words occur only once. (...) But if once is too little, how many tokens are enough to discover the patterns of collocation, of polysemy, of morphological characteristics and of the likelihood of occurrence of a given form or meaning? (Kennedy, 1998:67)

Selon Kennedy (1998), le nombre d'occurrences requis pour mener une étude descriptive généralisable est fixé à au moins 1.000 occurrences observées du phénomène étudié dans le corpus. Par ailleurs, Kennedy précise que ce chiffre varie d'une étude à l'autre, en fonction de la méthode adoptée et des objectifs spécifiques visés par le chercheur. Alors que pour des phénomènes linguistiques peu fréquents, tels que les collocations, de très larges corpus sont nécessaires, les études portant sur des unités lexicales caractérisées par une haute fréquence dans l'usage nécessitent des corpus moyens, i.e. comprenant, selon l'auteur, un demi à un million de mots (Kennedy, 1998: 68). La rareté qui caractérise les unités phraséologiques hautement figées est donc ici compensée par la haute fréquence de verbes *prendre* et *donner*. Plusieurs raisons linguistiques et didactiques ont ensuite motivé notre choix, qui sont à la base du projet de recherche collectif dans le cadre duquel

⁶² Un corpus *brut* est un corpus non étiqueté, i.e. qui n'a fait l'objet d'aucune annotation.

s'insère ce travail⁶³: les verbes à haute fréquence se caractérisent par leur nature polysémique (De Cock & Granger, 2004) et leur capacité à générer un grand nombre d'unités phraséologiques (Howarth, 1998b) ; les verbes à haute fréquence occupent une place centrale dans l'acquisition du vocabulaire, tant dans l'apprentissage d'une L1 que d'une L2 (Schmitt & Carter, 2004), puisqu'ils font l'objet d'un apprentissage précoce et tendent à être suremployés ou sous-employés à un niveau avancé d'apprentissage (Altenberg & Granger, 2001 – cf. II.4.2.3., pp. 130-131) ; on retrouve des équivalents de ces verbes dans beaucoup d'autres langues, ce qui peut occasionner leur suremploi par l'apprenant quand il y a similitude formelle et sémantique, et favoriser des mécanismes de transfert positif ou négatif. Nesselhauf (2005, cf. citation *infra*) montre pour sa part que les constructions verbo-nominales sont centrales en acquisition d'une L2, puisqu'elles sont par excellence le lieu de l'expression en termes de prédicat-argument (ou thème-rhème), elle-même au fondement de toute communication langagière.

Verb-noun combinations, i.e. combinations such as *make an attempt* or *take sth. into account*, have been chosen because they are not only frequent (...) and among the most difficult for the learner (...), but also particularly important, since “they tend to form the communicative core of utterances where the most important information is placed” (Altenberg 1993:227). In addition, unlike in adverb-adjective or adjective-noun collocations, it is not possible for the learner to leave out the difficult element or replace it with a safe choice (such as *very*); and paraphrasing is also often impossible (...). (Nesselhauf, 2005: 9-10)

Après avoir abordé les raisons qui ont motivé le choix de notre objet d'étude, nous allons évoquer maintenant la méthode utilisée pour extraire nos données des corpus.

Pour extraire de nos corpus les 906 séquences verbo-nominales (SVN) contenant les verbes *prendre* et *donner* ainsi que leurs formes fléchies, nous avons utilisé la fonction *Concord* du concordancier développé par Mike Scott (*Oxford WordSmith Tools 4.0*). Rappelons qu'un *concordancier* est un programme informatique permettant d'extraire d'un corpus informatisé un item linguistique, ou *mot-clé*, en contexte. À chaque forme de mot (*type*) qui a été entrée en tant que terme de recherche correspondent une ou plusieurs occurrences (*tokens*).

A concordance is a formatted version or display of all the occurrences or tokens of a particular type in a corpus. The type is usually called a keyword but is sometimes referred to as a target item, node word or search item. (Kennedy, 1998: 259-250)

Cet item peut être, selon le programme utilisé, un mot graphique avec ou sans ses formes dérivées/ fléchies, une séquence de mots, une catégorie grammaticale ou un lemme. Pour

⁶³ Pour rappel, ce projet collectif est intitulé *Apprentissage des langues étrangères: niveaux phraséologique et discursif* (ARC 03/08-301) et porte sur l'appropriation des langues étrangères (français, néerlandais et anglais), essentiellement en situation de production écrite.

notre étude, nous avons créé des fichiers bruts (.txt) listant les différentes formes verbales conjuguées pour les deux verbes, y incluant les formes non natives potentielles (par ex. **prennant* pour *prenant*). Le nombre de lexèmes est en effet limité quand on veut entrer simultanément plusieurs termes de recherche pour une seule concordance. La concordance désigne quant à elle la liste des fragments textuels successifs qui ont été extraits du corpus analysé (cf. Figure 7, ci-dessous).

N	Concordance
1	problèmes communautaires et les députés dans la Chambre des Représentants prennent les décisions. Si on veut savoir l'opinion des gens, on de
2	ant le dernier semestre. A la suite de cette flambée, la direction du journal a pris des décisions urgentes et nécessaires. La moitié des journali
3	. Il y a aussi de la revanche du côté des militaires amicaux au Mabutu. Ils ont pris position des points stratégiques, entre autres de chaînes de t
4	ar exemple, les enquêtes pour faire des conclusions sur l'opinion du public. On prend un sondage et tous les gens qui sont dedans doivent repon
5	La majorité dans la Chambre et le Sénat veut l'abdication du Roi à cause de sa prise de position en faveur du pouvoir exécutif. Selon la Constitutio
6	train de dégager des lois, des régularités dans la vie sociale; on doit d'abord prendre un échantillon de la population. Après on doit établir un c
7	les ministres et les parlementaires ont le droit d'initiative, quand un ministre prend l'initiative, on parle d'un projet de loi. Par contre, l'initiative d'un
8	ologue a effectué un étude concernant les habitudes anormales des wallons. Il a pris un échantillon de 10000 répondants dans la Région wallone. E
9	pour une loi peut énoncer du gouvernement ou du parlement : si l'initiative est prise par le gouvernement on parle d'un projet de loi, sinon il s'agit

Figure 7. Extrait de la concordance effectuée pour le verbe prendre dans les corpus non natifs et natifs

Dans une concordance, chaque fragment est présenté sur une seule ligne, au centre de laquelle se trouve le *mot-clé*.

A concordance, or rather KWIC (Key Word In Context) concordance, is a list of unconnected lines of text, which have been summoned by the concordance program from a computer corpus, that is a collection of texts held in a form which is accessible to the computer. At the centre of each line is the item being studied (keyword or node). The rest of each line contains the immediate co-text to the left and right of the keyword. (Partington, 1998: 9)

En plus de la fonction *Concord*, décrite ci-dessus, nous nous sommes servie de la fonction du programme qui donne accès à la source textuelle dont est tirée la ligne de concordance (*View*). Cette fonction nous a permis d'élargir le cotexte quand cela s'avérait indispensable, notamment pour lever des ambiguïtés sémantiques lors de l'analyse paramétrique.

À cette première étape d'extraction automatique succède un tri manuel des données. Une étape de désambiguïsation syntaxique et sémantique nous a ainsi permis de mieux cibler notre objet d'étude. Rappelons que la *désambiguïsation*, qui consiste « à choisir entre un certain nombre de possibilités concurrentes » (Habert *et al.*, 1997 : 12), peut être lexicale/ sémantique, syntaxique ou morphosyntaxique. Notre objectif étant d'analyser de manière la plus exhaustive possible la combinatoire des SVN contenant les verbes *prendre* et *donner*, nous n'avons gardé de la concordance initiale que les 906 fragments où l'item avait une fonction verbale, excluant les formes nominales et

déverbaux⁶⁴. Ces dernières sont en effet régies par des règles syntaxiques et lexicales qui leur sont propres.

Exemple d'emploi déverbal nominal (FL2-NL, <TNL022>):

(14) La majorité dans la Chambre et le Sénat veut l'abdication du Roi à cause de sa **prise** de position en faveur du pouvoir exécutif.

Exemple d'emploi nominal (FL2-AN, <TAN072>):

(15) Une enquête ouverte par la police indique que l'incendie commença comme résultat d'une **prise** de courant défectueuse, mais ce n'est pas la seule cause possible

Ensuite, afin de procéder à une analyse homogène basée sur des séquences comparables syntaxiquement, nous avons opéré une seconde sélection en conservant uniquement les fragments dont la structure syntaxique était un syntagme verbal (SV) avec ses compléments nominaux essentiels (SN1 et/ou SN2), soit sous forme (partiellement) pronominale avec référent repérable dans le contexte,

Exemple d'emploi pronominal avec référent explicite (FL1, <TFRB280>, occ. 782):

(16) La famille est un concept qui traverse les âges et dont chaque génération **prend conscience** très tôt dans sa vie



soit sous forme nominale lexicalisée (avec ou sans préposition).

Exemples de SVN avec compléments a-prépositionnel (SN1) et prépositionnel (SN2) :

(17) Oberon **a donné** une fleur rouge [=SN1] à Puck [=SN2], parce qu'il est en colère contre quelqu'un qui s'appelle Titania, la reine des fées (FL2-AN, <TAN039>, occ. 330)

(18) Un exemple simple est celui des adolescents : combien **prennent** pour modèle [=SN2] une star [=SN1] et vont jusqu'à recopier parfaitement faits et gestes de leur idole ? (FL1, <TNL022>, occ. 722)

Exemple de SVN avec complément a-prépositionnel (SN1) (FL1, <TFRB146>, occ. 705) :

(19) Lire un ouvrage **prend** un temps variable selon les envies du lecteur alors que le film est fort court.

Contrairement à notre étude pilote sur le degré de figement en français L1 (cf. I.1.4.), nous avons conservé ici les structures passives, notre objet n'étant plus d'appliquer des transformations syntaxiques aux SVN étudiées. Nous obtenons ainsi 906 SVN construites avec les verbes *prendre* ou *donner*, ou l'une de leurs formes fléchies, extraites des deux corpus d'apprenants non natifs et du corpus natif.

⁶⁴ Contrairement à notre étude pilote sur le degré de figement des UP avec le verbe *prendre* en français L1 (cf. I.1.4.), nous avons exclu de notre champ d'étude les structures déverbaux, à fonction nominale, pour pouvoir nous centrer exclusivement sur les unités ayant une fonction syntaxique verbale. Il serait néanmoins intéressant d'étudier ces déverbaux dans une perspective d'analyse de la complexité syntaxique des emplois phraséologiques, puisque la nominalisation semble être un trait tardif, voire natif, en acquisition d'une L2 (Tutin, 2005 ; Forsberg, 2006).

SVN	Corpus d'étude		Corpus de contrôle	Total
	FL2-AN 185.000 mots 237 textes	FL2-NL 90.000 mots 221 textes	FL1 185.000 mots 330 textes	
Prendre	178	75	217	470
Donner	211	94	131	436
Total	389	169	348	906

Tableau 13. Distribution des séquences à structure verbo-nominale dans les corpus non natifs et natifs

Pour procéder à l'analyse proprement dite des SVN extraites de nos corpus, nous adopterons une démarche qui prend appui sur trois approches distinctes en acquisition d'une L2, abordées ici de manière complémentaire: les étapes méthodologiques issues de l'analyse d'erreurs, l'analyse de fréquence des formes déviantes et non déviantes de l'analyse contrastive interlangagière (héritée de la linguistique de corpus), et l'analyse des interlangues en termes acquisitionnels de *complexité*, *précision* (et *fluidité*). Nous décrivons dans le Chapitre IV la seconde étape prônée en analyse d'erreurs, en l'occurrence l'*identification* et la *description* des données d'analyse, qui consistera dans ce travail à déterminer quelles sont les SVN phraséologiques (*vs* libres) et quelles sont les déviances phraséologiques commises par les locuteurs natifs et non natifs.

Pages 169 à 224 non consultables

UP déviantes produites par les locuteurs natifs, 18,7% des UP déviantes produites par les anglophones et 10,5% par les néerlandophones.

(121) Le cinéma nous **|donne|** et transmet des émotions fortes par les images et par la mise en scène des personnages et cela se passe aussi dans les romans par l'écriture et par notre propre identification, attachement aux personnages (FL1, <TFRB273>, occ. 916).

→ *Le cinéma **suscite** en nous/ nous **procure** des émotions fortes*

Nous présentons également pour cet exemple (121) un tableau détaillé des étapes d'analyse paramétrique effectuées à partir de nos données de corpus :

Nombre de déviations dans l'UP <i>GravDév10Total</i>	Type de déviation <i>GravDév11Type</i>	Localisation de la déviation <i>GravDév12Local</i>	Indice d'IM en L1 <i>GravDév13IMUP</i>	FqR en L1 <i>GravDév14FqUP</i>
Gravité 1	Gravité 1	Gravité 2	Gravité 2	Gravité 1
1 déviation	Déviation quantitative	Déviation sur V	$1,029 \leq IM \leq 2,959$	$0 \leq FqR \leq 10,86$

Tableau 45. Détail de l'analyse d'une UP déviante dont le degré de gravité est faible (occ. 916)

Dans le cadre de l'étude de *précision* que nous effectuerons dans le Chapitre VI de ce travail, nous aurons l'occasion de procéder à une manipulation supplémentaire à partir de ces paramètres (en l'occurrence au recodage de la variable *GravDév16Global*) pour pouvoir appliquer à nos données de corpus la méthode proposée par Foster & Wigglesworth (2007) (cf. Chapitre VI).

IV.4. Conclusion intermédiaire

Par le biais de l'analyse paramétrique que nous avons effectuée de nos données de corpus, nous avons pu dégager des résultats qui nous ont permis de formuler quelques tendances et de vérifier quelques-unes des hypothèses émises en III.2.. Nous proposons ci-dessous une synthèse de ces résultats.

Après avoir effectué une analyse collocationnelle fine des SVN observées (cf. IV.2.), nous avons constaté qu'une part très importante de nos données (77,6% des cas) appartenait au champ phraséologique, avec une différence de proportion entre le corpus natif qui comprend 85,1% d'UP, le corpus non natif néerlandophone qui en comprend 81,1% et le corpus non natif anglophone 69,4% [Première partie de l'Hypothèse 4 vérifiée].

Ensuite, nous avons procédé à une analyse détaillée des déviations phraséologiques au sein des corpus L2 et L1 (cf. IV.3.), en décrivant (i) les *types* de déviations en cause (formelles, contextuelles et/ou quantitatives), (ii) la *localisation* des déviations au sein des UP et (iii) leur *degré de gravité*.

Premièrement, il est ressorti de cette analyse de déviance que les locuteurs natifs, qui commettent peu de déviations phraséologiques (6,8%), produisent 40% d'UP déviantes de type quantitatif, contrairement aux apprenants non natifs qui produisent essentiellement des déviations formelles (40,7% pour les anglophones et 52,6% pour les néerlandophones) et secondairement des déviations contextuelles (26,4% pour les anglophones et 21,1% pour les néerlandophones) [Première partie de l'Hypothèse 6 vérifiée].

Deuxièmement, les catégories grammaticales qui sont le plus souvent le lieu de la déviance – hormis les déviations, dites *multiples*, qui portent sur plusieurs constituants⁸⁷ (excluant les déviations portant sur toute l'UP, cf. *supra*) et concernent 30% des cas – sont d'abord les verbes pour tous les apprenants (L2 et L1) (plus de 25% des UP déviantes), puis, dans les corpus non natifs, les constituants nominaux (11% pour les anglophones et 16% pour les néerlandophones) [Première partie de l'Hypothèse 5 vérifiée et Hypothèse 7 partiellement vérifiée].

Troisièmement, nous avons constaté que les UP présentant un degré important de gravité étaient rares (5,4% des déviations), alors que les déviations moyennement graves concernent 75,8% des UP déviantes. Le degré de gravité le plus bas a quant à lui été donné à un peu moins de 20% des cas [Seconde partie de l'Hypothèse 5 partiellement vérifiée et seconde partie de l'Hypothèse 6 vérifiée]. Les proportions semblent en outre relativement similaires entre les deux corpus non natifs, qui montrent une majorité de déviations moyennement graves (74,7% pour les anglophones et 86,8% pour les néerlandophones), par contraste avec les déviations ayant un degré de gravité faible ou élevé, qui concernent moins d'1 déviance sur 5. Notons que cette analyse du degré de gravité des déviations sera utile pour l'étude de précision que nous effectuerons dans le chapitre VI de ce travail.

Dans le chapitre suivant, nous tenterons de cerner quelle influence peuvent avoir certains paramètres linguistiques et extralinguistiques sur la production d'UP (déviantes ou non) par des apprenants du français écrit L2.

⁸⁷ Rappelons que les déviations *multiples* n'incluent pas les cas de déviations portant sur toute l'UP, i.e. à la fois sur le verbe et le nom dans une même séquence.

Chapitre V: Résultats 2 – Variables linguistiques et extralinguistiques

Notre analyse sur corpus d'apprenants a porté sur l'étude de 906 séquences verbo-nominales (SVN), dont 703 UP, tirées des corpus d'apprenants en L2 et du corpus de contrôle en français L1. Il s'agira d'étudier ici chacune de ces UP du point de vue des quelques variables linguistiques et extralinguistiques les plus susceptibles de jouer un rôle dans la production d'unités phraséologiques (UP), déviantes ou non, par des apprenants du français L2. Nous opérons tout d'abord une distinction entre les variables d'ordre formel (cf. V.1.) et celles d'ordre sémantique (cf. V.2.), qui nous permettra d'élaborer un profil syntaxique et sémantique des UP étudiées. Nous aborderons ensuite deux facteurs externes d'acquisition, en l'occurrence le médium langagier (cf. V.3.1.) et le type de texte (cf. V.3.2.).

V.1. Caractéristiques formelles

V.1.1. *Prendre vs donner*

V.1.1.1. Suremploi et sous-emploi

V.1.1.2. Perspective explicative : système flexionnel

V.1.1.3. Mésusage

V.1.1.4. Synthèse des résultats

V.1.2. Profil syntaxique des syntagmes nominaux

V.2. Caractéristiques sémantiques

V.2.1. Classification sémantique

V.2.1.1. Noms concrets, vivants et abstraits

V.2.1.2. Emplois prédicatifs : états, événements et actions

V.2.2. Profil sémantique des UP

V.3. Médium langagier et types de textes

V.3.1. Code oral

V.3.2. Types de textes

V.3.2.1. Distribution des UP

V.3.2.2. Profil syntaxique et sémantique des UP

V.3.2.3. UP avec *prendre vs donner*

V.4. Conclusion intermédiaire

V.1. Caractéristiques formelles

Les deux informations que nous étudierons de manière plus approfondie, sur le plan formel, sont la nature du verbe à haute fréquence (cf. la variable *VHF* du Tableau 46, ci-dessous) et le type de structure syntaxique du syntagme nominal (SN) postverbal constitutif des UP étudiées (cf. la variable *SyntaxSNUP* du Tableau 46, *infra*).⁸⁸ Nous analysons successivement chacun de ces paramètres dans les pages qui suivent.

Nom de la variable (<i>Name</i>) ⁸⁹	Variable (<i>Label</i>)	Valeurs (<i>Values</i>)
VHF	Nature du verbe à haute fréquence	1. Prendre 2. Donner
SyntaxSNUP	Structure syntaxique du SN postverbal de l'UP	1. Dét zéro + N + (Prép) 2. Dét Poss + N + (Modifieur) 3. Dét + N + (Modifieur) 4. <i>N de N</i> 5. N + P Sub 6. Prép + SN

Tableau 46. Extrait de la liste paramétrique – Variables formelles

V.1.1. Prendre vs donner

Rappelons qu'à l'instar de S. Granger (2004), nous pensons que le suremploi et le sous-emploi de formes linguistiques, qu'elles soient déviantes ou non, « contribute to the foreign-soundingness of advanced interlanguage even in the absence of downright errors » (Granger, 2004: 132). Dès lors, l'objectif de ce point sera double : (i) décrire ce qui, d'un point de vue statistique, distingue les productions écrites d'apprenants avancés du français L2 par rapport à des productions natives, en termes de suremploi et de sous-emploi ; (ii) déterminer quelles seraient les déviances spécifiques aux deux verbes étudiés, en termes de mésusage.

V.1.1.1. Suremploi et sous-emploi

Il est ressorti des premiers résultats obtenus par le biais de notre analyse paramétrique que les apprenants du français L2, comparés aux natifs, ne suremployaient ni ne sous-employaient les UP construites avec les deux verbes à haute fréquence *prendre* et *donner* (cf. IV.2.2., p. 193). Nous avons néanmoins mis en évidence que les LNN anglophones

⁸⁸ Rappelons que ces valeurs sont toujours attribuées en regard des productions originales des apprenants, quel que soit le statut, déviant ou non, de l'UP en question.

⁸⁹ Dans les tableaux présentant les variables paramétriques, les mots entre parenthèses et en italique renvoient aux termes anglais utilisés par le logiciel *SPSS* (cf. III.3.2.).

produisant deux fois plus de SL que les locuteurs natifs. Le sous-emploi/ suremploi des UP n'avait donc pas été démontré à ce stade de l'analyse.

Cependant, après avoir étudié de plus près la distribution des emplois phraséologiques des deux verbes dans les trois corpus, nous avons constaté une différence de fréquence entre les SVN construites avec le verbe *prendre* et celles construites avec le verbe *donner* (cf. Tableau 47).

	FL2-AN	FL2-NL	FL1	Total
Prendre	65,68 [135] 29,45% (-) 50,0%	62 27,79% (-) 45,3%	95,35 [196] 42,75% 66,2%	223,03 [393] 100% 55,9%
Donner	65,68 [135] 34,7% (+) 50,0%	75 39,6% (+) 54,7%	48,65 [100] 25,7% 33,8%	189,33 [310] 100% 44,1%
Total	131,36 [270] 100,0%	137 100,0%	144 [296] 100,0%	412,36 [703] 100,0%

Tableau 47. Distribution des UP dans les corpus non natifs et natifs, en fonction du verbe à haute fréquence

Nous avons mis en évidence un sous-emploi des UP avec *prendre* et un suremploi des UP avec *donner*. Les deux verbes, bien qu'étant tous les deux des verbes à haute fréquence à haut potentiel phraséologique, se différencient donc pourtant l'un de l'autre dans les productions des LNN. Ce résultat paradoxal est le reflet de notre tout aussi paradoxale Hypothèse 2 :

Hypothèse 2

La compétence phraséologique des apprenants d'une L2, concernant les UP construites avec des verbes à haute fréquence, se caractériserait paradoxalement par un écart en termes de (a) suremploi et/ou de (b) sous-emploi par rapport à l'usage natif.

a) Les apprenants avancés non natifs auraient tendance à recourir à des *doudous polylexicaux*⁹⁰, ce qui se traduirait à la fois par un suremploi de certaines UP *types* et par une diminution de la richesse lexicale des productions phraséologiques des apprenants en L2.

b) Les constructions délexicales verbo-nominales seraient sous-employées par les apprenants avancés.

Selon les résultats statistiques obtenus, la différence d'emploi des UP construites avec *prendre* et *donner*, en L2 vs L1, est hautement significative ($\chi^2 = 13.716$, ddl = 2, $p < *0.005$). En effet, la proportion d'UP construites avec *prendre* représente en L1 2/3 des UP pour 1/3 seulement avec *donner*. Par contre, l'apprenant anglophone avancé produit la même proportion d'UP avec *prendre* (50%) qu'avec *donner* (50%) et l'apprenant néerlandophone une proportion relativement équivalente, avec néanmoins légèrement plus d'UP construites avec *donner* (54,7%) qu'avec *prendre*. Ces résultats très significatifs traduisent donc une tendance des apprenants avancés du français L2

⁹⁰ Nous rappelons que nous avons proposé cette appellation pour traduire en français le concept anglais de *phrasal teddy bears* (Jaglinska, 2006), s'inspirant lui-même du concept de *lexical teddy bears* (Hasselgren, 1994 - cf. II.4.2.3., pp. 130-131).

anglophones ($\chi^2 = 7.439$, ddl = 1, $p < *0.01$) et néerlandophones ($\chi^2 = 12.517$, ddl = 1, $p < *0.001$) au suremploi des UP avec le verbe *donner* et un sous-emploi des UP avec le verbe *prendre*, par comparaison avec la fréquence des emplois en L1. Notons que la tendance au suremploi/ sous-emploi est légèrement plus marquée dans le cas des apprenants néerlandophones.

Comment expliquer ce *paradoxe* ? Il nous paraissait incontournable, malgré la limitation de notre champ d'étude à la description de données langagières, de mettre en avant quelques perspectives explicatives susceptibles de faire l'objet de recherches ultérieures. Nous proposons ainsi plusieurs explications liées à la spécificité du système linguistique de chacun de ces verbes à haute fréquence⁹¹ : (i) leur caractère plus ou moins polysémique ; (ii) une différence de système flexionnel (cf. point suivant) ; (iii) leur comportement plus ou moins phraséologique (cf. Chapitre VI sur la complexité). Ces caractéristiques pourraient expliquer (en partie) la différence de fréquence d'emploi entre les UP construites avec *prendre* et celles construites avec *donner*.

La première explication relève du champ d'investigation des processus de transfert qui interviennent notamment en situation de production d'une L2 (cf. Altenberg & Granger, 2001). Nous avons eu l'occasion d'aborder dans les pages précédentes la notion de transfert (cf. pp. 103-104, p. 108 et p. 130), considéré comme affectant principalement des formes linguistiques de la langue cible ayant un correspondant proche formellement et sémantiquement dans la langue source, et se caractérisant par une certaine prototypicalité (en anglais, *cognates*). L'idée de départ est que les verbes à haute fréquence sont, selon ces critères, d'excellents candidats au transfert.

A small number of typologically unmarked verbs called *nuclear verbs* tended to appear early and to be favoured in production both with respect to overall frequency and with respect to their tendency to be used in a wide range of semantic overextensions. (Viberg, 1998: 176)

En plus d'être très fréquents dans l'usage, les *verbes nucléaires* ont tendance à posséder de nombreux hyponymes (par ex. *offrir*, *livrer*, *prêter* comme verbes hyponymiques de *donner*), à être fortement polysémiques et à développer des sens grammaticaux (par ex. *étant donné que*) (Viberg, 2002b : 54). Ces verbes hautement polysémiques posséderaient néanmoins des emplois sémantiques légèrement différents entre langues, ce qui pourrait engendrer le suremploi ou le sous-emploi d'une certaine forme de la langue cible en

⁹¹ Ces explications ne sont certainement pas suffisantes pour expliquer un phénomène dans lequel interviennent une multitude de facteurs, tant internes qu'externes, mais le fait de les citer ici ouvre à nos yeux d'intéressantes perspectives pour des recherches ultérieures.

calquant (i.e. en transférant) le potentiel sémantique de la forme source sur celui de la forme cible. Selon cette hypothèse explicative, on pourrait donc légitimement penser que le verbe *prendre* serait plus polysémique que les verbes *take* en anglais et *nemen* en néerlandais, puisqu'il est sous-employé par les non-natifs, à l'inverse du verbe *donner* qui serait moins polysémique que les verbes *give* en anglais et *geven* en néerlandais, puisqu'il est suremployé par les non-natifs. Cette hypothèse reposant sur les mécanismes de transfert pourrait être testée par exemple en procédant à une analyse lexicographique des acceptions principales de *give/geven/donner* et *take/nemen/prendre*, couplée à l'analyse de corpus parallèles français-anglais⁹² et néerlandais/ français, en termes d'indice de traductibilité⁹³. Ce type d'étude contrastive n'a toutefois pas été mené dans le cadre de ce travail pour deux raisons. D'une part, l'examen, dans la langue source, du caractère plus ou moins polysémique des verbes nécessite une bonne connaissance préalable des langues concernées, ce qui n'est pas notre cas en tant que spécialiste de la langue française. D'autre part, ces analyses à visée explicative dépasseraient largement le cadre de ce travail qui s'est donné pour objectif la description systématique des unités phraséologiques (semi-figées) en français L1 et L2.

Nous proposons de développer dans le point qui suit la seconde de ces hypothèses, à savoir la nature du système flexionnel de chacun des verbes. Cette hypothèse explicative a été explorée ici, contrairement à la première hypothèse explicative émise plus haut, car elle touche au fonctionnement même de la langue française et ne requiert pas la mise en œuvre d'outils d'investigation spécifiques autres que ceux dont nous disposons déjà pour notre recherche.

La troisième hypothèse sera quant à elle abordée, bien que de manière indirecte, dans le point de notre Chapitre VI consacré à l'étude de la *complexité* des productions des apprenants du français L2. Nous supposons en effet que la *complexité* des SVN construites avec les verbes à haute fréquence *prendre* et *donner* résiderait dans le potentiel plus ou moins phraséologique de ces verbes. Cette *complexité* sera étudiée du point de vue de la complexité structurelle et de la diversité lexicale des deux verbes.

⁹² Il existe un corpus de bitextes anglais-français, nommé *BAF*, qui est disponible gratuitement en ligne. Ce corpus a été constitué par l'équipe de TAO (*Traduction Assistée par Ordinateur*) et par Michel Simard à l'université de Montréal (<http://rali.iro.umontreal.ca>).

⁹³ Nous renvoyons aux analyses contrastives menées par Viberg (1998, 2002a) et Källkvist (1998) pour plus de détails sur les méthodes utilisées pour mesurer cette différence d'emploi de verbes correspondants entre langues.

V.1.1.2. Perspective explicative : système flexionnel

Nous examinons dans ce point l'hypothèse de l'influence du système flexionnel plus ou moins irrégulier des verbes *prendre* vs *donner* sur les stratégies acquisitionnelles de l'apprenant (suremploi vs sous-emploi). Du point de vue flexionnel, nous pouvons ainsi supposer que la régularité de la conjugaison du verbe *donner* (en *-ER*) engendrerait un emploi plus aisé, voire un suremploi, de ce verbe par l'apprenant du français L2. Par opposition, l'irrégularité de la conjugaison du verbe *prendre* favoriserait la stratégie d'évitement et le sous-emploi des UP qui le contiennent. Pour vérifier cette hypothèse, nous avons étudié la fréquence d'occurrence de vingt verbes à haute fréquence⁹⁴, dont dix verbes en *_ER* et dix verbes en *_ENDRE*, dans les corpus d'étude et de contrôle. Nous formulons l'hypothèse que les verbes en *_ER* feraient l'objet d'un suremploi par les apprenants, tant anglophones que néerlandophones, par opposition aux verbes en *_ENDRE* qui seraient sous-employés.

Comme pour nos deux verbes à haute fréquence *prendre* et *donner*, nous avons extrait des corpus les formes de ces verbes via le concordancier *WordSmith Tools 4.0*. en entrant comme items de recherche une liste brute de leurs différentes réalisations incluant la forme infinitive, les formes conjuguées et les formes potentiellement déviantes (par ex. **prendu* pour *pris*) de ces verbes. Chacune des 20 concordances a fait l'objet d'une désambiguïisation syntaxique, en excluant les emplois nominaux, déverbaux, adverbiaux et prépositionnels de ces verbes, qui ne remplissaient pas une fonction syntaxique verbale à strictement parler. Nous avons ensuite additionné les occurrences de ces verbes en les classant en fonction de leur système flexionnel (*_ER* vs *_ENDRE*) et du corpus dans lequel elles apparaissaient. Les Tableaux (48) et (49) (ci-dessous) donnent un aperçu de la répartition de ces occurrences dans les trois corpus⁹⁵.

⁹⁴ Pour choisir ces verbes, nous nous sommes référée à une liste de fréquence de mots du français écrit contemporain (XIX^e - XX^e siècles), disponible en ligne à l'adresse [<http://www.eduscol.education.fr>]. Ces verbes sont les 10 verbes les plus fréquents pour chaque type de verbes.

⁹⁵ Nous avons calculé une fréquence relative des occurrences de nos corpus par 90.000 mots pour pouvoir comparer nos données et tirer des conclusions valables de l'examen de celles-ci. La fréquence notée entre crochets est quant à elle la fréquence absolue d'occurrence des verbes.

Verbes en _ER	FL2-AN	FL2-NL	FL1	Total
1. trouver	112,87 [232]	75	76,87 [158]	[465]
2. parler	77,35 [159]	33	62,76 [129]	[321]
3. passer	57,89 [119]	55	59,84 [123]	[297]
4. regarder	24,81 [51]	27	21,41 [44]	[122]
5. aimer	34,54 [71]	15	23,35 [48]	[134]
6. demander	49,62 [102]	60	49,62 [102]	[264]
7. rester	96,32 [198]	75	77,84 [160]	[433]
8. penser	84,16 [173]	65	139,14 [286]	[524]
9. arriver	48,16 [99]	21	35,51 [73]	[193]
10. entrer	30,65 [63]	22	30,65 [63]	[148]
Total	616,37 [1267]	448	576,99 [1186]	[2901]

Tableau 48. Distribution des 10 verbes en _ER les plus fréquents en L1, dans les corpus non natifs et natifs

Verbes en _ENDRE	FL2-AN	FL2-NL	FL1	Total
1. entendre	25,78 [53]	15	26,27 [54]	[122]
2. comprendre	32,11 [66]	47	64,22 [132]	[245]
3. rendre	64,7 [133]	85	57,41 [118]	[336]
4. attendre	25,78 [53]	13	12,16 [25]	[91]
5. reprendre	2,43 [5]	3	8,76 [18]	[26]
6. descendre	5,84 [12]	3	0,97 [2]	[17]
7. apprendre	27,24 [56]	38	47,68 [98]	[192]
8. tendre	1,46 [3]	1	13,14 [27]	[31]
9. vendre	2,43 [5]	12	7,78 [16]	[33]
10. étendre	4,87 [10]	4	7,3 [15]	[29]
Total	192,64 [396]	221	245,69 [505]	[1122]

Tableau 49. Distribution des 10 verbes en _ENDRE les plus fréquents en L1, dans les corpus non natifs et natifs

Ayant comparé le total des occurrences dans chaque catégorie flexionnelle en fonction des corpus, nous avons obtenu des résultats statistiques très significatifs ($\chi^2 = 16.146$, ddl = 2, $p < *0.001$) qui appuient (en partie) l'hypothèse de l'influence du système flexionnel des verbes à haute fréquence sur la fréquence d'emplois de ces verbes par les apprenants du français L2.

	FL2-AN	FL2-NL	FL1	Total
ER	616,37	448	576,99	1641,36
ENDRE	192,64	221	245,69	659,33
Total	809,01	669	822,68	2300,69

Tableau 50. Distribution des catégories flexionnelles en _ER et en _ENDRE dans les corpus non natifs et natifs

Les verbes en _ER, au même titre que le verbe *donner*, ont ainsi tendance à être suremployés de manière significative par les apprenants anglophones, alors que les verbes en _ENDRE tendent à être sous-employés par ceux-ci, comme c'est d'ailleurs le cas du verbe *prendre* ($\chi^2 = 7.606$, ddl = 1, $p < *0.01$). Par contraste, les résultats portant sur le corpus néerlandophone montrent une tendance non significative ($\chi^2 = 1.724$, ddl = 1, $p > 0.1$ (NS)). En effet, les apprenants néerlandophones utilisent les deux types de verbes, en _ER et en _ENDRE (avec respectivement 448 occ. et 221 occ.), dans une proportion similaire à celle des natifs, ce qui contredit notre hypothèse puisqu'il n'y a pas de

suremploi des verbes en _ER (448 occ. contre 577 occ. en L1). Notons toutefois que ces derniers types de verbes sont employés plus fréquemment que les verbes en _ENDRE.

Ces résultats confirment donc, bien que partiellement, l'hypothèse de l'influence du degré de complexité flexionnelle des verbes sur la tendance au sous-emploi par les apprenants, i.e. à l'évitement, des formes *a priori* plus difficiles à acquérir d'un point de vue flexionnel. En résumé, nous pourrions dire que les apprenants anglophones seraient plus sensibles à la complexité flexionnelle des verbes qui les pousserait à suremployer les verbes en _ER et à sous-employer les verbes en _ENDRE. Les apprenants néerlandophones seraient quant à eux également sensibles à cette complexité flexionnelle des verbes, puisqu'ils emploient plus de verbes en _ER qu'en _ENDRE, mais dans une moindre mesure, puisqu'ils tendent de manière générale à employer les verbes à haute fréquence dans une proportion similaire à celle des natifs. Rappelons que l'écart de fréquence, en termes de suremploi des UP construites avec le verbe *donner* et de sous-emploi des UP avec *prendre*, était pourtant davantage marqué chez les apprenants néerlandophones que chez les anglophones. D'autres facteurs⁹⁶, plus décisifs, semblent donc intervenir dans la production des UP construites avec le verbe *donner* par les locuteurs non natifs néerlandophones.

V.1.1.3. Mésusage

Nous avons examiné précédemment la proportion des déviations phraséologiques produites par l'apprenant avancé du français L2 (*vs* L1) en fonction des *types* de déviance, de la *localisation* des déviations et du *degré de gravité* de ces UP déviantes (cf. chapitre précédent, sous IV.3.). Il s'agira ici de déterminer la proportion des déviations en L2 portant sur les UP avec *prendre* par opposition aux UP avec *donner*.

Tout d'abord, selon l'hypothèse formulée à partir des résultats obtenus pour l'anglais L2 (Nesselhauf, 2004) (cf. l'Hypothèse 8, ci-dessous), nous nous attendions à une proportion de 25% d'UP déviantes (*vs* non déviantes) construites avec le verbe *prendre* et 25% d'UP déviantes construites avec *donner*.

Hypothèse 8

Les UP verbo-nominales construites avec les verbes *prendre* et *donner* auraient un comportement en partie comparable aux UP construites avec les deux verbes *take* et *give* en anglais. Selon ce point de vue, ¼ des UP construites avec *prendre* et *donner* seraient déviantes.

⁹⁶ Notre étude ne permet malheureusement pas de mettre le doigt sur ces facteurs « cachés » et c'est donc dans un état de curiosité inassouvie que nous laissons ici le lecteur.

Or, nous avons remarqué que la proportion de déviations portant sur *donner* correspondait aux chiffres attendus, avec 26,5% d'UP déviantes, alors que les déviations sur le verbe *prendre* représentaient 17% des UP produites avec ce verbe. Ces résultats confirment donc, tout en la nuanciant, l'hypothèse formulée ci-dessus.

	Prendre	Donner	Total
UP déviante	67 45,0% 17,0%	82 55,0% 26,5%	149 100,0% 21,2%
UP non déviante	326 58,8% 83,0%	228 41,2% 73,5%	554 100,0% 78,8%
Total	393 55,9% 100,0%	310 44,1% 100,0%	703 100,0% 100,0%

Tableau 51. Distribution des UP construites avec les verbes *prendre* vs *donner*, en fonction de leur statut déviant ou non déviant

Ensuite, selon la répartition proposée dans le tableau ci-dessous (cf. Tableau 52), nous avons observé que la proportion d'UP déviantes chez le locuteur natif (LN) était identique pour *prendre* (Ex. 122) et *donner* (Ex. 123) (10 occ. pour chaque verbe).

(122) De l' autre côté, il y a les autres : Ils sont différents parcequ' ils possèdent une autre culture, avec une religion et des coutumes particulières. Pourquoi une culture autre, qui devrait être ?|**prise**| comme un atout serait-elle une barrière sociale ? (FL1, <TFRB178>, occ. 719)

→ Reconstruction native : *Pourquoi une culture autre, qui devrait être **perçue/ considérée** comme un atout, serait-elle une barrière sociale ?*

(123) Si tout le monde imite une seule personne, il faut donc faire tout ce que celle-ci dit, comme elle le dit. lonesco nous **donne** un ?|**merveilleux**| exemple d'un fait similaire. Dans sa pièce intitulée « Rhinocéros », il présente la montée du nazisme et de ses idées auxquelles tout le monde adhère. (FL1, <TFRB093>, occ. 845)

→ Reconstruction native : *lonesco nous **donne** un bel exemple d'un fait similaire*

- L'indice d'association lexicale du cooccurrent syntaxique *donner* + *exemple* (IM = 4,572) est supérieur au seuil de collocabilité fixé à une IM ≥ 3 . Nous avons également noté la présence, dans la liste du LADL, de l'expression correspondante *donner le exemple de « N »* à et constaté que cette SVN était attestée dans l'usage L1 (FqR *Le Soir* = 0,97 et FqR dans *Valibel* = 63,4). Nous avons donc considéré cette SVN comme étant phraséologique.
- Néanmoins, l'UP telle qu'elle est produite dans le corpus étudié, i.e. avec le modifieur adjectival *merveilleux*, ne renvoie à aucun cooccurrent syntaxique *donner* + *merveilleux exemple* dans *Les Voisins De Le Monde*, alors que le couple concurrent *donner* + *beau exemple* montre un indice d'association lexicale élevé (IM = 5,274).
- Par le biais d'un examen de la fréquence de ces deux couples de cooccurrents sur *Google.fr*, l'un avec l'adjectif *merveilleux* et l'autre avec l'adjectif *bel*, nous n'avons trouvé que 2 p. contenant l'expression « donne un merveilleux exemple de/d' », pour 1.104 p. contenant l'expression « donne un bel exemple de/d' ». L'UP construite avec l'adjectif *merveilleux* semble pour sa part préférer les constructions avec le verbe *être*, puisque nous avons trouvé 161 p. contenant l'expression « est/sont/être un merveilleux exemple de ». Il semblerait donc que dans le contexte de l'occurrence étudiée (avec *donner*), la construction avec l'adjectif *bel* soit plus appropriée pour exprimer un sens 'illustratif' que celle avec l'adjectif *merveilleux*.

La proportion de déviations phraséologiques chez le locuteur non natif (LNN) est quant à elle un peu plus élevée avec *donner* (Ex. 124 et 125) (52,6% chez les néerlandophones et 57,1% chez les anglophones) qu'avec *prendre* (Ex. 126 et 95) (47,4% chez les néerlandophones et 42,9% chez les anglophones).

Exemples d'UP déviantes avec le verbe *donner* en français L2 :

(124) Anne était très heureuse. Elle l'invita de passer la nuit chez nous. Je *|**donnai**| un coup d'oeil à ma amie. (FL2-AN, <TAN081>, occ. 344)

→ Reconstruction native : Je **jetai** un coup d'oeil à ma amie.

(125) Ce que Gide appelle 'journalisme' ne peut pas entraîner le rejet des oeuvres contemporaines. Mais cela doit nous inciter à revoir notre comportement envers les 'classiques' : La littérature - par exemple - est plus que ces classiques et les nouvelles évolutions devraient en tous cas être mentionnées ! Parce qu'il est vrai que *|**donner**| un jugement concernant la valeur journalière n'est pas toujours facile. (occ. 1070)

→ Reconstruction native : Parce qu'il est vrai que **porter** un jugement concernant la valeur journalière n'est pas toujours facile.

Exemples d'UP déviantes avec le verbe *prendre* en français L2 :

(126) Votre santé est ni un sujet qui devrait être éviter, ni pas un sujet qui peut ?|**prendre** la deuxième place|. (FL2-AN, <TAN106>, occ. 118)

→ Reconstruction native : Votre santé est ni un sujet qui devrait être éviter, ni pas un sujet qui peut **passer/être (relégué/ laissé) au second plan**

(95) Lui aussi commet des crimes sans s'en être conscient, mais il finit par **prendre** *|sa responsabilité| et il s'arrache les yeux. (FL2-NL, <TNL118>, occ. 958)

→ Reconstruction native : mais il finit par **prendre ses responsabilités** et il s'arrache les yeux

Mais cette légère différence au sein de la variété avancée du français L2 est-elle pour autant statistiquement significative ?

	FL2-AN	FL2-NL	FL1	Total
Prendre	18,97 [39] 42,9%	18 47,4%	4,86 [10] 50,0%	[67] 45,0%
Donner	25,3 [52] 57,1%	20 52,6%	4,86 [10] 50,0%	[82] 55,0%
Total	44,27 [91] 100,0%	38 100,0%	9,73 [20] 100,0%	[149] 100,0%

Tableau 52. Distribution des UP déviantes construites avec les verbes *prendre* et *donner* dans les corpus non natifs et natif

Comme l'indiquent nos résultats statistiques, cette légère différence au sein de la variété avancée du français L2 n'est en fait pas significative (χ^2 de Yates = 0.046, ddl = 2, $p > 0.5$ (NS)): la nature du verbe à haute fréquence utilisé par l'apprenant du français L2 ne semble pas avoir d'influence sur le nombre d'UP déviantes produites au sein des corpus. Nous pouvons dès lors considérer que les UP déviantes montrent une proportion relativement équivalente de constructions avec *prendre* et *donner* dans nos trois corpus respectifs.

Enfin, nous avons examiné le rapport entre chacun de ces deux verbes et la nature des déviances produites.

Premièrement, nous nous attarderons sur la description des déviances en *types* de déviances. D'un point de vue général, considérant l'ensemble de nos trois corpus, il ne semble pas y avoir d'influence de la nature du verbe utilisé sur le type de déviances produites ($\chi^2 = 3.615$, ddl = 3, $p > 0.3$ (NS)). Nous avons ainsi observé plus ou moins la

même proportion de cas déviants construits avec *prendre* et *donner* concernant les déviations formelles et contextuelles. Seules les déviations quantitatives portent sur une proportion plus élevée d'UP construites avec le verbe *donner* (25 occ.) (Ex. 127) que celles construites avec *prendre* (14 occ.) (Ex. 128).

(127) Le monde romanesque se perd. On peut y insister |en **donnant** l'exemple de| Montaigne.
(FL2-NL, <TNL067>, occ. 1067)

→ Reconstruction native : On peut y insister en **prenant** l'exemple de Montaigne.

(128) Le plan national de passage à l'Euro regroupe toutes les décisions et recommandations |**prises**|, au niveau national, pour assurer dans les meilleures conditions de la monnaie unique en chaque pays (FL2-NL, <TAN142>, occ. 231)

→ Reconstruction native : Le plan national de passage à l'Euro regroupe toutes les décisions prises et les recommandations faites

Rappelons que cette légère différence n'est cependant pas significative comparée aux autres types de déviance.

	Prendre	Donner	Total
Déviance formelle	33	30	63
Déviance contextuelle	17	20	37
Déviance quantitative	14	25	39
Déviance mixte	3	7	10
Total	67	82	149

Tableau 53. Distribution des UP déviante construites avec prendre vs donner, en fonction du type de déviance

Du point de vue des corpus d'apprenants du français L2 pris séparément, aucune différence significative n'a été observée concernant la relation entre les types de déviance et les verbes utilisés. Retenons simplement que les apprenants anglophones produisent une même proportion d'UP déviantes au sein de chaque catégorie de déviance pour les deux verbes, si ce n'est la catégorie des déviations quantitatives qui compte deux fois plus d'UP déviantes construites avec le verbe *donner* (16 occ.) qu'avec le verbe *prendre* (8 occ.).

Deuxièmement, nous examinerons la répartition des UP déviantes construites avec *prendre* vs *donner* du point de vue de la *localisation* des déviations au sein de l'UP.⁹⁷ Il ressort de nos observations quelques différences de répartition entre les deux verbes. Concernant les déviations sur le nom et le déterminant, celles-ci portent en majorité sur des UP construites avec *donner* : on dénombre 12 déviations avec *donner* portant sur le constituant nominal (Ex. 129), contre 7 avec *prendre* (Ex. 130), et 5 déviations avec *donner* portant sur le déterminant (Ex. 131), contre une seule déviance avec *prendre* (Ex. 132).

⁹⁷ Les occurrences peu nombreuses ne nous ont pas permis de procéder ici à des tests statistiques valables (la fréquence attendue est inférieure à 5 dans plus de 20% des cases de la table de contingence).

Exemples d'UP contenant une déviance sur le constituant nominal :

(129) Biensûr, cet héritage a évolué en moi et je me suis imprégnée également d' autres valeurs issues de mon expérience de vie. Mais ma famille a joué le rôle majeur de **donner** ?[à ma construction] des ?[piliers] solides. (FL1, <TFRB252>, occ. 895)

→ Reconstruction native : *Mais ma famille a joué le rôle majeur de **donner** des bases solides à/ pour mon épanouissement/ développement personnel*

(130) Donc les pauvres et la classe ouvrière a pris des *[classes] très élémentaires dans les écoles des catholiques (FL2-AN, <TAN238>, occ. 203)

→ Reconstruction native : *Donc les pauvres et la classe ouvrière a pris **a pris** des cours très élémentaires dans les écoles des catholiques*

Exemples d'UP contenant une déviance sur le déterminant :

(131) C'était la même année que le gouvernement québécois a voté le loi 63, qui **donnait** *[un] garanti que les parents pouvaient dans quelle langue leurs enfants étaient enseignés. (FL2-AN, <TAN224>, occ. 451)

→ Reconstruction native : *la loi 63 (...) **donnait** la garantie que les parents pouvaient dans quelle langue leurs enfants étaient enseignés.*

(132) Car les adultes, eux non plus, ne font le sport pour le sport. " Le sport est l'art par lequel l'homme se libère de soi-même " a dit Giraud. **Prise** dans ?[le] sens négatif, il n'y a rien à redire à cette phrase. (FL2-NL, <TNL107>, occ. 954)

→ Reconstruction native : ***Prise** dans un/son sens négatif, il n'y a rien à redire à cette phrase*

Ces déviances, qui portent sur le nom ou le déterminant d'UP construites avec *donner*, sont surtout le fait d'apprenants anglophones, qui produisent respectivement 9 déviances sur le nom et 4 déviances sur le déterminant. On compte néanmoins 6 UP déviantes portant sur le nom dans le corpus néerlandophone, mais qui présentent, à l'inverse du corpus anglophone, une majorité d'UP déviantes portant sur des UP construites avec le verbe *prendre* (5 occ. sur 6 occ.) (Ex. 133).

Exemple d'UP avec le verbe *prendre*, contenant une déviance sur le nom (corpus néerlandophone):

(133) Dans le cas d'une erreur judiciaire enfin, les conséquences sont catastrophiques. Espérons que tous ceux qui favorisent avec la peine de mort **prennent** un *[moment] de réflexion et se rendent compte du caractère inhumain de cette décision. (FL2-NL, <TNL152>, occ. 1016)

→ Reconstruction native : ***prennent** un temps de réflexion*

Concernant les déviances portant sur les particules prépositionnelles ou conjonctives, celles-ci portent par contre majoritairement sur des constructions avec le verbe *prendre*, pour lesquelles nous comptons 8 déviances (Ex. 134), contre 3 seulement avec *donner* qui sont par ailleurs toutes les trois identiques (sélection déviante de la conjonction *pour* au lieu de particule *de*) et sont produites par le même apprenant (Ex. 135, 136 et 137).

Exemple d'UP avec le verbe *prendre*, contenant une déviance sur la particule prépositionnelle/ conjonctive:

(134) Et maintenant nous veulent une seule personne qui solve ce problème et **prendre** les autres problèmes de la communauté ?[pour] lui seulement ? (FL2-AN, <TAN003>, occ. 55)

→ Reconstruction native : ***prendre** les autres problèmes de la communauté sur lui seulement ?*

UP avec le verbe *donner*, qui contiennent une déviance sur la particule prépositionnelle/ conjonctive:

(135) Si vous nous **donnez** l'occasion *|pour| l'améliorer, nous le ferons. (<TAN104>, occ. 382)

→ Reconstruction native : *Si vous nous **donnez** l'occasion de l'améliorer, nous le ferons.*

(136) Comme résultat, vous avez besoin de nous **donner** le temps et donc l'occasion *|pour| le rectifier. (<TAN104>, occ. 383)

→ Reconstruction native : *vous avez besoin de nous **donner** le temps (...) de le rectifier*

(137) Comme résultat, vous avez besoin de nous **donner** le temps et donc l'occasion *|pour| le rectifier. (<TAN104>, occ. 384)

→ Reconstruction native : *Par conséquent, vous devez nous **donner** (...) l'occasion de le rectifier.*

Concernant les déviances portant sur le verbe, les apprenants anglophones produisent autant d'UP déviantes avec *prendre* (14 occ.) (Ex. 138) qu'avec *donner* (12 occ.) (Ex. 139), alors que les néerlandophones produisent 10 UP déviantes avec le verbe *donner* (Ex. 140) et 5 UP déviantes avec *prendre* (Ex. 81).

Exemple d'UP avec le verbe *prendre*, contenant une déviance sur le verbe (corpus anglophone):

(138) Donc les aristocrates, les riches garderaient leur puissance et leur statut en empêchant les autres classes d'avoir les connaissances et l'éducation requis pour *|prendre| les hauts postes de la société. (FL2-AN, <TAN165>, occ. 132)

→ Reconstruction native : *(...) en empêchant les autres classes d'avoir les connaissances et l'éducation requis pour **occuper** les hauts postes de la société.*

Exemple d'UP avec le verbe *donner*, contenant une déviance sur le verbe (corpus anglophone):

(139) La société nous *|donne| des rôles traditionnels pour l'homme et pour la femme. (FL2-AN, <TAN130>, occ. 397)

→ Reconstruction native : *La société **assigne** des rôles traditionnels à l'homme et à la femme*

Exemple d'UP avec le verbe *donner*, contenant une déviance sur le verbe (corpus néerlandophone):

(140) Même si les méthodes d'exécution sont plus mécanisées et même plus humaines aujourd'hui qu'auparavant, il y a toujours quelqu'un qui doit appuyer sur le bouton ou *|donner| l'injection létale, ce qui entraîne sans aucun doute un sentiment de culpabilité et de responsabilité. (FL2-NL, <TNL171>, occ. 1194)

→ Reconstruction native : *il y a toujours quelqu'un qui doit appuyer sur le bouton ou **pratiquer/ procéder à** faire l'injection létale*

Exemple d'UP avec le verbe *prendre*, contenant une déviance sur le verbe (corpus néerlandophone):

(81) Quand un enquêteur veut réaliser une enquête, il a besoin d'une audience pour qu'il veuille rassembler des données des individus. On utilise par exemple, les enquêtes pour faire des conclusions sur l'opinion du public. On *|prend| un sondage et tous les gens qui sont dedans doivent répondre à un questionnaire. (FL2-NL, <TNL020>, occ. 924)

→ Reconstruction native : *On **réalise/ effectue/ fait** un sondage*

Troisièmement, un test de rang (test de Mann-Whitney) comparant la nature du verbe à haute fréquence (variable nominale) et le degré de gravité des UP déviantes (variable ordinale) a montré qu'il n'y avait pas d'influence de l'une de ces variables sur l'autre. Ces résultats, qui traduisent un degré de gravité des déviances légèrement supérieur pour les UP construites avec *prendre* (rang moyen = 75.84) par contraste aux UP avec *donner* (rang moyen = 74.32), ne sont cependant pas significatifs ($U = 2691.000$; $p > 0.5$ (NS)). Il semblerait donc qu'il n'y ait pas d'influence de la nature du verbe utilisé par les apprenants sur le degré de gravité plus ou moins important des UP déviantes.

V.1.1.4. Synthèse des résultats

Dans ce point, nous avons étudié la fréquence d'emplois des UP en comparant le comportement phraséologique des deux verbes *prendre* vs *donner*. Nous avons ainsi observé que les apprenants du français L2 suremployaient les UP construites avec le verbe *donner* et sous-employaient les UP construites avec le verbe *prendre* [Hypothèse 2 vérifiée]. Suite à ce constat, nous avons suggéré quelques hypothèses explicatives et vérifié partiellement l'une d'entre elles, en l'occurrence l'influence de la complexité flexionnelle des verbes irréguliers en _ENDRE par opposition aux verbes à conjugaison régulière en _ER, sur la fréquence d'emplois des UP construites avec des verbes à haute fréquence. Comparant ensuite ces deux verbes quand ils portaient sur des UP déviantes, nous avons noté que les UP étaient déviantes dans 26,5% des cas quand elles étaient construites avec le verbe *prendre* et dans 17% des cas quand elles étaient construites avec *donner* [Hypothèse 8 partiellement vérifiée]. Pour terminer, à l'instar de notre examen des UP déviantes en fonction de leur répartition dans les trois corpus, nous avons mis en exergue quelques différences de comportement des deux verbes par rapport à la nature des déviances en jeu dans les UP, du point de vue (i) des *types* de déviances en cause (formelles, contextuelles et/ou quantitatives), (ii) de la *localisation* des déviances au sein des UP et (iii) de leur *degré de gravité*.

Premièrement, aucune influence de la nature du verbe sur le type de déviance n'a été démontrée, mais nous avons néanmoins observé que la catégorie des déviances quantitatives comptait deux fois plus d'UP déviantes avec le verbe *donner* qu'avec le verbe *prendre*.

Deuxièmement, concernant les catégories grammaticales sur lesquelles portent les déviances, nous avons observé qu'une majorité des déviances construites avec *donner* portait sur le nom ou le déterminant et étaient principalement produites par les apprenants anglophones. Les UP construites avec le verbe *prendre*, pour lesquelles la déviance porte sur le nom, sont quant à elles majoritairement le fait d'apprenants néerlandophones. Ce sont aussi les UP avec *prendre* qui comptent le plus de déviances portant sur les particules prépositives ou conjonctives. Les UP déviantes portant sur le constituant verbal concernent de manière égale les UP avec *prendre* (21 occ.) et les UP avec *donner* (25 occ.), mais les néerlandophones produisent néanmoins 2/3 des déviances verbales avec le verbe *donner* pour seulement 1/3 avec *prendre*.

Troisièmement, nous avons constaté qu'il n'y avait pas de corrélation significative entre la nature du verbe utilisé par les apprenants et le degré de gravité plus ou moins important des déviations.

Ayant réussi à faire émerger une image plus claire du comportement de ces deux verbes dans leurs emplois phraséologiques respectifs, nous nous proposons maintenant de dresser un profil syntaxique des syntagmes nominaux constitutifs des UP en nous centrant cette fois, non plus sur leur constituant verbal (*prendre vs donner*), mais sur leur constituant nominal du point de vue de leur structure syntaxique.

V.1.2. Profil syntaxique des syntagmes nominaux des UP

La variable que nous étudions dans ce point, nommée *SyntaxSNUP*, étudie la structure syntaxique des syntagmes nominaux (SN) postverbaux constitutifs des UP. Nous avons attribué différentes valeurs à cette variable qui rendent compte des structures effectivement observées dans les corpus de français L2 et L1.

Nom de la variable (Name)	Variable (Label)	Valeurs (Values)
SyntaxSNUP	Structure syntaxique du SN postverbal de l'UP	1. à 3. Dét zéro + N + (Modifieur/ <i>de</i> N/ P Sub) 4. Dét + <i>N de</i> N 5. Dét + N + P Sub 6. Dét Poss + N + (Modifieur) 7. Prép + SN 8. Dét + N + (Modifieur)

Tableau 54. Extrait de la liste paramétrique – Variable formelle : structure syntaxique des SN postverbaux

Nous illustrons ci-dessous quelques-unes des valeurs attribuées à nos données de corpus. Ainsi, on attribuera à l'exemple (82) la valeur 4 pour la variable *SyntaxSNUP*, puisque le syntagme nominal s'y caractérise par une structure en *N de N*.

(141) **Prenons** de nouveau le cas de "Kinépolis"⁹⁸ (FL2-NL, <TNL057>, occ. 935)

Dans le cas de structures nominales prépositionnelles, les SVN ont été considérées comme telles quand le figement s'exerçait sur le syntagme nominal prépositionnel, considéré dès lors comme un complément essentiel⁹⁹ du verbe. C'est notamment le cas de l'exemple (103), comprenant le syntagme prépositionnel *dans l'intrigue*, auquel nous avons attribué la valeur 8.

⁹⁸ Rappelons encore que les caractères gras et les soulèvements dans les exemples sont de nous. Le caractère gras marque le verbe, i.e. la base de la SVN, le soulèvement simple concerne ses compléments essentiels sans préposition et le double soulèvement ses compléments essentiels prépositionnels.

⁹⁹ Comme nous l'avons déjà rappelé dans la note 24 (p. 46), les compléments sont dits *essentiels* quand ils font partie de la structure valencielle du verbe, par opposition aux compléments dits *non essentiels* ou *circonstanciels*.

(103) Au fur et à mesure de cette histoire, le côté irréel va prendre de plus en plus de place, mais cela n'a aucune importance : le lecteur est pris dans l'intrigue et tout peut arriver (FL1, <TFRB265>, occ.770)

Nous donnons ci-dessous (cf. Tableau 55) un aperçu de la répartition des types de structures syntaxiques dans les corpus, concernant l'ensemble de nos données phraséologiques. Plusieurs observations ressortent de ce tableau. Une tendance apparaît en ce qui concerne les structures sans déterminant (i.e. avec déterminant « zéro ») (valeurs 1 à 3 dans la variable *SyntaxSNUP*) et celles avec un déterminant quelconque (valeur 8 dans la variable *SyntaxSNUP*). La fréquence des structures sans déterminant est supérieure dans le corpus natif (18,2%), nettement inférieure dans le corpus néerlandophone (11,7%) et dans le corpus anglophone (8,9%). A l'inverse, la fréquence d'emploi des structures avec un déterminant quelconque est moindre dans le corpus natif (36,1%), supérieure dans le corpus néerlandophone (47,4%) et nettement supérieure dans le corpus anglophone (54,4%). On pourrait dès lors suggérer qu'il y aurait un sous-emploi des structures sans déterminant (Ex. 142) et un suremploi des structures avec déterminant (autre que zéro ou possessif) par les apprenants du français L2 (Ex. 38).

(142) je veux dire qu'il a **prit** connaissance de ses émotions et peut donc les contrôler (FL2-AN, <TAN166>, occ. 133)

(38) D'autres explications sont également **données** dans cet article et inclus " un déficit d'information, un facteur démographique et la vantardise masculine " (FL2-AN, <TAN160>, occ. 541)

Cette tendance au suremploi est également observable pour les structures avec une proposition subordonnée (Ex. 143) (valeur 5 dans la variable *SyntaxSNUP*), alors que nous observons de nouveau une tendance inverse au sous-emploi pour les syntagmes nominaux prépositionnels (Ex. 144) (valeur 7 dans la variable *SyntaxSNUP*).

(143) Le garanti de pouvoirs accrus dans la matière de l'immigration **donnerait** au Québec le pouvoir de contrôler le flux d'immigration et de sanctionner les immigrants qui sont déjà installés au Canada. (FL2-AN, <TAN223>, occ. 444)

(144) De cette façon, au lieu de créer un climat de violence, on crée une société dans laquelle on **prend en considération** la psychologie humaine. (FL2-NL, <TNL217>, occ. 1039)

Dans le cas des UP prépositionnelles, ce sont toutefois les apprenants néerlandophones qui montrent la fréquence d'emplois la plus faible (5,8%) par rapport aux natifs (15,9%).

	FL2-AN	FL2-NL	FL1	Total
Dét zéro + N + (Modifieur/ <i>de N/ P Sub</i>)	11,68 [24] (-) 8,9%	16 (-) 11,7%	26,27 [54] 18,2%	53,95 [94]
Dét + N <i>de N</i>	12,65 [26] (-) 9,6%	25 18,2%	23,35 [48] 16,2%	61 [99]
Dét + N + P Sub	18,97 [39] (+) 14,4%	16 (+) 11,7%	12,16 [25] 8,4%	47,14 [80]
Dét Poss + N + (Modifieur)	5,84 [12] 4,4%	7 5,1%	7,3 [15] 5,1%	20,14 [34]
Prép + SN	10,7 [22] (-) 8,1%	8 (-) 5,8%	22,86 [47] 15,9%	41,57 [77]
Dét + N + (Modifieur)	71,51 [147] (+) 54,4%	65 (+) 47,4%	52,05 [107] 36,1%	188,57 [319]
Total	131,35 [270] 100,0%	137 100,0%	144 [296] 100,0%	412,35 [703]

Tableau 55. Distribution des structures syntaxiques du SN des UP dans les corpus non natifs et natifs

De manière générale, la différence entre les locuteurs non natifs (LNN) anglophones et les locuteurs natifs (LN) dans leur manière de structurer le SN des UP s'est avérée très significative ($\chi^2 = 35.713$, ddl = 5, $p = *0.000$), et la différence entre les LNN néerlandophones et les LN significative ($\chi^2 = 11.503$, ddl = 5, $p < *0.05$). La tendance au suremploi et au sous-emploi de certaines structures syntaxiques par les LNN anglophones et néerlandophones a donc été confirmée par ces résultats, avec une différence moindre pour les LNN néerlandophones qui semblent plus proches des LN que les anglophones dans leur structuration syntaxique des syntagmes nominaux au sein des UP.

De manière plus spécifique, nous examinons ci-dessous quel est le profil syntaxique des syntagmes nominaux dans les UP construites respectivement avec les verbes *prendre* vs *donner*.

L'usage natif montre une différence très significative ($\chi^2 = 40,750$, ddl = 5, $p = *0.000$) entre les UP avec *prendre* et celles avec *donner*, en fonction de la structure syntaxique du SN. Selon la répartition présentée dans le Tableau (56) ci-dessous, les UP construites avec *prendre* préfèrent les constructions moins marquées, i.e. avec un déterminant et/ou un modifieur quelconques, dans 1/3 des cas (34,2%) (Ex. 145).

(145) Sous cette perspective, l'union européenne **prend** une tout autre dimension et rappelle furieusement les thèses célèbres de Darwin: seules les espèces les mieux adaptées à leur environnement survivent, les autres sont condamnées à l'extinction. (FL1, <TFRB300>, occ. 590)

Elles montrent ensuite une préférence pour les constructions prépositionnelles dans un peu plus d'1/5 des cas (21,9%) (Ex. 146), pour les constructions en *N de N* (18,9%) (Ex. 147) et les structures avec un déterminant « zéro » (17,9%) (Ex. 148) dans un peu moins d'1/5 des cas.

(146) Si on prend au pied de la lettre la phrase, on pourrait penser que celle ci véhiculerai une valeur communiste, le fait d' être tous de la même classe social, pas de patron. Il faut prendre le « ressembler », dans un sens plus large. (FL1, <TFRB046>, occ. 642)

(147) Ainsi, ce n'est pas le premier ilote venu qui aurait été capable d'écrire un scénario tel que celui du « Le fabuleux destin d'Amélie Poulain ». En effet, prenons l'exemple de ce film : Il regroupe à lui tout seul une grande partie des genres littéraires, la poésie, le théâtre de par sa mise en scène, la philosophie... ce qui transmet une émotion unique aux spectateurs. (FL1, <TFRB164>, occ. 715)

(148) La chance que nous possédons de pouvoir exprimer, plus ou moins, librement notre identité, doit nous faire prendre conscience de notre impact personnel et collectif sur la vie d'une société que nous représentons. (FL1, <TFRB186>, occ. 724)

Les UP construites avec le verbe *donner* présentent quant à elles principalement, comme pour le verbe *prendre*, des SN avec un déterminant et un modifieur quelconques (40% des cas) (Ex. 149).

(149) L'Europe de 92 implique donc un avenir enrichissant, dont l'objectif sera de donner une dynamique nouvelle à la Communauté. (FL1, <TFRB295>, occ. 788)

Elles montrent ensuite une préférence, dans 1/5 des cas, pour des SN construits avec une proposition subordonnée (20%) (Ex. 150) ou avec un déterminant « zéro » (19%) (Ex. 151).

(150) Beaucoup de gens se ressemblent fortement que ce soit au niveau vestimentaire ainsi que dans leur manière d'agir. Ils nous donnent l'impression d'avoir été formé dans le même moule. (FL1, <TFRB130>, occ. 854)

(151) Toutefois, certains esprits auront tendance à prétendre que ces adaptations dénaturent la saveur initiale des originaux et, malgré que je ne puisse leur donner tort à l'occasion, les transpositions se font souvent pour le mieux. (FL1, <TFRB033>, occ. 832)

	Prendre	Donner	Total
Dét zéro + N + (Modifieur/ de N/ P Sub)	35 64,8% 17,9%	19 35,2% 19,0%	54 100,0% 18,2%
Dét + N de N	37 77,1% 18,9%	11 22,9% 11,0%	48 100,0% 16,2%
Dét + N + P Sub	5 20,0% 2,6%	20 80,0% 20,0%	25 100,0% 8,4%
Dét Poss + N + (Modifieur)	9 60,0% 4,6%	6 40,0% 6,0%	15 100,0% 5,1%
Prép + SN	43 91,5% 21,9%	4 8,5% 4,0%	47 100,0% 15,9%
Dét + N + (Modifieur)	67 62,6% 34,2%	40 37,4% 40,0%	107 100,0% 36,1%
Total	196 66,2% 100,0%	100 33,8% 100,0%	296 100,0% 100,0%

Tableau 56. Profil syntaxique des syntagmes nominaux dans les UP construites avec prendre vs donner en français L1

Nous soulignons également le fait que les structures prépositionnelles apparaissent dans 91,5% des cas, ainsi que les structures en *N de N* dans 77,1% des cas, dans des UP avec le verbe *prendre*. A l'inverse, les structures avec une proposition subordonnée complément du nom apparaissent essentiellement dans les UP incluant le verbe *donner* (80% des cas). Suite à ces observations, nous pouvons donc conclure que le profil syntaxique des SN dans les UP dans l'usage natif est très différent selon que l'on utilise le verbe *prendre* ou *donner*.

Mais que font les apprenants du français L2 ? Prennent-ils en considération cette différence de comportement entre les deux verbes dans leurs emplois phraséologiques ? C'est ce que nous allons tenter de déterminer maintenant. Observons successivement la répartition des UP avec *prendre* vs *donner* dans chacun des corpus non natifs.

L'usage des LNN anglophones ne semble pas refléter l'usage que les LN font des UP construites avec le verbe *prendre* : la différence entre les UP produites par les LNN anglophones et les LN est en effet significative du point de vue de la structuration des SN dans les UP construites avec le verbe *prendre* ($\chi^2 = 24.123$, ddl = 5, $p < *0.001$). Par contre, cette différence ne s'est pas avérée significative dans le cas des UP construites avec *donner* (χ^2 de Yates = 9.653, ddl = 5, $p > 0.05$ (NS)). L'usage des UP avec *donner*, du point de vue de leur profil syntaxique, semble donc relativement similaire à celui des locuteurs natifs. Nous avons noté, suite à ce constat, quelques observations qui émanent du tableau présenté ci-dessous (cf. Tableau 57), parmi lesquelles quelques tendances au suremploi (marquées par un + devant les %) ou au sous-emploi (marquées par un - devant les %).

	Prendre		Donner	
	FL2-AN	FL1	FL2-AN	FL1
Dét zéro + N + (Modifieur/ de N/ P Sub)	12 (-) 8,9%	35 17,9%	12 (-) 8,9%	19 19,0%
Dét + N de N	13 (-) 9,6%	37 18,9%	13 9,6%	11 11,0%
Dét + N + P Sub	12 (+) 8,9%	5 2,6%	27 20,0%	20 20,0%
Dét Poss + N + (Modifieur)	4 3,0%	9 4,6%	8 5,9%	6 6,0%
Prép + SN	22 (-) 16,3%	43 21,9%	0 (-) 0%	4 4,0%
Dét + N + (Modifieur)	72 (+) 53,3%	67 34,2%	75 (+) 55,6%	40 40,0%
Total	135 100,0%	196 100,0%	135 100,0%	100 100,0%
Chi carré	$\chi^2 = 24.123$, ddl = 5, p < *0.001		χ^2 de Yates = 9.653, ddl = 5, p > 0.05 (NS)	

Tableau 57. Distribution des structures syntaxiques du SN des UP dans le corpus anglophone (vs natif) en fonction du verbe utilisé

Examinons tout d'abord le cas des UP construites avec le verbe *prendre*. Les tendances observées concernent, d'une part, le suremploi des UP incluant le verbe *prendre* quand elles sont construites de manière relativement neutre, i.e. avec un déterminant et/ou un modifieur quelconque(s) (Ex. 152) (53,3% pour 34,2% d'UP natives), ou avec une proposition subordonnée complément du nom (Ex. 153) (8,9% pour 2,6% d'UP natives) et, d'autre part, le sous-emploi de ces UP quand elles présentent une structure prépositionnelle (Ex. 154) (16,3% pour 21,9% d'UP natives), en *N de N* (Ex. 155) (9,6% pour 18,9% d'UP natives) ou sans déterminant (Ex. 25) (8,9% pour 17,9% d'UP natives).

(152) nous sommes retournés au Pavillon de Naro Moru, où nous **avons pris** le meilleur dîner qu'on ait jamais vu.(FL2-AN, <TAN050>, occ. 91)

(153) En juin 1990 ceux qui l'ont déjà ratifié, ont **pris** la décision que l'accord ne pouvait pas être encore changé. (FL2-AN, <TAN230>, occ. 184)

(154) Je vous dit - vous avez maintenant la chance de **prenez en charge** votre vie (FL2-AN, <TAN101>, occ. 116)

(155) Nous **avons pris** la route de Naro Moru qui se trouve au sud de la montagne. (FL2-AN, <TAN050>, occ. 89)

(25) la contestation de la gauche **a pris** vigueur avec la formation du Parti Populaire Martiniquais et les Partis Communistes en Martinique et Guadeloupe. (FL2-AN, <TAN221>, occ. 169)

Ensuite, les résultats concernant les UP construites avec le verbe *donner* employées par les LNN anglophones, bien que non significatifs, montrent que ces apprenants utilisent, d'une part, davantage d'UP construites avec un déterminant et/ou un modifieur quelconque(s) que les LN (Ex. 156) (55,6% pour 40% d'UP natives) et, d'autre part, aucune structure prépositionnelle (contre 4% chez les natifs) et moins d'UP construites sans déterminant (Ex. 157) (8,9% pour 16% d'UP natives).

(156) Des impératifs sont très commun dans ce type de texte, pour **donner des ordres** au destinataire. (FL2-AN, <TAN087>, occ. 345)

(157) La peur de perdre de l'argent par les erreurs de conversion est un soupçonnable valable. Cette difficulté d'estimer les prix nouveaux, **donnera lieu** à une sorte de xenophobie. (FL2-AN, <TAN154>, occ.536)

Contrairement aux anglophones, l'usage des apprenants néerlandophones semble pour sa part refléter de manière plus fidèle l'usage que font les natifs des UP construites avec les verbes *prendre* et *donner*, du point de vue de leur structuration syntaxique. Si l'on observe le Tableau (58) ci-dessous, aucune différence significative ne semble en effet émerger de la comparaison entre les deux corpus néerlandophone et francophone natif.¹⁰⁰ Nous avons néanmoins mis en évidence quelques différences d'emploi (non significatives) entre les productions non natives et les productions natives, notées par un (+) ou un (-) dans le Tableau (58) (ci-dessous).

	Prendre		Donner	
	FL2-NL	FL1	FL2-NL	FL1
Dét zéro + N + (Modifieur/ de N/ P Sub)	7 (-) 11,3%	17,03 [35] 17,9%	9 (-) 12,0%	9,24 [19] 19,0%
Dét + N de N	15 (+) 24,2%	18 [37] 18,9%	10 13,3%	5,35 [11] 11,0%
Dét + N + P Sub	1 1,6%	2,43 [5] 2,6%	15 20,0%	9,73 [20] 20,0%
Dét Poss + N + (Modifieur)	2 3,2%	4,38 [9] 4,6%	5 6,7%	2,92 [6] 6,0%
Prép + SN	8 (-) 12,9%	20,92 [43] 21,9%	0 (-) 0,0%	1,95 [4] 4,0%
Dét + N + (Modifieur)	29 (+) 46,8%	32,59 [67] 34,2%	36 (+) 48,0%	19,46 [40] 40,0%
Total	62 100,0%	95,35 [196] 100%	75 100,0%	48,65 [100] 100%

Tableau 58. Distribution des structures syntaxiques du SN des UP dans le corpus néerlandophone (vs natif) en fonction du verbe utilisé

La fréquence d'emploi des UP dont le SN est construit avec un déterminant « zéro » est inférieure à celle de ces UP dans le corpus natif, avec le verbe *prendre* (11,3% pour 17,9% d'UP natives) (Ex. 158) et avec le verbe *donner* (12% pour 19% d'UP natives) (Ex. 159).

(158) De nos jours, la culture classique passe bien souvent pour admirable et extrêmement élevée. Pourtant, il y en a beaucoup qui disent que cette civilisation a bien eu son temps et qu'elle, par conséquent, est devenu aujourd'hui de moindre importance dans notre vie. (...) La valeur de celle-ci est maintenant indiscutablement prouvée, (...). Cependant, il faut **prendre garde** de ne pas exagérer la prospérité et surtout l'humanité de la société classique. (FL2-NL, <TNL126>, occ. 962)

¹⁰⁰ Nous n'avons pas pu vérifier cette observation de manière statistique, vu le nombre restreint d'occurrences. Ce nombre s'est encore vu réduit par le calcul des fréquences relatives ($n/90.000$ mots), ce qui a donné lieu à des cas de figure où plus de 20% des cases du tableau de contingence avaient une fréquence attendue inférieure à 5, ainsi qu'à des cas de figure où au moins une fréquence attendue était inférieure à 1. Or, dans ces cas de figure, rappelons-le, le test de croisement en chi carré n'est pas validé.

(159) A quoi sert la lecture si l'on peut lire Antigone le matin et être tortionnaire l'après-midi ? Cette question pertinente, posé par un philosophe éminent, **donne lieu** à une réflexion sur la problématique de l'influence des différents aspects de l'éducation et d'une culture classique en particulier. (FL2-NL, <TNL112>, occ. 1095)

La fréquence des UP avec un SN prépositionnel est également inférieure chez les LNN néerlandophones : nous observons 12,9% d'UP prépositionnelles avec *prendre* (Ex. 160) pour 21,9% d'UP natives et l'absence d'UP prépositionnelles avec *donner* pour 4% d'UP prépositionnelles natives.

(160) Jeunes élèves pris en otage et assassinés en Russie, membre de Médecins sans frontières enlevé et décapité en Iraq, deux jeunes amies, Holly et Jessica, tuées dans des conditions cruelles en Angleterre,... Ce ne sont que quelques exemples de crimes terribles commis récemment. (FL2-NL, <TNL168>, occ. 1018)

A l'inverse, les LNN néerlandophones utilisent plus fréquemment certaines structures du SN des UP que les LN, à savoir les structures avec un déterminant et/ou un modifieur quelconque(s) avec les deux verbes *prendre* (46,8% d'UP) (Ex.161) et *donner* (48% d'UP) (Ex. 162), ainsi que les UP construites avec le verbe *prendre* dont le SN est en *N de N* (24,2% d'UP) (Ex. 163).

(161) Et pourtant, il convient de **prendre une distance intellectuelle** par rapport à ces cris du « vox populi » et de se poser des questions, qui montreront que du point de vue éthique et moral, la peine de mort n'est guère défendable (FL2-NL, <TNL216>, occ. 1037)

(162) Les conséquences ne doivent pas être toujours si tragiques : la vie du footballeur Gilles De Bilde a changé sérieusement après l'incident avec le coup de poing : au cours d'un match il **a donné** un autre footballeur un méchant coup-de-poing. (FL2-NL, <TNL091>, occ. 1079)

(163) Les deux frères Bush par exemple, gouverneur chacun d'un Etat, semblent ainsi faire la course pour effectuer le plus d'exécutions. La peine de mort **a** par conséquent **pris l'allure d'un moyen de propagande** comme c'était le cas du pain et des jeux pendant l'ère romaine. (FL2-NL, <TNL147>, occ. 1012)

Pour conclure, nous tenons à souligner l'importance de cette variable formelle qui sera par ailleurs prise en compte lors de notre examen de la *complexité* des productions non natives (*vs* natives) (cf. Chapitre VI).

V.2. Caractéristiques sémantiques

Outre ces caractéristiques formelles, notre approche linguistique s'est aussi intéressée aux caractéristiques sémantiques des UP. A partir de la classification sémantique que nous effectuons des 906 SVN extraites de l'ensemble des corpus L2 et L1 (cf. V.2.1.), nous tenterons de dessiner un profil sémantique des UP utilisées en français L1 et de vérifier dans quelle mesure les emplois phraséologiques des apprenants en L2 reflètent cet usage natif (cf. V.2.2.).

V.2.1. Classification sémantique

Dans notre étude pilote portant sur 200 occurrences de SVN construites avec le verbe *prendre* en français L1 (400.000 mots – corpus journalistiques et littéraires) (cf. I.1.4.), nous avons procédé à une analyse détaillée de chaque SVN en termes de traits sémantiques inhérents ou afférents, actualisés ou non dans l'énoncé en fonction du co(n)texte de production langagière (cf. I.1.4.3., pp. 49-50). Ce travail étant très coûteux en temps, nous avons appliqué à nos 906 SVN une méthode d'analyse sémantique adaptée à la quantité importante des données à traiter. Le critère de base est ici d'ordre cognitif et repose sur une classification des *procès* ou *prédicats* héritée de Vendler (1967, cité dans Recanati & Recanati, 1999) et de Harris (1976, cité dans Kokochkina, 2004). Vendler (1967) a établi sa classification des procès en adoptant à la fois un point de vue cognitif et linguistique visant à prédire la manière dont vont se comporter les unités linguistiques à partir de la manière dont tout sujet cognitif appréhende les événements. Selon Recanati & Recanati (1999 : 167), on peut ainsi distinguer trois types de prédicats (qu'ils soient adjectivaux, verbaux et/ou nominaux) : les actions/ activités, les événements (accomplissement et achèvement) et les états.

Premièrement, les *actions* ou *activités* (par ex., *regarder, courir, marcher, pousser un chariot, vivre*)¹⁰¹ se distinguent des deux autres types de prédicats par leur agentivité. Elles se caractérisent aussi par leur dynamicité, leur caractère non borné dans le temps et leur non-punctualité.

(164) Jean traduit le texte du français en anglais (Ex tiré de Kokochkina, 2004 : 68)

Deuxièmement, les *événements* se caractérisent par leur caractère temporellement borné, dynamique et télique. Ils peuvent être soit des *accomplissements*, avec un aspect duratif non ponctuel (par ex., *accoucher, construire une maison, traverser la rue, persuader, dessiner un cercle*), soit des *achèvements*, avec un aspect terminatif ou ponctuel, lié au résultat de l'événement (par ex., *reconnaître, trouver, gagner la course, atteindre le sommet, naître, mourir*).

(165) Le concert de Pavarotti aura lieu à 9 heures du soir (Ex. tiré de Kokochkina, 2004 : 68)

¹⁰¹ Les exemples entre parenthèses sont tirés de Recanati & Recanati (1999 : 167). Il est important de souligner que ces exemples sont décontextualisés et que leur emploi est donc parfois ambigu (par ex., *vivre* est repris ici dans les actions, alors qu'il possède aussi les traits dynamique, borné et télique des événements/ accomplissement). C'est le chercheur qui tranchera entre l'un ou l'autre sens plus ou moins marqué en fonction du contexte.

Troisièmement, les *états* se caractérisent par leur caractère non borné, statique, a-télique et non ponctuel (par ex., *savoir, aimer, être malade, désirer, dominer*)¹⁰². C'est principalement la dynamicité (absente des prédicats d'*états*) qui permettra d'opposer les descriptions d'états à celles d'événements.

(166) Les invités sont las (Ex. tiré de Kokochkina, 2004 : 68)

Nous présentons ci-dessous un organigramme hiérarchique (cf. Figure 14) rendant compte de la classification à laquelle nous sommes arrivée à partir de ces notions. Cet organigramme a été élaboré à partir de nos données de corpus, qui sont authentiques et contextualisées. Cette catégorisation repose sur l'examen des sens effectifs du nom postverbal qui ont été observés dans leur contexte d'énonciation.

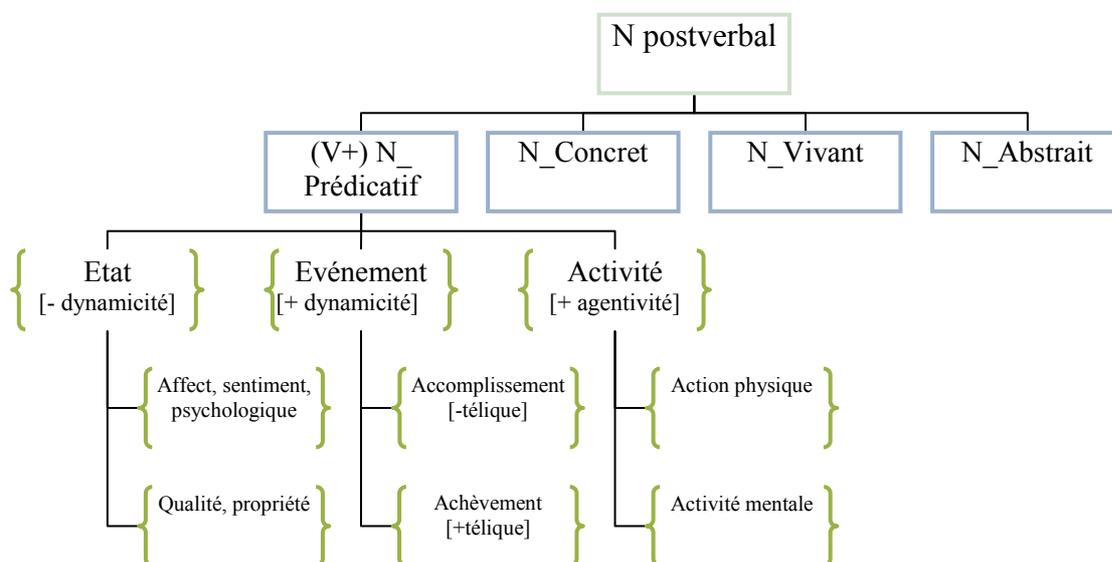


Figure 14. Classification sémantique « à la Vendler » des SVN construites avec *prendre et donner*.

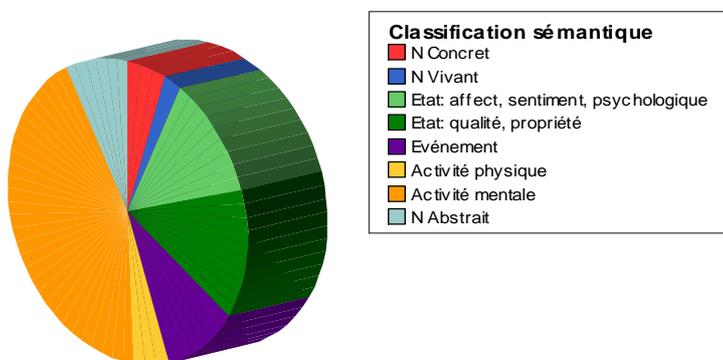
A partir de cette hiérarchisation sémantique, nous avons organisé et listé les valeurs à attribuer aux séquences étudiées. Nous avons ainsi créé une variable générale qui regroupe les classes sémantiques les plus saillantes par rapport à nos données de corpus. A noter que nous avons subdivisé certaines classes pour différencier les emplois dont la distribution était importante : d'une part, la classe des prédicats d'*états* a été divisée en prédicats *affectifs* et *qualifiants*, et, d'autre part, la classe des prédicats d'*actions* distingue les *actions physiques* des *activités mentales*.

¹⁰² Nous verrons que notre classification, bien que largement inspirée de la réflexion des auteurs sur le sujet, s'en éloigne quelque peu. Ainsi, nous considérerons certains prédicats classés par les auteurs dans les prédicats d'*états* (par ex. *savoir*), comme appartenant à une sous-classe d'*activités* (cf. notre classification dans le Tableau 59, *infra*). C'est le trait de l'agentivité qui s'est avéré discriminant dans ce cas.

Nom de la variable (<i>Name</i>)	Variable (<i>Label</i>)	Valeurs (<i>Values</i>)
Sém1ClassGénR2	Classification sémantique générale	1. N_Concret 2. et 3. N_Vivant : N_Humain et N_Animal/Végétal 4. Etat : affect, sentiment, psychologique 5. Etat : qualité, propriété 6. Evénement 7. Activité : action physique 8. Activité mentale : cognitif, épistémique, volitif 9. N_Abstrait

Tableau 59. Extrait de la liste paramétrique – Classification sémantique générale

Ces subdivisions se sont révélées pertinentes puisque représentatives des emplois les plus fréquents dans les corpus. Le graphe ci-dessous (Graphique 3) donne un aperçu de la répartition des occurrences dans l'ensemble des corpus (L1 et L2) en fonction de leur appartenance à l'une des huit catégories sémantiques précitées (*Sém1ClassGénR2*) : 42,3% des SVN étudiées appartiennent majoritairement à la catégorie des activités mentales (en orange dans le Graphique 3), pour un peu moins de 5% d'actions physiques (en jaune) ; 26,7% des occurrences concernent ensuite l'expression d'états (en vert clair et foncé) ; enfin, 10,5% représentent les prédicats d'événements, alors que les catégories restantes contiennent chacune moins de 10% des SVN étudiées.



Graphique 3. Distribution des différentes catégories sémantiques dans l'ensemble des corpus

Alors que pour les noms concrets, vivants et abstraits leur appartenance à une classe sémantique a été décidée de manière relativement non ambiguë, l'appartenance des noms à valeur prédicative (états, événements et activités) s'est révélée plus complexe à déterminer. En effet, le sens prédicatif d'une SVN peut reposer sur l'élément nominal ou verbal, mais peut aussi émaner de la combinaison des deux éléments. Rappelons en effet

que, par définition, dans une structure *délexicale* ou à *verbe support*, le constituant verbal est (partiellement) désémantisé (cf. I.1.3.2., pp. 42-44). L'appartenance à une classe sémantique dépend dans ce cas du sens global (i.e. du *signifié global* partiellement non compositionnel) de la SVN étudiée, et non plus seulement du sens de l'élément nominal postverbal.

V.2.1.1 Noms concrets, vivants et abstraits

Pour chaque grande classe sémantique (N_Concret, N_Vivant, N_Abstrait, N_Prédicatif) (cf. Figure 14, *supra*), nous avons créé une liste paramétrique catégorisant les emplois rencontrés dans nos corpus de manière plus spécifique. Nous présentons ci-dessous les trois premières classes relatives aux noms concrets (*Sém2Concret*), vivants (*Sém3Vivant*) et abstraits (*Sém5Abstrait*). Quelques exemples de cas rencontrés dans notre corpus sont proposés entre parenthèses sous la forme V + N, i.e. *prendre* (noté P) ou *donner* (noté D) suivi du nom postverbal sous sa forme *type* lemmatisée.

Nom de la variable (<i>Name</i>)	Variable (<i>Label</i>)	Valeurs (<i>Values</i>)
Sém2Concret	Classification sémantique des N concrets, inanimés	1. Lieu, habitation (Ex. : P+ <i>appartement</i> ; D+ <i>chambre</i>) 2. Objet utilitaire (Ex. : P+ <i>serviette</i> ; D+ <i>meuble</i>) 3. Vestimentaire, bijou (Ex. : P+ <i>montre/ sac</i> ; D+ <i>chemise</i>) 4. Technologie, télécommunication (Ex. : P+ <i>portable</i>) 5. Librairie, papeterie (Ex. : P+ <i>livre</i> ; D+ <i>carte d'identité</i>) 6. Pécuniaire (Ex. : D+ <i>virement</i> , D+ <i>argent</i>) 7. Produits alimentaires (Ex. : P+ <i>fruit/ biscuit</i> ; D+ <i>chocolat</i>)
Sém3Vivant	Classification sémantique des N vivants	1. Nom propre (Ex. : P+ <i>Hoffman/ Mohammed Ali</i>) 2. Anonyme (Ex. : P+ <i>personne</i>) 3. Relation familiale, affective (Ex. : P+ <i>femme/ enfant/ parent</i>) 4. Fonction sociale (Ex. : P+ <i>élève/ artiste</i> ; D+ <i>président</i>) 5. Qualificatif (Ex. : P+ <i>as</i>) 6. Collectif (P+ <i>gens</i>) 7. Partie du corps (Ex. : P+ <i>par le bras/ sang</i> ; D+ <i>cerveau</i>) 8. Animal (Ex. : P+ <i>chien</i>) 9. Végétal (Ex. : D+ <i>fleur/ bouquet</i>)
Sém5Abstrait	Classification sémantique des N abstraits	1. Matière, sujet, thème, pensée (Ex. : P+ <i>matière</i> ; D+ <i>sujet/ matière/ (3)idée</i>) 2. Repère, modèle, croyance (Ex. : D+ <i>base/ modèle/ repère/ religion/ (2)exemple/ canon de beauté/ pilier/ (2)valeur/ vérité</i>) 3. Arts (Ex. : P+ <i>cinéma/ documentaire</i>) ¹⁰³ 4. Société (Ex. : P+ <i>travail/ emploi</i> ; D+ <i>administration/ programme/ (2)droit/ (1)pouvoir/ service de la santé</i>) 5. Relation métonymique, pluralité/singularité (Ex. : P+ <i>élément/ échantillon</i> ; D+ <i>pléthore</i>)

Tableau 60. Extrait de la liste paramétrique – Classification sémantique des N_Concrets, N_Vivants et N_Abstrait

¹⁰³ Ces cas de SVN sont inclus dans la catégorie des N_Abstrait par le sens général de 'discipline artistique' qu'ils recouvrent dans le contexte d'énonciation en cause.

Nous tenons à souligner que les exemples cités sont ici *décontextualisés*. Il faut cependant garder à l'esprit que leur appartenance à l'une ou l'autre classe sémantique a été déterminée en fonction de leur *contexte* d'énonciation. Prenons le cas de la SVN *donner* *Dét idée*, notée ici D+(3)*idée*, renvoyant au sens de 'transmettre/conférer des idées', que nous illustrons dans l'exemple (167) ci-dessous.

Exemple d'une SVN avec le nom *idée*, classé dans la catégorie des N_Abstrait:

(167) Mais pour les musulmans, leur religion leur **donne** certaines idées au sujet des femmes et du mariage (FL2-AN, <TAN241>, occ. 501)

Nous retrouvons cette SVN *type* dans deux autres catégories de prédicats. D'une part, l'acception notée D+(2)*idée* appartient à la catégorie des *événements* mettant l'accent sur le résultat ou la finalité (Ex. 50), i.e. les *achèvements* (valeur 8 dans *Sém4Prédicatif*, cf. Tableau 63, *infra*).

Exemple d'une SVN avec le nom *idée*, classé dans la catégorie des événements:

(50) La révolution à **donné** aux Français l'idée d'un Etat-nation qu'on ne veut pas perdre (FL2-AN, <TAN251>, occ.522)

D'autre part, l'acception notée D+(1)*idée* appartient à la catégorie des *affects* qui véhiculent un sens passif d' 'impression' (Ex. 51) (valeur 2 dans *Sem4Prédicatif*, cf. Tableau 61, *infra*).

Exemple d'une SVN avec le nom *idée*, classé dans la catégorie des affects:

(51) Le contraste entre les deux livres dans ce domaine est frappant et nous **donne** une idée de l'influence de la religion il y a cent ans, comparé avec celle d'aujourd'hui (FL2-AN, <TAN221>, occ.440)

De nouveau, c'est l'examen de l'occurrence en cotexte qui nous a permis de discerner les emplois sémantiquement distincts pour une même séquence *type*. Quand c'était le cas pour une séquence, nous l'avons chaque fois signalé dans les exemples en notant devant le constituant nominal le numéro de l'acception à laquelle il renvoyait. Par exemple, D+(2)*exemple* correspond à l'acception de 'modèle' (Ex. 168), par opposition à celle d' 'illustration' (notée D+(1)*exemple*) (Ex. 169), qui renvoient respectivement aux deux catégories sémantiques des N_Abstrait (cf. Tableau 60, ci-dessus) et des activités mentales relatives à l'organisation conceptuelle ou langagière (cf. valeur 18 dans le Tableau 65, *infra*).

(168) En dernier lieu, il faut se rendre compte de l'exemple que l'on donne aux jeunes à travers l'application de la peine de mort. (FL2-NL, <TNL186>, occ. 1206) [Acception de 'modèle': D+(2)*exemple*]

(169) Le texte est principalement au passé composé parce qu'il parle de la vie d'une femme, principalement son histoire ; « elle a épousé, « ? est empressée? », « ?s'est illustré?. » pour **donner** des exemples. (FL2-AN, <TAN112>, occ. 374) [Acception d' 'illustration': D+(1)*exemple*]

Nous examinons ci-dessous les classes sémantiques des emplois prédicatifs qui recouvrent le plus grand nombre de nos données de corpus, c'est-à-dire 84,2% des occurrences.

V.2.1.2. Emplois prédicatifs : états, événements et actions

Rappelons tout d’abord que les emplois prédicatifs comprennent des constructions verbales dites *délexicales* ou à *verbe support*, dans lesquelles c’est le constituant nominal qui possède la valeur prédicative généralement attribuée au verbe, comme c’est le cas dans l’exemple (170), répondant à la définition traditionnelle que l’on donne de ce type de construction.

(170) Le Sénat réflète sur les problèmes communautaires et les députés dans la Chambre des Représentants **prennent** les décisions (FL2-NL, <TNL001>, occ. 921)

Mais, comme nous l’avons déjà signalé précédemment, d’autres emplois sont moins tranchés. Rappelons que cette analyse sémantique porte sur toutes les séquences extraites de notre corpus, qu’elles soient plus ou moins libres ou figées. Nous avons ainsi classé dans la catégorie des emplois prédicatifs des constructions qui ne sont pas *délexicales* à proprement parler, mais qui possèdent un sens prédicatif via la combinatoire verbo-nominale. Le sens prédicatif naît alors de la co-apparition en contexte des deux éléments verbal et nominal (Ex. 171, ci-dessous).

(171) Une standardisation des moeurs et de la culture peut amener à étouffer les richesses mais, d’ autre part, des échanges et des confrontations entre les richesses culturelles des différents pays peut **donner** des résultats inattendus (FL1, <FRFB312>, occ. 797)

Dans le cas de ce type d’emplois prédicatifs, la valeur prédicative est en quelque sorte partagée, souvent de manière inégale, entre les deux constituants.

Plusieurs catégories ont été mises en exergue pour les occurrences étudiées dans notre travail. Nous les répertorions dans les tableaux récapitulatifs présentés dans les pages qui suivent. Un premier tableau (cf. Tableau 61, ci-dessous) extrait de notre liste paramétrique concerne les prédicats d’états relatifs aux affects et aux sentiments, ainsi qu’aux états psychologiques et relationnels.

Nom de la variable (Name)	Variable (Label)	Valeurs (Values)
Sém4Prédicatif	Classification sémantique des N prédicatifs	<p><u>Etat</u>: Affect (12,7% des SVN)</p> <p>1. Emotion, sentiment (Ex. : D+<i>confiance/ émotion/ courage/ (1)sentiment/ envie/ contentement/ satisfaction/ goût</i> ; P+<i>prédilection</i>)</p> <p>2. Impression, point de vue [-agentivité] (Ex. : D+(1)<i>idée/ image/ (1)impression/ illusion/ (2)sentiment</i>)</p> <p>3. Psychologique, relationnel (Ex. : P+<i>pour modèle/ responsabilité/ en charge/ soin/ risque/ garde/ pouvoir</i>; D+<i>autonomie/ indépendance</i>)</p>

Tableau 61. Extrait de la liste paramétrique – Classification sémantique des N_Predicatifs d’état affectif/ sentimental et psychologique/ relationnel

Nous soulignons la présence ici d'une catégorie que l'on rattacherait plutôt aux prédicats d'activité mentale de type épistémique. Il s'agit en fait d'une sous-catégorie des prédicats de sentiment exprimant une 'impression' (Ex. 172 et 48, ci-dessous) dont le caractère *non dynamique* et *non agentif* les distingue des prédicats d'événement et d'action (cf. la définition des états, p. 251).

(172) En effet, voir d'autres personnes agir comme nous nous **donne l'impression qu'elles partagent nos désirs, nos aspirations, nos angoisses**, (48) nous **donne le sentiment d'être nous-même compris**, ce qui nous sécurise (FL1, <TFRB199>, occ. 868 et 869)

Cette catégorie est donc à distinguer de la catégorie des prédicats d'activité mentale repris sous une appellation similaire qui expriment le 'point de vue' ou la 'perspective' (valeur 20 dans le Tableau 65, *infra*). Ceux-ci se caractérisent en effet par leur caractère *dynamique* et leur *agentivité*, comme dans l'exemple (173) ci-dessous.

(173) Le roman a de nos jours une connotation péjorative : bon nombre de gens trouvent que le contenu et l'atmosphère qui l'entourent peuvent être décrits beaucoup plus bref. C'est pourquoi l'évolution cinématographique a connu un tel essor puisqu'elle pouvait lancer immédiatement les spectateur dans un monde irréel et exotique. (...) J'ai seulement voulu vous **donner quelques impressions** afin de renouveler le genre romanesque. Ce qui est l'essentiel, c'est que la littérature est à retrouver dans l'évolution cinématographique, plus que nous nous en rendons compte ! (FL2-NL, <TNL067>, occ. 1068)

Une autre catégorie d'états relève de la caractérisation et de la qualification. Nous avons subdivisé cette catégorie en trois types de propriétés différentes : les propriétés portant sur la qualité, la faculté ou l'apparence (valeur 4 de la variable *Sém4Prédicatif*, cf. le Tableau 62, ci-dessous) ; les propriétés portant sur l'intensité et l'importance (valeur 5 dans le le Tableau 62, ci-dessous) ; les propriétés portant sur des rapports de force du type avantage vs inconvénient, ou des rapports de supériorité vs infériorité (valeur 6 dans le Tableau 62, ci-dessous).

Nom de la variable (Name)	Variable (Label)	Valeurs (Values)
Sém4Prédicatif	Classification sémantique des N prédicatifs	<p><u>Etat</u>: Caractérisation, propriétés (14% des SVN)</p> <p>4. Qualité, forme, apparence, faculté (Ex. : P+<i>expression/ trait/ allure/ (1)forme</i> ; D+<i>bon sens/ façon de penser/ personnalité/ compétence/ statut</i>)</p> <p>5. Intensité, importance, force, mesure (Ex. : P+<i>ampleur/ proportion/ importance/ (1)valeur/ envol/ essor</i> ; D+<i>coup de collier/ dynamique/ prestige/ impulsion/ second souffle/ (2)pouvoir</i>)</p> <p>6. Avantage/Inconvénient, Supériorité/Infériorité, Avance/Retard (Ex. : P+<i>avantage/ part du gâteau/ position prédominante/ longueur d'avance</i> ; D+<i>arme supplémentaire/ apport/ faveur/ priorité</i>)</p>

Tableau 62. Extrait de la liste paramétrique – Classification sémantique des N_Prédicatifs d'état de qualité/ propriétés

Nous attirons l'attention du lecteur sur le cas de la SVN construite avec le verbe *prendre* et le lexème nominal *forme*. Les occurrences de cette séquence *type* sont en effet incluses

dans la catégorie des états uniquement quand elles sont construites avec un déterminant, comme c'est le cas de l'exemple (174).

(174) Mais les divisions internes entre Européens **prennent** parfois la forme, moins tragique et plus sournoise à la fois, de mesquineries politiques entre peuples (FL1, <TFRB284>, occ. 545)

Par contraste, les constructions sans déterminant appartiennent à la catégorie des événements du type *achèvement* (valeur 8 dans le Tableau 63, ci-dessous), puisqu'elles signifient le plus souvent 'naître' ou 'apparaître', comme c'est le cas dans l'exemple (175). Ces emplois se caractérisent donc par leur caractère borné dans le temps et leur orientation téléique.

(175) L'Europe de 1992 se rapproche de nous à grands pas. Ce vaste programme économique, accueilli avec un concert d'éloges et d'un intérêt vital pour tous les citoyens de la Communauté, est un vieux rêve qui deviendra donc réalité. Cependant, tout reste encore à faire, de ferventes éloges ne suffisent donc pas pour que le projet **prenne forme**. (FL1, <TFRB318>, occ. 578)

Nom de la variable (<i>Name</i>)	Variable (<i>Label</i>)	Valeurs (<i>Values</i>)
Sém4Prédicatif	Classification sémantique des N prédicatifs	<u>Événement</u> [+dynamique] (10,5% des SVN) 7. Accomplissement (duratif) (Ex. : P+ <i>habitude/ lecture</i> ; D+ <i>cours/ enseignement/ formation/ conférence de presse</i>) 8. Achèvement (résultat) (Ex. : P+(2) <i>forme/ fin</i> ; D+(1) <i>naissance/ jour/ résultat/ (2)idée</i>)

Tableau 63. Extrait de la liste paramétrique – Classification sémantique des N_Predicatifs d'événement

Au sein des *actions* ou *activités*, nous distinguons les *activités physiques* des *activités mentales*. Rappelons que les premières, catégorisées dans le Tableau (64) (ci-dessous) ne représentent que 4,7% des occurrences étudiées, alors que les *activités mentales* concernent quasiment une occurrence sur deux (42,3% des occurrences).

Nom de la variable (<i>Name</i>)	Variable (<i>Label</i>)	Valeurs (<i>Values</i>)
Sém4Prédicatif	Classification sémantique des N prédicatifs	<u>Activité</u> [+agentivité]: Action physique (4,7% des SVN) 9. Perception, sens (Ex. : D+ <i>coup d'œil</i>) 91. Mettre au monde, accoucher (Ex. : D+(2) <i>naissance/ vie</i>) 10. Mouvement, préhension, transfert (Ex. : P+ <i>dans ses mains/ des mains de X / route</i> ; D+ <i>injection létale</i>) 11. Photographie (Ex. : P+ <i>photo</i>) 12. Hygiène (Ex. : P+ <i>bain/ douche</i>) 13. Aggression, guerre (Ex. : P+ <i>les armes</i> ; D+ <i>coup de poing</i>) 14. (Télé-)communication (Ex. : P+ <i>sondage</i> ; D+ <i>coup de fil</i>) 15. Consommable (Ex. : P+ <i>thé/ dîner/ consommation</i>) 16. Moyen de transport (Ex. : P+ <i>train/ bac/ autobus/ voiture</i>)

Tableau 64. Extrait de la liste paramétrique – Classification sémantique des N_Predicatifs d'action physique

Les activités mentales sont de loin les plus nombreuses au sein de nos emplois prédicatifs. Cette prédominance semble liée aux types de textes qui constituent nos corpus, à savoir des *compositions académiques* rédigées par des étudiants universitaires dans le cadre de

leur cursus scolaire. Rappelons en outre que la majorité de nos textes sont *expositifs* et que ces derniers ont pour fonction première la transmission et le partage d'informations avec un interlocuteur ou un public averti. L'objectif est, en d'autres termes, de s'exprimer de manière claire et structurée pour rendre l'information la plus lisible et le texte le plus efficace possible. Rien d'étonnant, dès lors, que nous trouvons un grand nombre de SVN à contenu prédicatif ayant trait à la nécessité de transmettre des connaissances (valeur 17 du Tableau 65, ci-dessous), d'organiser ses pensées en structurant son langage (valeur 18 du Tableau 65), de se positionner (valeurs 19 et 21 du Tableau 65) et de donner du relief à son propos (valeur 20 du Tableau 65), ou encore la nécessité d'affirmer sa volonté et ses décisions (valeur 22 du Tableau 65). Nous aurons l'occasion d'étudier plus en détail l'influence potentielle du type de textes sur la production d'UP dans le point V.3.2.

Nom de la variable (<i>Name</i>)	Variable (<i>Label</i>)	Valeurs (<i>Values</i>)
Sém4Prédicatif	Classification sémantique des N prédicatifs	<p><u>Activité [+agentivité]</u>: Activité mentale (42,2% des SVN)</p> <p>17. Cognitif, conscience (Ex. : P+<i>en compte/ conscience</i>)</p> <p>18. Organisation conceptuelle, langagière (Ex. : P+<i>dans un N sens/ N au pied de la lettre</i> ; D+<i>appellation/ titre/ nom/ sens/ terme/ (1)exemple/ description/ définition/ explication/ information/ résumé/ réponse/ message/ argument/ parallèle</i>)</p> <p>19. Croyance, attitude (Ex. : P+<i>position/ part/ parti/ avec mesure/ exemple(sur)/ en modèle</i> ; D+<i>garantie/ jugement/ raison/ soutien</i>)</p> <p>20. Point de vue, perspective [+agentivité] (Ex. : P+<i>dans son ensemble</i> ; D+ <i>aperçu/ vision/ (2)impression</i>)</p> <p>21. Permissif, choix, opportunité (Ex. : P+<i>chance</i> ; D+<i>choix/ possibilité/ opportunité/ occasion/ (1)droit/ (3)pouvoir/ son aval</i>)</p> <p>22. Volitif, conatif (Ex. : P+<i>décision/ initiative/ mesure</i> ; D+<i>conseil/ instruction/ preuve/ solution</i>)</p>

Tableau 65. Extrait de la liste paramétrique – Classification sémantique des N_Prédicatifs d'activité mentale

V.2.2. Profil sémantique des UP

Ayant analysé chacune des SVN extraites des corpus non natifs et natif, nous donnons ci-dessous une idée de la distribution générale de ces séquences. Observant de plus près la répartition des 906 occurrences, on constate à première vue une relative homogénéité entre les pourcentages de chaque sous-corpus.

	FL2-AN	FL2-NL	FL1	Total
N concret	18,49 [38] 9,8%	0 0%	3,41 [7] 2%	[45] 5,0%
N vivant	5,84 [12] 3,1%	5 3,0%	1,46 [3] ,9%	[20] 2,2%
N prédicatif: état	45,73 [94] 24,2%	42 24,9%	51,57 [106] 30,5%	[242] 26,7%
N prédicatif: événement	14,11 [29] 7,5%	31 18,3%	17,03 [35] 10,1%	[95] 10,5%
N prédicatif: activité	87,57 [180] 46,3%	76 45,0%	82,7 [170] 48,9%	[426] 47,0%
N abstrait	17,51 [36] 9,3%	15 8,9%	13,13 [27] 7,8%	[78] 8,6%
Total	189,24 [389] 100,0%	169 100,0%	169,3 [348] 100,0%	[906] 100,0%

Tableau 66. Distribution des 906 SVN dans les corpus non natifs et natifs, en fonction de leur répartition en catégories sémantiques

Les locuteurs non natifs et natifs utilisent principalement des SVN à contenu prédicatif d'action (plus de 45% des SVN dans chaque corpus) et secondairement des SVN à contenu prédicatif exprimant un état (environ 25% des SVN dans chaque corpus). Nous avons néanmoins mis en évidence, dans le tableau ci-dessus, quelques tendances : d'une part, les LNN anglophones emploient plus de constructions avec un nom concret (9,8% des SVN) (Ex. 176) que les LN (2% des SVN) et, d'autre part, les LNN néerlandophones emploient plus de constructions avec un nom prédicatif d'événement (18,3% des SVN) (Ex. 177) que les LN (10,1% des SVN).

(176) Je lui **ai donné** une chemise neuve et des chocolats. C'était une bonne journée! (FL2-AN, <TAN051>, occ. 332 et 333) [= SL]

(177) Ces derniers ont prétendu que le sphinx près de Luxor daterait de 12500 avant Jésus Christ au lieu de 4000 avant Jésus Christ. Pour démontrer cela, ils invoquent entre autres les deux types d'érosion que montrent la partie supérieure et inférieure du sphinx. La partie inférieure est érodée par le coulement d'eau (qui **a pris** des décennies et des décennies). (FL2-NL, <TNL120>, occ. 960) [= SL]

Notre objectif étant de dresser un profil sémantique des UP, nous ne retiendrons de ces classes sémantiques que celles qui concernent la majorité des UP (surlignées en gris dans le Tableau 67), i.e. les emplois à valeur prédicative d'activité mentale (52,7% des UP), d'état (28,6% des UP) et d'événement (10,5% des UP). Notons que nous n'avons pas gardé, au sein des prédicats d'action, les actions physiques qui représentent une proportion infime de nos UP (5%).

	Fréquence	%
N concret	3	0,4
N vivant	1	0,1
N prédicatif: état	201	28,6
N prédicatif: événement	74	10,5
N prédicatif: activité mentale	371	52,7
N prédicatif: action physique	35	5,0
N abstrait	18	2,6
Total	703	100,0

Tableau 67. Fréquence des 703 UP des corpus non natifs et natifs, en fonction de leur répartition en catégories sémantiques

De la répartition des données phraséologiques dans les trois corpus (cf. Tableau 68, ci-dessous), il ressort une tendance des LNN néerlandophones à suremployer les UP d'événement (Ex. 178) (21,5% d'UP contre 10,5% en L1).

(178) Même dans nos loisirs nous devons nous dépêcher, car la pression sociale ne nous pousse pas uniquement à la lecture, mais également à pratiquer un sport et à la charité. Par conséquent, nous sommes toujours en train de faire quelque chose, nous n'arrivons plus à **prendre le temps**, à réfléchir sur un livre qu'on a lu, sur son sens, sur l'intention de son auteur. (FL2-NL, <TNL103>, occ. 952)

Il y a en outre un emploi moins fréquent, par ces apprenants, d'UP appartenant aux catégories prédicatives d'activité mentale (Ex. 179) (49,6% d'UP contre 57,1% en L1) et d'état (Ex. 180) (28,9% d'UP contre 32,4% en L1).

(179) Pour cela, il fallait une modification de la mise en page. Les rédacteurs ont confirmé que le porte-parole du directeur **a donné son accord**. (FL2-NL, <TNL005>, occ. 1058)

(180) Essayer d'atteindre son but, d'obtenir des résultats en réalisant un idéal, son idéal, peut inciter à agir et **donner du courage** à vivre pour soi même et pour les autres (FL2-NL, <TNL086>, 1077)

Les LNN anglophones, pour leur part, montrent la tendance inverse à celle des néerlandophones, puisqu'ils tendent à suremployer les UP d'activité mentale (Ex. 181) (61,8% des UP).

(181) Ils pensent que dans le proche avenir ce parlement **prendra les décisions politiques et économiques** pour chaque pays. (FL2-AN, <TAN193>, occ. 144)

Les résultats concernant les deux autres types de prédicats ne semblent pas montrer de différence marquée par rapport à la fréquence d'emploi des natifs.

	FL2-AN	FL2-NL	FL1	Total
N prédicatif: état	35,51 [73] 30,7%	35 28,9%	45,24 [93] 32,4%	[201] 31,1%
N prédicatif: événement	8,76 [18] 7,6%	26 (+) 21,5%	14,59 [30] 10,5%	[74] 11,5%
N prédicatif: activité mentale	71,51 [147] (+) 61,8%	60 49,6%	79,78 [164] 57,1%	[371] 57,4%
Total	115,78 [238] 100,0%	121 100,0%	139,62 [287] 100,0%	[646] 100,0%

Tableau 68. Distribution des 646 UP prédicatives (sans les actions physiques) dans les corpus non natifs et natifs

La différence entre les corpus s'est avérée significative ($\chi^2 = 11.803$, ddl = 4, $p < *0.05$) du point de vue de l'appartenance des UP à une catégorie sémantique de prédicats particulière. Néanmoins, quand on compare le corpus anglophone au corpus de contrôle, la différence ne s'avère pas significative ($\chi^2 = 0.864$, ddl = 2, $p > 0.5$ (NS)), alors que la différence entre les néerlandophones et les natifs est significative ($\chi^2 = 6.015$, ddl = 2, $p < *0.05$). Ceci tend à appuyer la thèse du suremploi des UP à valeur événementielle et au sous-emploi des UP à valeur cognitive (i.e. d'activité mentale) par les LNN néerlandophones, contrairement aux LNN anglophones qui se rapprocheraient donc de l'usage phraséologique natif sur le plan sémantique. Examinant de plus près le cas des néerlandophones, nous observons que les UP d'événement sont dans 3 cas sur 4 des *accomplissements* (Ex. 182) pour seulement 1 cas sur 4 chez les LN, alors que les UP d'*achèvement* (Ex. 183) montrent une proportion d'emplois équivalente dans les deux corpus. Le suremploi serait donc lié à la fréquence d'emplois des UP véhiculant en particulier un sens événementiel d'*achèvement*.

(182) Peut-être que le roman va devenir moins populaire, parce que lire un roman **prend beaucoup de temps** et dans la société actuelle tout doit avancer. (FL2-NL, <TNL055>, occ. 933)

(183) Le premier aspect économique du grand marché **donnera lieu** ensuite à une politique extérieure commune. (FL1, <TFRB296>, occ.790)

Nous pensons que ces différences de fréquence entre les trois corpus, du point de vue du sémantisme des UP, ne peuvent être interprétées qu'à la lumière d'une étude mettant en rapport le sémantisme des UP et la thématique du texte. Cette différence tiendrait bien plus à la nature de l'information véhiculée par le texte qu'à des lacunes portant sur une capacité de l'apprenant à associer un sens à une UP. Cependant, il nous paraît indispensable d'étudier les UP d'un point de vue sémantique à partir du moment où l'on étudie les emplois phraséologiques de deux verbes, en les comparant. Nous verrons donc ci-dessous si les UP construites avec les verbes *prendre* et *donner* se distinguent les unes des autres au niveau sémantique, en fonction des corpus non natifs ou natif où elles apparaissent. Autrement dit, nous avons tenté d'établir ci-dessous le profil sémantique des

UP construites avec les verbes *prendre* vs *donner*, en limitant l'étude aux cas d'UP couvrant les catégories sémantiques les plus fréquentes, i.e. les emplois prédicatifs d'événement, d'état et d'activité mentale.

L'usage natif montre une différence très significative ($\chi^2 = 20.413$, ddl = 2, $p = *0.000$) entre les UP construites avec le verbe *prendre* et celles construites avec *donner*, en fonction de leur catégorisation sémantique en classes de prédicats.

	Prendre	Donner	Total
N prédicatif: état	51	42	93
	54,8%	45,2%	100,0%
	26,2%	45,7%	32,4%
N prédicatif: événement	15	15	30
	50,0%	50,0%	100,0%
	7,7%	16,3%	10,5%
N prédicatif: activité mentale	129	35	164
	78,7%	21,3%	100,0%
	66,2%	38,0%	57,1%
Total	195	92	287
	67,9%	32,1%	100,0%
	100,0%	100,0%	100,0%

Tableau 69. Profil syntaxique des syntagmes nominaux dans les UP construites avec *prendre* vs *donner* en français L1

Selon la répartition présentée dans le Tableau (69) ci-dessus, les UP construites avec *prendre* sont en majorité des UP à valeur cognitive (66,2% des cas) (Ex. 184), puis des UP prédicatives d'état (26,2% des cas) (Ex. 185) et enfin, dans une minorité de cas, des UP événementielles (7,7% des cas) (Ex. 186).

Exemple d'UP à valeur cognitive :

(184) L' Angleterre de Madame Thatcher a protesté l' an dernier lors de l' accord concernant l' union monétaire entre les Etats membres. Cette réaction nationale nous montre que certains pays **prennent conscience de leur perte d'identité et d'indépendance progressive.** (FL1, <TFRB310>, occ. 561)

Exemple d'UP à valeur d'état (propriété, qualité, forme):

(185) Celle-ci concrétisera-t' -elle ce vieux rêve d'une Europe unie? Rêve qui **prit des allures sérieuses,** lorsqu' en 1951, la Belgique, l' ancienne République fédérale d' Allemagne, la France, l' Italie et les Pays-bas, signèrent le traité instaurant la communauté européenne du charbon et de l' acier. (FL1, <TFRB301>, occ. 591)

Exemple d'UP à valeur événementielle (achèvement):

(186) Le mouvement européen est bien présent, mais il doit s'imposer. Lentement, il imprènera les personnes, il mouillera le noeud des identités et y **prendra place.** (FL1, <TFRB290>, occ. 582)

Les UP construites avec le verbe *donner* concernent quant à elles principalement des emplois prédicatifs d'état (45,7% des cas) (Ex. 187). Elles montrent ensuite une préférence pour des emplois prédicatifs à valeur cognitive (38% des cas) (Ex. 188) et enfin, comme pour le verbe *prendre*, une minorité de cas d'emplois événementiels (16,3% des cas) (Ex. 183).

Exemple d'UP à valeur d'état (affect, sentiment) :

(187) Ensuite, le cinéma, par son impact sur les spectateurs permet de faire découvrir à un très large public, de grandes œuvres littéraires lorsque le film est une adaptation cinématographique d'un livre. Ces adaptations peuvent **donner** l'envie aux spectateurs, après avoir vu le film, de découvrir l'œuvre originale. (FL1, <TFRB010>, occ. 827)

Exemple d'UP à valeur cognitive (épistémique, attitude) :

(188) Ils ont la circonstance atténuante de n' avoir pas eu le choix et d' avoir été "annexés" sans **avoir donné leur avis**. (FL1, <TFRB309>, occ. 795)

Exemple d'UP à valeur événementielle (achèvement):

(183) Le premier aspect économique du grand marché **donnera lieu** ensuite à une politique extérieure commune. (FL1, <TFRB296>, occ.790)

Soulignons également que les emplois cognitifs concernent, dans plus de 3 cas sur 4 (78,7%), les UP incluant le verbe *prendre*, alors que les autres catégories de prédicats sémantiques concernent les deux verbes dans une proportion identique. Suite à ces observations, nous pouvons conclure que le profil sémantique des UP dans l'usage natif s'avère différent pour le verbe *prendre* et pour le verbe *donner*.

Observons maintenant les emplois de ces verbes, en termes de catégorisation sémantique, dans les corpus non natifs.

Premièrement, comme le montrent les résultats présentés dans le Tableau (70) ci-dessous, la différence d'emploi sémantique des UP entre les LNN anglophones et les LN n'est pas significative dans le cas du verbe *prendre* ($\chi^2 = 0.558$, ddl = 2, $p > 0.5$ (NS)). Par contre, dans le cas du verbe *donner*, l'usage des LNN ne semble pas refléter l'emploi des LN: la différence entre les UP produites par les LNN anglophones et les LN est significative du point de vue de la catégorisation sémantique des UP ($\chi^2 = 13.216$, ddl = 2, $p < *0.01$). Cette différence réside dans la tendance que les anglophones ont de suremployer les UP cognitives (61,6%) construites avec le verbe *donner* (Ex. 96) par rapport à la fréquence d'emploi dans le corpus natif (38%).

Exemple d'UP à valeur cognitive (organisation conceptuelle, langagière) :

(96) Les verbes conditionnel sont aussi importantes dans ce type de texte. Pour **donner** *|*Dét zéro*| exemple, « si l'ours se dresse? » ou « si vous n'êtes pas dans un véhicule? » (FL2-AN, <TAN>, occ. 371)

L'explication de cette différence d'emploi résiderait dans une surgénéralisation de l'emploi cognitif du verbe *donner*, liée à un mécanisme de transfert sémantique par les apprenants du verbe *give* pour lequel ce type d'emploi prédicatif à valeur cognitive serait plus courant en anglais L1 qu'en français L1. Cette hypothèse liée au caractère polysémique des verbes à haute fréquence (cf. les hypothèses explicatives émises en V.1.1.1., pp. 231-232) reste néanmoins à vérifier dans des recherches ultérieures.

Nous observons également une tendance au sous-emploi portant sur les UP événementielles qui se traduit par 6,4% seulement d’emplois dans le corpus anglophone pour 16,3% d’UP événementielles en L1.

	Prendre		Donner	
	FL2-AN	FL1	FL2-AN	FL1
N prédicatif: état	33 29,2%	51 26,2%	40 32,0%	42 45,7%
N prédicatif: événement	10 8,8%	15 7,7%	8 (-) 6,4%	15 16,3%
N prédicatif: activité mentale	70 61,9%	129 66,2%	77 (+) 61,6%	35 38,0%
Total	113 100,0%	195 100,0%	125 100,0%	92 100,0%
Chi carré	$\chi^2 = 0.558$, ddl = 2, p > 0.5 (NS)		$\chi^2 = 13.216$, ddl = 2, p < *0.01	

Tableau 70. Distribution des catégories sémantiques prédictives des UP (sans les actions physiques) dans le corpus anglophone (vs natif) en fonction du verbe utilisé

Deuxièmement, l’usage des apprenants néerlandophones semble refléter de manière assez fidèle les emplois sémantiques des verbes *prendre* et *donner*. Si l’on observe le Tableau (71) ci-dessous, aucune différence ne semble en effet émerger de la comparaison entre le corpus néerlandophone et le corpus natif, ce qui est confirmé par les résultats statistiques non significatifs (p > 0.1 (NS) et p > 0.05 (NS)).

	Prendre		Donner	
	FL2-NL	FL1	FL2-NL	FL1
N prédicatif: état	18 34,0%	24,81 [51] 26,2%	17 25,0%	20,43 [42] 45,7%
N prédicatif: événement	8 15,1%	7,3 [15] 7,7%	18 26,5%	7,3 [15] 16,3%
N prédicatif: activité mentale	27 50,9%	62,76 [129] 66,2%	33 48,5%	17,03 [35] 38,0%
Total	53 100,0%	94,86 [195] 100,0%	68 100,0%	44,76 [92] 100,0%
Chi carré	$\chi^2 = 3.812$, ddl = 2, p > 0.1 (NS)		$\chi^2 = 5.376$, ddl = 2, p > 0.05 (NS) χ^2 de Yates = 4.155, p > 0.1 (NS)	

Tableau 71. Distribution des catégories sémantiques prédictives des UP (sans les actions physiques) dans le corpus néerlandophone (vs natif) en fonction du verbe utilisé

En conclusion, comme nous l’avons signalé à plusieurs reprises dans ce travail, les résultats obtenus doivent être nuancés en prenant en considération la nature du type de texte en cause. C’est cette perspective que nous aborderons sous V.3.2., dans les dernières pages de ce chapitre.

V.3. Medium langagier et types de textes

A côté des variables syntaxiques et sémantiques, que nous avons examinées dans les pages précédentes, nous avons pris en compte quelques informations extralinguistiques dont nous disposons. Nous avons ainsi intégré dans notre liste paramétrique les variables suivantes : le numéro de l'occurrence étudiée (*NumOcc*) et le fragment textuel auquel il renvoie (*Frag*), la nature de la langue source de l'apprenant (*TexteLS*), l'identité codée du texte (*TexteCode*) et la taille de celui-ci (*TexteTaille*).

Nom de la variable (<i>Name</i>)	Variable (<i>Label</i>)	Valeurs (<i>Values</i>)
NumOcc - Frag	Numéro de l'occurrence et fragment textuel	<xy>
TexteLS	Langue source	1. Anglais 2. Néerlandais 3. Français
TexteCode	Code du texte	<xy> (Ex. : <TAN001>)
TexteTaille	Taille du texte (<i>n</i> mots)	<xy>

Tableau 72. Extrait de la liste paramétrique – Variables extralinguistiques

Nous développons ici les paramètres concernant le type de texte et le medium langagier, en testant leur hypothétique influence sur la production d'UP par les apprenants du français L2 (cf. Hypothèse 3).

Hypothèse 3

Il y aurait une influence (a) du medium langagier et (b) du type de textes sur la fréquence et le type d'UP utilisées par l'apprenant avancé d'une L2.

a) On observerait un emploi fréquent, par les apprenants, d'UP provenant du code oral, ce qui donnerait lieu à un écart stylistique par rapport à l'usage conventionnel en L1.

b) Les types de textes nécessitant un style plus impersonnel (par ex. les textes réflexifs) comprendraient moins d'UP que les textes plus subjectifs (par ex. les textes argumentatifs).

Nous examinerons successivement la première partie de l'hypothèse précitée, portant sur l'influence, dans la production d'UP, du code oral (V.3.1.), puis la seconde partie de cette hypothèse, portant sur l'influence des types de textes (V.3.2.).

V.3.1. Code oral

Pour étudier dans quelle mesure il y aurait influence du code oral sur les UP étudiées, nous nous sommes appuyée sur l'information donnée par la fréquence relative d'occurrences observées dans le corpus de français oral *Valibel* (cf. variable *FqRelValibel*).

Dans un premier temps, nous avons sélectionné via le logiciel *SPSS* toutes les UP qui étaient attestées en français oral avec une fréquence relativement élevée, i.e. avec une fréquence relative minimale supérieure à 10 occurrences par million de mots¹⁰⁴. Nous avons ainsi ciblé notre objet d'étude, en ne gardant que les UP susceptibles d'apparaître couramment à l'oral. Notons que ces UP n'appartiennent cependant pas exclusivement au code oral, et peuvent tout aussi bien avoir une fréquence élevée d'occurrences dans le corpus de presse écrite (cf. Ex. 189).

(189) Lire un roman demande un certain effort et une grande attention. C'est aussi une activité qui peut prendre beaucoup de temps. (FL2-NL, <TNL070>, occ. 938) [*Prendre Dét temps (Modifieur)*: FqR en français oral L1 = 30,27 / FqR en français écrit L1 = 20,30]

Nous supposons, sur la base de l'Hypothèse 3 citée plus haut, que ces UP qui sont courantes en français oral auraient tendance à être employées plus fréquemment dans les écrits académiques par les LNN que par les LN. Or, étonnamment, nous observons que ces UP courantes à l'oral (155 occ.) sont utilisées par les LN dans une plus grande proportion que par les LNN. Sur l'ensemble de ces UP attestées fréquemment en français oral, 43,8% sont le fait de LN, pour 37,5% qui sont le fait de LNN néerlandophones et 18,7% de LNN anglophones.¹⁰⁵ Cette distribution contredit donc l'Hypothèse 3, du moins en ce qui concerne la fréquence des emplois issus de l'oral par les apprenants : les LN utilisent davantage d'UP se caractérisant par une fréquence élevée dans l'usage parlé que les LNN.

Dans un second temps, pour préciser notre étude, nous avons sélectionné les UP qui se caractérisaient par leur fréquence relative élevée dans l'usage parlé L1 (comme précédemment), mais qui se caractérisaient aussi par leur faible fréquence dans l'usage natif écrit. Cette seconde sélection nous a permis de focaliser notre attention sur les UP plus spécifiques au code oral (92 occ.), comme c'est le cas notamment de l'exemple (172) que nous présentons ici. Cet exemple se caractérise en effet par une faible fréquence dans l'usage écrit L1 (FqR = 1,72), représenté par le corpus de presse écrite du *Soir*, mais par une haute fréquence dans l'usage oral L1 (FqR = 15), représenté par le corpus oral *Valibel*.

(172) En effet, voir d'autres personnes agir comme nous nous donne l'impression qu'elles partagent nos désirs, nos aspirations, nos angoisses, nous donne le sentiment d'être nous-même compris, ce qui nous sécurise. (FL1, <TFRB199>, occ. 868)

¹⁰⁴ La moyenne des fréquences relatives attribuées aux UP du corpus de contrôle, qui sont attestées en français parlé, nous a servi de référence pour déterminer le caractère plus ou moins fréquent des UP dans l'usage (FqR moyenne à l'oral = 13,9651).

¹⁰⁵ Rappelons que ces proportions ont été obtenues après avoir calculé la fréquence relative d'occurrences pour 90.000 mots dans chaque corpus, afin de pouvoir rendre les résultats comparables.

Nous avons ainsi étudié l'influence de la fréquence des UP dans l'usage oral L1 sur la production d'UP par les LNN vs LN. Or, les résultats ont montré qu'il n'y avait pas de différence significative entre les différents corpus (test χ^2 de Kruskal-Wallis = 2.690, ddl = 2, $p > 0.2$ (NS)), du point de vue de l'indice de fréquence relative dans l'usage L1 attribué aux UP relevant spécifiquement de l'oral. La tendance semble néanmoins aller dans le sens des observations émises plus haut, puisque ce sont les LN qui montrent le rang moyen de fréquence des UP dans l'usage oral le plus élevé (49,73). Par contre, le rang moyen intermédiaire (47,54) concerne les LNN anglophones et le rang moyen le plus bas (39,70) est attribué aux LNN néerlandophones. En conclusion, malgré cette dernière observation, l'impact du code oral sur la fréquence des UP produites ne s'est pas avéré significatif : les LN et les LNN ont recours, les uns comme les autres, à des UP spécifiques au code oral dans leurs productions écrites, ce qui contredit en partie l'Hypothèse 3 formulée plus haut.

Dans un troisième temps, nous avons tenté de vérifier si le recours à des UP spécifiques au code oral pouvait donner lieu à des déviations par rapport à l'usage conventionnel en L1. Cet examen tend à répondre à la seconde partie de l'Hypothèse 3.a. selon laquelle l'intrusion d'UP spécifiquement orales dans les textes écrits produirait une déviation stylistique (i.e. une déviation *quantitative*). En examinant la distribution des UP spécifiques à l'oral (92 occ.) en fonction des types de déviations, nous n'avons extrait que 7 UP déviantes de l'ensemble des corpus L2 et L1. Sur ces 7 UP déviantes, 5 sont produites par des LNN néerlandophones, parmi lesquelles 3 déviations quantitatives (Ex. 190) et 2 déviations formelles (Ex. 83).

(190) La Belgique a changé beaucoup les dernières années. Non seulement c'est devenu un état fédéral, on a aussi modifié des institutions. On peut 'donner l'exemple du Parlement. Depuis la naissance de la Belgique, il y avait un système bicaméral avec deux chambres plus ou moins équivalent. Aujourd'hui il n'y a plus d'équilibre. (FL2-NL, <TNL004>, occ. 1057)

→ Reconstruction native : On peut **prendre** l'exemple du Parlement

(83) Est-ce bien un cas où la violence est la seule façon d'assurer la justice sociale ? Il n'est pas facile d'y 'donner une réponse. La loi est très stricte dans le cas d'un meurtre : personne n'a le droit de le faire et s'il le fait, il devra être pénalisé. (FL2-NL, <TNL091>, occ. 1117)

→ Reconstruction native : Il n'est pas facile d'y **apporter** une réponse

Soulignons que ces résultats n'ont pas pu être généralisés, vu le nombre beaucoup trop restreint d'occurrences observées. L'Hypothèse 3 reste donc en suspens. La voie est néanmoins ouverte à d'éventuelles recherches qui voudraient approfondir cette approche, en termes de fréquence d'emploi liée au médium langagier, dans une perspective

phraséologique.¹⁰⁶ Il serait ainsi intéressant d'étudier l'influence de facteurs, tels que le degré de *xénité* des apprenants d'une langue étrangère (cf. II.1.2., pp. 72-73), sur la production de déviances phraséologiques quantitatives et/ou contextuelles, en relation avec l'intrusion dans l'interlangue de formes spécifiques à l'usage oral. Dans le cas des corpus étudiés ici, nous pourrions formuler l'hypothèse d'un recours plus fréquent aux emplois propres à l'oralité par les LNN néerlandophones par opposition aux LNN anglophones. Les néerlandophones, comparés aux anglophones, se caractérisent en effet par un degré de *xénité* moindre vis-à-vis de la langue française. Rappelons que les LNN néerlandophones de nos corpus sont belges, ce qui implique une proximité notamment géographique, sociopolitique et culturelle importante par rapport à la langue cible, ce qui n'est pas le cas des apprenants anglophones des corpus étudiés.

Pour conclure, nous mettons le doigt sur un problème méthodologique qui a résulté de notre méthode d'investigation. Nous avons opposé ici les UP spécifiques au code oral à celles relevant du code écrit, en conférant donc implicitement au corpus de français écrit investigué (i.e. le corpus du *Soir* 1997) le statut de corpus de référence pour notre analyse de l'influence du medium langagier. Bien que représentatif de la langue courante, ce corpus de presse ne permet pas de tenir compte des spécificités des types de textes écrits qui constituent les corpus étudiés, à savoir des textes académiques à visée expositive. La spécificité de certaines UP relève en effet parfois plus du type de texte qui les contient que du medium langagier susceptible ou non de favoriser leur emploi¹⁰⁷. Certains cas, qui sont caractérisés par une faible fréquence dans les écrits de presse, mais une haute fréquence dans le corpus oral, semblent en fait être spécifiques à l'écrit académique. Leur emploi par les apprenants, dans les textes académiques qui constituent nos corpus, n'est donc absolument pas le reflet d'une intrusion du code oral dans l'écrit. Bien plus, ces cas marquent la spécificité du type de texte académique par opposition à d'autres types de productions écrites (ici, la presse écrite). Certaines UP, qui ont pu, de cette façon, être considérées comme spécifiques à l'oral, sont en fait des UP tout à fait

¹⁰⁶ Nous profitons de cette mise en perspective pour rappeler l'intérêt que représentent les études sur les effets de fréquence, menées entre autres par N. Ellis (2002, 2006), pour les études en acquisition de la compétence phraséologique en L2.

¹⁰⁷ Cette objection pourrait s'appliquer à notre étude collocationnelle et à notre analyse de déviance, qui reposent entre autres sur l'examen des UP en français L1 à partir de corpus de presse (*Le Soir* et *Les Voisins De Le Monde*). Nous répondrons en disant que, dans notre analyse collocationnelle et de déviance, nous avons combiné expressément plusieurs types de ressources pour le français L1 dans le but d'éviter cet écueil et de couvrir de la manière la plus large possible le champ phraséologique dans l'usage natif. Rappelons en outre que notre corpus de contrôle est, quant à lui, constitué uniquement de textes académiques et que c'est à partir de ce corpus comparable que nous avons tiré des conclusions quant à la compétence phraséologique des apprenants non natifs.

courantes dans les textes académiques, mais peu usuelles dans la presse écrite. L'exemple (191) nous semble ainsi représentatif de ce type d'emplois, à la fois courants à l'oral et spécifiques aux écrits expositifs, mais peu usuels en presse écrite.

(191) **Prenons l'exemple de scientifiques ou de penseurs émettant des idées novatrices.** (FL1, <TFRB055>, occ. 651) [*Prendre Dét exemple (Modifieur)* : FqR en français oral L1 = 55/ FqR en français écrit L1 = 1,57]

Nous devons dès lors relativiser les résultats présentés plus haut : la haute fréquence d'emplois spécifiques à l'oral, par opposition à une faible fréquence d'emplois dans les écrits de presse, n'est pas interprétable uniquement en termes de medium langagier, mais aussi en termes de type de textes. Ceci nous amène au point V.3.2., qui porte sur le rôle que jouent les types de textes dans la production phraséologique d'apprenants du français L2.

V.3.2. Types de textes

Les informations concernant le type de texte académique où apparaissent les données phraséologiques ont été quantifiées, afin de tester l'hypothèse qui a été formulée à ce propos (cf. Hypothèse 3) et de mettre en relief certaines observations tirées des analyses effectuées précédemment.

Hypothèse 3

Il y aurait une influence [(a) du medium langagier] et (b) du type de textes sur la fréquence et le type d'UP utilisées par l'apprenant avancé d'une L2.

(a) On observerait un emploi fréquent, par les apprenants, d'UP provenant du code oral, ce qui donnerait lieu à un écart stylistique par rapport à l'usage conventionnel en L1.]

b) Les types de textes nécessitant un style plus impersonnel (par ex. les textes réflexifs) comprendraient moins d'UP que les textes plus subjectifs (par ex. les textes argumentatifs).

Rappelons tout d'abord brièvement la définition que nous avons donnée des textes *académiques* en III.5. (pp. 161-163). Ces textes sont le plus souvent écrits dans un style formel et impersonnel qui découle de leur fonction de transmission d'informations. Ils sont habituellement produits dans un environnement institutionnel (*académique*). La plupart des textes L2 et L1 étudiés sont en outre des textes *expositifs*, se caractérisant par une structure organisationnelle *hiérarchisée* autour de la notion de topique. Nous avons distingué six types de *compositions académiques* auxquels nous avons attribué une valeur dans la liste paramétrique (cf. Tableau 73, ci-dessous) : les textes comparatifs, persuasifs, réflexifs, narratifs, de synthèse et argumentatifs.¹⁰⁸

¹⁰⁸ Nous renvoyons le lecteur à la définition de ces catégories sous III.5. (pp. 161-163).

Nom de la variable (Name)	Variable (Label)	Valeurs (Values)
TexteType	Types de texte académique	1. Comparatif – Analyse de textes littéraires 2. Persuasif – Discours politique 3. et 4. Réflexif – Histoire de la francophonie/ Economie 5. Narratif 6. Synthèse – Résumé de presse 7. Argumentatif – Dissertation

Tableau 73. Extrait de la liste paramétrique – Variables extralinguistiques : types de texte académique

Pour une description détaillée des types de textes académiques au sein des corpus respectifs, nous renvoyons le lecteur à la description qui en est faite en III.5. (pp. 162-163). Rappelons néanmoins que le corpus L1 est composé exclusivement de dissertations, alors que le corpus L2 néerlandophone est constitué principalement de dissertations (76%) et secondairement de textes de synthèse sur des sujets de presse écrite (24%). Le corpus d'apprenants anglophones montre, à l'inverse des deux autres corpus, une grande diversité de types de textes de laquelle émanent deux types principaux, à savoir les types argumentatif (1/3 des textes) et réflexif (1/3 des textes). L'ensemble des trois corpus (anglophone, néerlandophone et francophone) inclut une nette majorité de textes argumentatifs (73,9% des textes).

V.3.2.1. Distribution des UP

Considérant la répartition des UP au sein des corpus (cf. IV.2.2.), nous avons constaté que le corpus anglophone comptait 69,4% d'emplois phraséologiques (*vs* séquences libres), alors que le corpus néerlandophone en comptait 81,1 % et le corpus natif 85,1%. La différence de fréquence entre les corpus, qui s'est révélée significative, a ainsi mis en exergue la prédominance des UP construites avec les verbes *prendre* et *donner* dans les corpus L2 et L1, par opposition à l'emploi de séquences libres (SL). En outre, aucun suremploi ou sous-emploi des UP n'avait alors été mis au jour dans les corpus non natifs, puisque la répartition des UP entre les trois corpus est équivalente.

Après avoir examiné la distribution des UP dans les corpus en fonction des types de textes, nous nuancions ces premiers résultats.

Au sein du corpus anglophone, nous constatons que la proportion d'UP relatives aux textes de type argumentatif, qui représentent 1/3 de ce corpus, est en fait de 80,4%. Ce pourcentage est donc très proche de la proportion d'UP dans le corpus de contrôle en L1 (85,1%), qui est composé exclusivement de textes argumentatifs, contrairement à la proportion globale d'UP que nous avons observée pour l'ensemble du corpus académique anglophone (69,4%). La différence significative que nous avons mise au jour

entre les corpus anglophone et natif doit donc être relativisée.¹⁰⁹ La proportion des UP au sein des types de textes, autres qu'argumentatifs, semble en effet influencer les résultats portant sur l'ensemble des textes académiques. Les textes persuasifs et narratifs qui, bien que représentant une partie moindre du corpus anglophone (environ 10% des textes sont persuasifs et environ 20% sont narratifs), semblent infléchir la tendance vers le bas, puisqu'ils présentent une proportion équivalente d'UP et de SL. C'est aussi le cas des textes réflexifs, qui constituent 1/3 du corpus anglophone, et présentent une proportion d'UP élevée (66,3%), mais inférieure à la proportion d'UP chez les natifs. Seuls les textes comparatifs (moins de 5% des textes) présentent une proportion d'emplois phraséologiques importante (86,7%), mais leur impact sur la fréquence globale d'emplois phraséologiques semble être moindre au vu de leur nombre restreint. Selon ces observations, ce sont principalement les textes comparatifs et argumentatifs qui favoriseraient les emplois phraséologiques, et secondairement les textes réflexifs, par opposition aux textes persuasifs et narratifs qui présentent une proportion équivalente de séquences libres et phraséologiques.

Au sein du corpus néerlandophone, les textes argumentatifs montrent également une proportion d'UP très proche de celle des natifs, avec 81,8% d'UP pour 18,2% de SL. Mais les UP au sein des textes de synthèse sont elles aussi très présentes (76,2% des SVN). L'influence du type de textes sur la production d'UP, par opposition aux SL, semblerait ici moins importante.

En conclusion, nous répondons à l'Hypothèse 3 selon laquelle « les types de textes nécessitant un style plus impersonnel comprendraient moins d'UP que les textes plus subjectifs », en partant du principe qu'il y aurait bien une influence du type de texte sur la proportion d'UP au sein des textes. Les textes argumentatifs et comparatifs (*a priori* plus subjectifs) se caractérisent ainsi par un nombre très important d'UP, par comparaison aux textes narratifs et réflexifs moins « phraséologiques » (*a priori* plus impersonnels). Notons toutefois que la proportion d'UP est moins élevée dans les textes persuasifs, alors qu'ils sont le lieu de la subjectivité par excellence, et que les textes de synthèse *a priori* moins subjectifs, comprennent une part importante d'UP.

¹⁰⁹ Notons que, malgré les observations qui seront faites ici à propos de l'influence des types de textes sur la production d'UP, nous n'avons pas procédé dans notre étude au réexamen *complet* des résultats obtenus précédemment pour deux raisons : (i) les différents types de textes académiques des corpus non natifs ne sont pas comparables entre eux du point de vue de leur *taille* respective ; (ii) nous ne disposons pas actuellement de textes comparables, en termes de *type* textuel, en français L1. Par contre, les trois corpus étant en partie composés de textes argumentatifs, une étude contrastive sera menée sur ce type de textes académiques en particulier, dans le Chapitre VI de ce travail.

Ces observations, qui ne feront pas l'objet d'une étude approfondie dans ce travail, ouvrent des pistes de recherche intégrant la perspective phraséologique à la perspective de description linguistique des types textuels.

Nous tenons à souligner que la distribution des UP déviantes au sein des types de textes n'entre pas en contradiction avec les résultats obtenus par le biais de l'analyse de déviance effectuée dans le Chapitre IV. Ainsi, dans la quasi-totalité des types de texte composant les corpus non natifs, la proportion d'UP déviantes reflète les résultats globaux qui ont été mis en évidence précédemment (66,3% d'UP non déviantes dans le corpus anglophone et 72,3% dans le corpus néerlandophone). Seuls les textes persuasifs, qui ne constituent qu'1/10 du corpus anglophone, montrent une proportion inverse d'UP non déviantes, avec 28,6% d'UP non déviantes contre 71,4% d'UP déviantes produites par les LNN anglophones. Les autres types de textes ont une proportion d'UP non déviantes supérieure à 66%. Le corpus narratif compte même très peu d'UP déviantes, avec seulement 8,3% d'UP déviantes sur l'ensemble des UP qui en ont été extraites.

Nous nous intéressons dans le point suivant à la relation entre les types de textes et les traits syntaxiques/ sémantiques qui caractérisent les UP des corpus étudiés.

V.3.2.2. Profil syntaxique et sémantique des UP

Nous examinerons successivement l'impact des types de textes sur la structuration syntaxique du syntagme nominal (SN) constitutif des UP et sur les emplois sémantiques de ces UP.

Premièrement, la structure du SN des UP a été étudiée sous V.1.2. Nous avons alors constaté une différence statistiquement très significative ($p = *0.000$), du point de vue de la structure syntaxique du constituant nominal des UP, entre les productions des apprenants anglophones et celles des francophones natifs, ainsi qu'une différence un peu moins significative ($p < *0.05$) entre les productions des néerlandophones et celles des natifs.

Les résultats concernant les anglophones dans leur manière de structurer les SN dans les UP s'était avéré significatif (cf. V.1.2.) pour l'ensemble des textes académiques. Nous présentons donc ici les résultats relatifs au corpus anglophone (*vs* natif) qui ont fait l'objet d'un réexamen en fonction des types textuels, sur la base du Tableau (57) des pages précédentes en y incluant les données spécifiques aux textes argumentatifs (colonne 2 du Tableau 74, ci-dessous).

	FL2-AN Tout	FL2-AN Argumentatif	FL1
Dét zéro + N + (Modifieur/ <i>de N/ P Sub</i>)	24 (-) 8,9%	8 (-) 9,3%	54 18,2%
Dét + <i>N de N</i>	26 (-) 9,6%	8 (-) 9,3 %	48 16,2%
Dét + N + P Sub	39 (++) 14,4%	9 (+) 10,5%	25 8,4%
Dét Poss + N + (Modifieur)	12 4,4%	3 3,5%	15 5,1%
Prép + SN	22 (-) 8,1%	14 16,3%	47 15,9%
Dét + N + (Modifieur)	147 (++) 54,4%	44 (+) 51,2%	107 36,1%
Total	270 100,0%	86 100,0%	296 100,0%

Tableau 74. Distribution des structures du SN des UP dans les textes argumentatifs anglophones (fréquence absolue), par comparaison avec l'ensemble du corpus anglophone et le corpus natif

L'examen plus approfondi de la répartition des UP dans les types de textes du corpus anglophone montre que les résultats portant sur les textes argumentatifs, par comparaison avec les résultats globaux obtenus sous V.1.2., restent relativement similaires, avec généralement une réduction de l'écart entre la fréquence d'emploi des UP par les LNN et les LN. Malgré ce rapprochement avec la compétence phraséologique native, du point de vue de la structure du SN, les structures avec un déterminant et un modifieur quelconques semblent être suremployées dans les textes anglophones strictement argumentatifs (Ex. 192) (51,2% des UP), comme dans l'ensemble des textes académiques (54,4%).

(192) un gouvernement, démocratique, qui as le courage de **prendre** les mesures necessaire pour que l'economie se reprends. (FL2-AN, <TAN008>, occ. 56)

Les structures avec un déterminant « zéro » (Ex. 142) et en *N de N* (Ex. 193) semblent quant à elles faire l'objet d'un sous-emploi dans les textes argumentatifs, comme dans l'ensemble des textes académiques étudiés.

(142) je veux dire qu'il **a prit** connaissance de ses emotions et peut donc les controler (FL2-AN, <TAN166>, occ. 133)

(193) Il est possible de prévoir que les dirigeants européens aborderaient l'évaluation minutieuse de l'euro sous l'angle économique. Riché **donne** un aperçu de la signification et de l'objet de la monnaie. D après Riché, les quinze Etats-membres veulent que l'Europe soit émancipé de la [«] domination du dollar ». (FL2-AN, <TAN146>, occ. 532)

Par contraste, la catégorie des SN prépositionnels (Ex. 194) et celle des structures avec proposition subordonnée complément du nom (Ex. 195), pour lesquelles nous avons noté respectivement un sous-emploi (8,1% en L2 pour 15,9% en L1) et un suremploi (14,4% en L2 pour 8,4% en L1) dans l'ensemble des textes anglophones, font l'objet dans notre réexamen d'un emploi proche de celui qu'en font les natifs, lorsqu'on ne considère que les textes argumentatifs.

(194) Le succès de l'euro n'est pas quelque chose que nous pouvons **prendre à la légère**. (FL2-AN, <TAN162>, occ. 254)

(195) En conclusion, nous savons que qu'il y a plusieurs différences qui existent entre les sexes. A mon avis il faut les supprimer à tel point qu'on **donne** à chacun la possibilité de choisir sa propre vie selon ses capacités. (FL2-AN, <TAN130>, occ. 400)

Concernant les LNN néerlandophones, le réexamen des résultats en fonction des types textuels a montré que la différence entre les néerlandophones et les natifs restait sensiblement la même quand on ne prenait en considération que les textes de type argumentatif, avec une légère réduction de l'écart entre les LNN et les LN. La répartition des UP dans les textes néerlandophones strictement argumentatifs, en fonction de la structure du SN, est quasiment inchangée par rapport à la répartition globale dans les textes académiques. Le suremploi des structures avec un déterminant quelconque (Ex. 196), ainsi que le sous-emploi des structures avec un déterminant « zéro » (Ex. 31) et des structures prépositionnelles (Ex. 8), sont moins marqués quand on ne considère que les textes argumentatifs.

(196) " Political solutions don't work ". Avec cette déclaration, Woody Allen résume en quelques mots la nouvelle tendance qui apparaît à la fin de ce siècle : la désertion du politique. (...) Pourtant, il est indispensable que le politique **prend un nouvel essor**. Sinon, l'anarchie totale ou la dictature nous attend. (FL2-NL, <TNL092>, occ. 944)

(31) L'État peut-il décider de la vie et de la mort de ses sujets ? Pourrait-il jouer le rôle de Dieu ? D'après vous, l'Etat devrait **prendre parti** contre les actes criminels et inhumains. S'il tue pour assurer la justice, il perd de vue son rôle d'exemple. Comment lutter contre le terrorisme, en pratiquant l'assassinat ? (FL2-NL, <TNL174>, occ.1021)

(8) Si les partisans de la peine de mort crient à haute voix que la prison à vie est plus cruelle que la peine de mort, ils sont pris à leur propre piège. Le criminel, à qui on voudrait infliger la peine de mort, mérite le châtement le plus grand. La peine à perpétuité se présente alors comme le châtement suprême. (FL2-NL, <TNL172>, occ. 1020)

Seul le suremploi des structures nominales avec proposition subordonnée (Ex. 197) semble plus marqué dans les textes strictement argumentatifs, comparé à la fréquence d'emploi dans l'ensemble des textes académiques.

(197) Les livres qu'on achète n'importe où par contre ne méritent, d'après moi, même pas d'être comptés parmi la littérature. Ils ne sont en effet que prduits pour l'homme de la rue et ne traitent qu'une histoire banale, souvent à l'aide d'un simple schéma afin d'offrir au commun des mortels un peu de distraction et de lui **donner l'impression qu'il n'est pas le seul au monde qui ne puisse surmonter ses problèmes**. (FL2-NL, <TNL105>, occ. 1090)

Cette réduction de l'écart entre non-natifs et natifs dans les textes argumentatifs plaide en faveur d'une différence qui ne serait plus statistiquement significative pour ce type de textes, puisque le coefficient de contingence, concernant l'ensemble des textes

académiques, se situait déjà très légèrement sous le seuil de significativité ($p = *0.042$).¹¹⁰ De manière générale, nous pouvons conclure que les LNN néerlandophones semblent se rapprocher des natifs dans leur manière de structurer le SN des UP, en particulier du point de vue des textes argumentatifs, par comparaison avec l'ensemble des textes académiques pour lesquels la différence entre les productions non natives et natives s'est avérée significative.

En résumé, sur le plan de la structure syntaxique du SN, la différence entre les LNN et les LN semble moins importante quand on ne compare entre eux que les types de textes argumentatifs des corpus L2 et L1. Dans les textes argumentatifs comme dans les textes académiques, les LNN néerlandophones tendent néanmoins, d'une part, à sous-employer les structures avec déterminant « zéro » et les structures prépositionnelles et, d'autre part, à suremployer les structures nominales avec proposition subordonnée et celles avec un déterminant quelconque. Les LNN anglophones tendent, pour leur part, à sous-employer les structures avec déterminant « zéro » et en *N de N*, ainsi qu'à suremployer les structures avec un déterminant quelconque, dans l'ensemble des textes académiques comme dans les textes strictement argumentatifs.

Deuxièmement, nous avons proposé un profil sémantique des UP sous V.2.2. Nous avons alors soulevé la prédominance des emplois d'activité mentale (i.e. cognitifs) au sein des emplois prédicatifs, en suggérant que cette prédominance était liée au type de textes académiques, majoritairement expositifs, qui constituent les corpus L2 et L1. Cette hypothèse s'est vu confirmée par l'examen de la répartition des catégories sémantiques des UP au sein des différents types de textes dans les corpus, qui montrent presque tous une prédilection pour les emplois phraséologiques dans cette catégorie sémantique. Deux types de textes se démarquent néanmoins de l'ensemble du corpus anglophone: les textes comparatifs montrent une nette majorité d'emplois prédicatifs cognitifs (92,3% des UP) ; les UP dans les textes narratifs sont majoritairement des emplois prédicatifs d'état (66,7% des UP) et secondairement des emplois cognitifs (33,3%). Malgré ces différences, nous constatons que les résultats révisés en termes de type textuel ne semblent pas contredire les résultats globaux portant sur l'ensemble du corpus (cf. Tableau 75, ci-dessous).

¹¹⁰ Rappelons que nous ne pouvons pas vérifier, dans le cadre de ce travail, si ces tendances révisées sont significatives, principalement en raison du nombre d'occurrences observées qui est insuffisant pour pouvoir effectuer des tests statistiques valides.

	FL2-AN	FL2-AN Argumentatif	FL1
N prédicatif: état	73 30,7%	73 30,0%	93 32,4%
N prédicatif: événement	18 7,6%	4 (-) 5,0%	30 10,5%
N prédicatif: activité mentale	147 61,8%	52 (+) 65,0%	164 57,1%
Total	238 100,0%	238 100,0%	287 100,0%

Tableau 75. Distribution des UP prédicatives (sans les actions physiques) dans les textes argumentatifs anglophones (fréquence absolue), par comparaison avec l'ensemble du corpus anglophone et le corpus natif

La tendance (non significative) des anglophones à employer davantage d'UP d'activité mentale (Ex. 198) que les natifs semble être toujours présente, tandis qu'une tendance à utiliser moins de prédicats d'événement (Ex. 157) semble émerger des résultats révisés.

(198) Pour que les français arrivent à accepter l'euro, il y a aura besoin de projets de sorte qu'il s'intègre dans la culture et dans les traditions du pays. Quelque régions **ont** déjà **pris** l'initiative, comme Sarlat en Périgord, où ils ont lancé le double étiquetage francs-euros des produits de terroir tels que les truffes. (FL2-AN, <TAN141>, occ. 228)

(157) La peur de perdre de l'argent par les erreurs de conversion est un soupçonnable valable. Cette difficulté d'estimer les prix nouveaux, donnera lieu à une sorte de xenophobie. (FL2-AN, <TAN154>, occ. 536)

A l'inverse des résultats portant sur le corpus anglophone, nous avons obtenu des résultats significatifs pour le corpus néerlandophone, qui semblent se confirmer quand on ne considère que les textes de type argumentatif. Le suremploi des UP d'événement (Ex. 189) (22,7% d'UP pour 10,5% en L1) et le sous-emploi de prédicats cognitifs (Ex. 199) (47,3% d'UP pour 57,1% en L1) sont en effet confirmés, de manière légèrement plus marquée par rapport aux résultats antérieurs.

(189) Lire un roman demande un certain effort et une grande attention. C'est aussi une activité qui peut **prendre** beaucoup de temps. (FL2-NL, <TNL070>, occ. 938)

(199) A ce point de ce texte, il faut réfuter et compléter la description de la vie privée de Gilles Lipovetsky. En disant que la vie privée consiste en 'veiller à sa santé, préserver sa situation matérielle (...) et attendre les vacances' il en **donne** expres une description très pauvre. (FL2-NL, <TNL094>, occ. 1081)

	FL2-NL	FL2-NL Argumentatif	FL1
N prédicatif: état	35 28,9%	33 30,0%	93 32,4%
N prédicatif: événement	26 (+) 21,5%	25 (+) 22,7%	30 10,5%
N prédicatif: activité mentale	60 (-) 49,6%	52 (-) 47,3%	164 57,1%
Total	121 100,0%	238 100,0%	287 100,0%

Tableau 76. Distribution des UP prédicatives (sans les actions physiques) dans les textes argumentatifs néerlandophones (fréquence absolue), par comparaison avec l'ensemble du corpus néerlandophone et le corpus natif

Ceci confirme la thèse du suremploi des UP à valeur événementielle et du sous-emploi des UP à valeur cognitive par les LNN néerlandophones, tant du point de vue de l'ensemble du corpus que des textes spécifiquement argumentatifs qui le constituent.

V.3.2.3. UP avec *prendre* vs *donner*

Nous avons mis en évidence sous V.1.1.1. un sous-emploi des UP construites avec le verbe *prendre* et un suremploi des UP construites avec le verbe *donner*, dans les deux corpus de français L2 par opposition à la fréquence d'emploi des deux verbes en L1 (cf. Tableau 77, ci-dessous). Ceux-ci, bien qu'étant des verbes à haute fréquence, se différencient ainsi l'un de l'autre dans les textes académiques produits par les LNN.

	FL2-AN	FL2-NL	FL1	Total
Prendre	65,68 [135] 29,45% (-) 50,0%	62 27,79% (-) 45,3%	95,35 [196] 42,75% 66,2%	223,03 [393] 100% 55,9%
Donner	65,68 [135] 34,7% (+) 50,0%	75 39,6% (+) 54,7%	48,65 [100] 25,7% 33,8%	189,33 [310] 100% 44,1%
Total	131,36 [270] 100,0%	137 100,0%	144 [296] 100,0%	412,36 [703] 100,0%

Tableau 77. Distribution des UP dans les corpus non natifs et natifs, en fonction du verbe à haute fréquence

Nous avons procédé, comme dans les points précédents, à un examen de l'impact des types de textes sur la production d'UP par les apprenants du français L2, en comparant les UP construites avec *prendre* vs *donner*, en particulier dans le corpus anglophone qui est le plus hétérogène en termes de types de textes. Nous avons obtenu les résultats que nous présentons dans les tableaux ci-dessous.

	FL2-AN		
	Prendre	Donner	Total
Comparatif	0 0,0%	26 100,0%	26 100,0%
Persuasif	6 42,9%	8 57,1%	14 100,0%
Réflexif	56 46,7%	64 53,3%	120 100,0%
Narratif	21 87,5%	3 12,5%	24 100,0%
Argumentatif	52 60,5%	34 39,5%	86 100,0%
Total	135 (-) 50,0%	135 (+) 50,0%	270 100,0%

Tableau 78. Distribution des UP avec prendre vs donner dans le corpus anglophone, en fonction du type de texte

	FL2-NL		
	Prendre	Donner	Total
Synthèse	10 62,5%	6 37,5%	16 100,0%
Argumentatif	52 (-) 43,0%	69 (+) 57,0%	121 100,0%
Total	62 (-) 45,3%	75 (+) 54,7%	137 100,0%

Tableau 79. Distribution des UP avec prendre vs donner dans le corpus néerlandophone, en fonction du type de texte

Au vu de ces résultats, notre hypothèse du suremploi de *donner* vs *prendre* est remise en cause pour les apprenants anglophones, si l'on considère uniquement les textes académiques de type argumentatif. Alors que ces apprenants montrent un suremploi significatif des UP construites avec *donner* par opposition à *prendre* dans l'ensemble du corpus, on remarque la tendance inverse dans les textes anglophones spécifiquement argumentatifs: 60,5% des UP y sont construites avec le verbe *prendre* et 39,5% avec *donner*, ce qui est relativement similaire aux proportions du corpus natif comprenant 66,2% d'UP construites avec *prendre* et 33,8% avec *donner*. Par contre, les emplois phraséologiques des verbes dans le corpus néerlandophone restent similaires selon que l'on considère le corpus dans son entièreté (45,3% d'UP avec *prendre* pour 54,7% avec *donner*) ou uniquement les textes argumentatifs (43% d'UP avec *prendre* pour 57% avec *donner*). Le suremploi et le sous-emploi par les néerlandophones sont donc confirmés pour les textes argumentatifs, ce qui n'est pas le cas des anglophones qui tendent, dans les textes strictement argumentatifs, à se rapprocher de la compétence native en termes de fréquence d'emploi des verbes dans les UP.

Dans le cas du corpus anglophone, nous pensons que les résultats globaux (50% d'UP avec *prendre*) sont principalement biaisés par la présence des textes comparatifs, qui ne contiennent que des UP construites avec le verbe *donner* (26 occ.), mais aucune avec le verbe *prendre*, ainsi que par la fréquence légèrement plus élevée des UP construites avec *donner* dans les textes réflexifs et les textes persuasifs. Les occurrences plus fréquentes d'UP avec *donner* dans ces types de textes ont donc donné lieu à une proportion globale plus élevée d'UP avec *donner* dans le corpus non natif (50%) que dans le corpus natif (33,8%). Ce suremploi qui est observable dans une perspective globale

d'analyse de textes dits *académiques*, est donc remis en cause par un examen des types de textes en particulier. C'est ainsi que l'hypothèse explicative de l'influence du caractère irrégulier ou régulier du système flexionnel des verbes doit être elle aussi révisée : il semblerait finalement que ce paramètre ne soit pas crucial, puisque le verbe *prendre* fait l'objet d'un emploi plus fréquent que le verbe *donner* dans les textes argumentatifs d'apprenants anglophones. L'hypothèse d'un impact minime, mais non décisif, de ce facteur n'est cependant pas totalement à exclure. Nous avons en effet constaté que les apprenants néerlandophones utilisaient malgré tout plus d'UP avec *donner* qu'avec *prendre* et que, parmi les cinq types de textes du corpus anglophone, les trois types textuels comparatif, persuasif et réflexif montraient un emploi plus fréquent des UP avec *donner* qu'avec *prendre* (cf. Tableaux 78 et 79, ci-dessus).

Pour conclure, nous insistons sur le fait que, suite à ces nuances apportées par l'examen plus approfondi de l'influence des types de textes en cause sur la production d'UP, seule la généralisation des résultats relatifs au suremploi et au sous-emploi des UP dans les textes académiques semble avoir été affectée de manière cruciale. Tenant compte de l'impact non négligeable de ce facteur, nous effectuerons, dans le Chapitre VI, une analyse de complexité structurelle et de diversité lexicale des UP produites par les apprenants anglophones et néerlandophones, comparée à la compétence native, dans des textes strictement argumentatifs. Nous visons à mettre en évidence, de cette manière, certaines caractéristiques phraséologiques propres aux textes argumentatifs, qui pourront servir à l'élaboration d'une *norme pédagogique* (cf. la présentation de cette notion sous II.2.1.2.) exploitable dans une perspective didactique en L2. Nous reviendrons sur ces perspectives didactiques, envisageables à partir de nos résultats de recherche, dans le Chapitre VII de ce travail.

V.4. Conclusion intermédiaire

Sur le plan formel, nous avons tout d'abord, dans ce chapitre, comparé le comportement des deux verbes *prendre vs donner* qui constituent la base verbale des UP étudiées (703 occ.). Nous avons ainsi mis en évidence, sous V.1.1.1., un sous-emploi des UP construites avec le verbe *prendre* et un suremploi des UP construites avec *donner* dans les deux corpus non natifs, anglophone et néerlandophone [Confirmation de l'Hypothèse 2]. Ces résultats se sont ensuite vus nuancés par un examen plus approfondi de l'impact du type de textes sur la production d'UP par les apprenants (cf. V.3.2.). Nous avons ainsi

constaté que ces résultats, valables pour l'ensemble des textes académiques qui constituent le corpus anglophone, diffèrent quand on ne compare entre eux que les textes strictement argumentatifs. Les LNN anglophones y emploient davantage d'UP construites avec *prendre* qu'avec *donner*. Le suremploi des UP avec *donner* et le sous-emploi des UP avec *prendre* par les néerlandophones sont quant à eux confirmés pour l'ensemble des textes académiques comme pour les textes strictement argumentatifs, ce qui n'est pas le cas des anglophones qui tendent donc, dans les textes strictement argumentatifs, à se rapprocher de la compétence native en termes de fréquence d'emploi des verbes dans les UP.

Sur le plan syntaxique (cf. V.1.2.), nous avons observé une tendance significative des apprenants anglophones du français L2 à sous-employer certaines structures dans les unités phraséologiques (UP) qu'ils produisaient – i.e. les structures sans déterminant (i.e. avec déterminant « zéro »), les structures en *N de N* et les structures prépositionnelles –, et à en suremployer d'autres – i.e. les structures avec un déterminant quelconque, ainsi que les structures nominales avec proposition subordonnée. L'examen de la répartition des UP dans les types de textes du corpus anglophone a par ailleurs montré que cette différence entre LNN anglophones et LN était réduite quand on comparait la distribution des UP au sein des textes strictement argumentatifs. Cette légère influence du type de texte ne semble cependant pas contredire le suremploi et le sous-emploi constatés pour l'ensemble des textes académiques anglophones, sauf pour les structures prépositionnelles du syntagme nominal (SN) (notées *Prép + SN* dans la variable *SyntaxSNUP*) qui y sont sous-employées, mais qui ne le sont pas dans les textes strictement argumentatifs, et pour les structures avec proposition subordonnée complément du nom, qui sont suremployées dans l'ensemble du corpus anglophone, mais ont une fréquence d'emploi proche de celle des natifs dans les textes strictement argumentatifs. La différence d'emploi des UP par les néerlandophones en fonction de la structure syntaxique du constituant nominal s'est également avérée significative, mais dans une moindre mesure : nous avons noté, dans l'ensemble des textes académiques, d'une part, un suremploi des structures avec un déterminant quelconque et des structures avec une proposition subordonnée et, d'autre part, un sous-emploi des structures avec un déterminant « zéro » et des structures prépositionnelles. Ces différences semblent cependant moins marquées, voire non significatives, quand on prend en considération le facteur du type textuel, puisque dans les textes strictement argumentatifs, la fréquence d'emploi des structures du SN des UP par les LNN néerlandophones tend à se rapprocher fortement de celle des natifs.

Comparant les UP quand elles sont construites soit avec *prendre*, soit avec *donner*, nous avons constaté que les LN les utilisaient de manière significativement différente en termes de structuration syntaxique : les UP construites avec *prendre* comptent un nombre important de structures en *N de N* et de structures prépositionnelles (*Prép + SN*), alors que les UP avec *donner* présentent plutôt des structures avec proposition subordonnée complément du nom (*Dét + N + P Sub*). Nous avons néanmoins constaté qu'il y avait, pour les deux verbes, une majorité d'UP avec un déterminant quelconque (34,2% avec *prendre* et 40% avec *donner*) et environ 1/5 d'UP avec un déterminant « zéro » (17,9% avec *prendre* et 19% avec *donner*). Du point de vue des LNN anglophones, la différence entre les UP produites par les LNN et les LN est significative avec le verbe *prendre* mais pas avec *donner*, ce qui met en évidence que le sous-emploi et le suremploi observés précédemment pour l'ensemble des UP dans le corpus anglophone sont en fait dus à une différence de fréquence d'emploi portant sur les UP avec le verbe *prendre*. A l'inverse, l'emploi des UP avec *donner* par les anglophones, au niveau de la structuration syntaxique du SN, se rapproche de celui des LN. Malgré une différence non significative relative aux UP avec *donner* employées par les apprenants anglophones, nous avons constaté que ces derniers utilisaient davantage d'UP construites avec un déterminant quelconque par rapport aux natifs (55,6% pour 40% en L1), qu'ils n'utilisaient aucune structure prépositionnelle (pour 4% en L1) et produisaient moins d'UP construites avec un déterminant « zéro » (8,9% pour 19% en L1). Du point de vue des LNN néerlandophones, aucune différence significative entre les corpus néerlandophone et natif n'a pu être confirmée statistiquement concernant les emplois avec *prendre* vs *donner*. Il ressort néanmoins de nos observations que les LNN néerlandophones emploient, pour les deux verbes *prendre* et *donner*, davantage de structures avec un déterminant quelconque que les LN, mais moins de structures avec un déterminant zéro et de structures prépositionnelles. Par contre, nous avons mis au jour pour les néerlandophones une fréquence d'emploi des structures en *N de N* avec *prendre* supérieure à la fréquence d'emploi des natifs (24,2% pour 18,9% en L1), mais pas pour *donner* dont la fréquence d'emploi est proche en L2 et en L1 (13,3% pour 11% en L1).

Sur le plan sémantique (cf. V.2.2.), nous avons observé pour l'ensemble de nos données de corpus (906 SVN), tout comme pour l'ensemble de nos données phraséologiques (703 UP), une majorité d'emplois prédicatifs, avec principalement des activités mentales, puis des états et, enfin, des événements. Cette distribution des UP en catégories sémantiques, tenant compte de l'influence potentielle des types de textes sur la

production d'UP par les apprenants (cf. V.3.2.), a par ailleurs été confirmée par un examen de la répartition des catégories sémantiques des UP au sein des différents types de textes : les textes des trois corpus L2 et L1 manifestent tous une prédilection pour les emplois phraséologiques d'activité mentale, sauf les textes narratifs (représentant moins de 20% des textes du corpus anglophone) qui comptent majoritairement des emplois prédicatifs d'état. Ces résultats révisés en termes de type textuel ne contredisent donc pas les résultats globaux portant sur l'ensemble des corpus, du point de vue de la catégorisation sémantique des UP. Quelques tendances ont cependant été mises au jour : les anglophones emploient davantage d'UP d'activité mentale que les natifs, mais moins de prédicats d'événement que ces derniers (la différence n'est pourtant pas significative) ; les néerlandophones montrent une différence significative, que l'on prenne en considération l'ensemble du corpus ou seulement les textes argumentatifs, qui se traduit par un suremploi des UP d'événement et un sous-emploi d'emplois phraséologiques d'activité mentale.

Enfin, nous avons vu que le code langagier oral (*vs* écrit), à l'inverse du type de textes [Seconde partie de l'Hypothèse 3 partiellement confirmée], n'influait pas de manière importante la production d'UP par les apprenants [Première partie de l'Hypothèse 3 non confirmée]. En effet, les UP courantes en français oral ne semblent pas être plus fréquemment employées par les LNN que par le LN dans les écrits académiques (cf. V.3.1.). Premièrement, nous avons observé que les LN utilisaient davantage d'UP courantes dans l'usage parlé que les LNN. Deuxièmement, les résultats portant plus particulièrement sur les UP spécifiques au code oral (*vs* écrit) ont montré que les LN et les LNN ont recours, les uns comme les autres, à ces UP spécifiques au code oral dans leurs productions écrites, avec une plus haute fréquence de ces emplois par les LN, puis par les LNN anglophones et, enfin, par les LNN néerlandophones. Troisièmement, nous avons tenté de vérifier si le recours à des UP spécifiques au code oral pouvait donner lieu à des déviations (quantitatives) par rapport à l'usage conventionnel en L1. Cette partie de l'Hypothèse 3 n'a cependant pu être vérifiée au vu du nombre trop restreint (7 occ.) d'UP déviantes spécifiques à l'oral, extraites des corpus L2 et L1.

Nous adoptons, dans le chapitre VI, un nouvel angle d'approche de l'analyse sur corpus d'apprenants menée dans les pages précédentes. Tenant compte des différences entre les types de textes constitutifs des corpus L2, principalement du corpus anglophone, nous analyserons nos données en ne prenant pour objet d'étude que les UP produites dans des textes de type argumentatif. Les résultats obtenus dans les chapitres IV et V portant

sur une compétence phraséologique des UP construites avec *prendre* et *donner* dans des textes académiques, va donc se focaliser sur cette compétence, mais portant uniquement sur des textes argumentatifs. Nous verrons ainsi comment identifier la nature de l'écart entre les interlangues d'apprenants avancés du français L2 et la norme pédagogique visée de la langue cible, en termes de *complexité structurelle phraséologique* et de *diversité (poly)lexicale*, ainsi qu'en termes de *précision phraséologique*.

Chapitre VI: Résultats 3 – Précision et complexité

L'objectif de ce chapitre est d'élargir le champ d'application de l'étude descriptive menée précédemment en adoptant une perspective d'analyse en termes de *précision* (cf. VI.1.) et de *complexité* (cf. VI.2.), héritée des recherches en acquisition d'une L2 (cf. II.3.). Ayant mis au jour dans le chapitre précédent une influence du type de textes sur la production d'unités phraséologiques (UP) par les apprenants du français L2, les résultats de ces deux analyses porteront uniquement sur les textes argumentatifs présents dans les corpus L2 et L1 (cf. Tableau 80, ci-dessous). Cette restriction de notre objet d'étude vise à dégager des traits phraséologiques propres à ce type de textes. Les données étudiées représentent, dans les corpus argumentatifs L2 et L1, un total de 503 UP (cf. Tableau 80, ci-dessous).

	Textes argumentatifs		
	Français L2		Français L1
	LNN anglophones 85 textes	LNN néerlandophones 167 textes	LN francophones 330 textes
	FRIDA (CECL)	FRIDA (CECL) 45.000 mots	CaFLAM 132.000 mots
		Extension FRIDA 30.000 mots	CODIS (CECL) 53.000 mots
Total : 503 UP	55.000 mots 86 UP	75.000 mots 121 UP	185.000 mots 296 UP

Tableau 80. Analyse de complexité et de précision – Description des corpus de textes argumentatifs

Les calculs de *précision* et de *complexité* traditionnellement proposés pour ces mesures de compétence langagière seront adaptés à notre objet d'étude, i.e. la compétence phraséologique des apprenants avancés. Les hypothèses formulées à ce propos sont l'Hypothèse 4 (cf. VI.2.) et l'Hypothèse 1 (cf. VI.1.) reproduites plus bas. Nous aborderons par ailleurs la question du suremploi d'UP par les apprenants avancés d'une L2, qui est supposé aller de pair avec un lexique moins riche (i.e. moins *complexe*) (cf. Hypothèse 2), en dégagant des textes argumentatifs non natifs les *structures collocationnelles* (déviantes et non déviantes) les plus représentatives en termes de fréquence d'emploi et/ou de mésusage par rapport au corpus de contrôle en français L1 (cf. VI.2.3.).

- VI.1. Analyse de précision phraséologique
- VI.2. Analyse de complexité phraséologique
 - VI.2.1. Complexité structurelle
 - VI.2.2. Diversité lexicale
 - VI.2.3. Structures collocationnelles
- VI.3. Conclusion intermédiaire

VI.1. Analyse de précision phraséologique

Comme nous l'avons vu précédemment (cf. II.3.2., pp. 114-115), différents moyens permettent de calculer le taux de *précision* des productions d'apprenants. Ces mesures sont traditionnellement appliquées au niveau soit syntaxique, soit lexical. Nous les avons adaptées ici au phénomène lexico-grammatical, afin de tester l'hypothèse selon laquelle il y aurait un écart, mesurable en termes de *précision*, entre la compétence phraséologique des apprenants avancés et la compétence phraséologique native (cf. Hypothèse 4, ci-dessous). Pour rappel, la *précision* correspond au « degré de coïncidence du système de l'IL avec le système langagier du locuteur natif (la norme) » (Van Daele *et al.*, 2007: diapositive 4).

Hypothèse 4

La compétence phraséologique des apprenants avancés d'une L2 se caractériserait par un écart en terme de mésusage par rapport à l'usage natif, avec une proportion plus importante d'UP déviantes que chez les locuteurs natifs. Cet écart serait mesurable via un calcul de précision : les productions non natives se caractériseraient par un degré de précision moins élevé que les productions natives.

Premièrement, comme nous étudions un phénomène linguistique particulier, en l'occurrence les UP, nous avons procédé à une analyse en termes d'*occasions obligatoires*, plutôt qu'en termes de *propositions* non déviantes (*error-free clauses*) (cf. II.3.2., p. 114) : il s'agit de calculer le pourcentage d'UP non déviantes (*error-free PUs*), en divisant le nombre d'UP non déviantes par le nombre total d'UP, multiplié par 100.

Deuxièmement, nous avons calculé le nombre d'UP déviantes pour 1000 mots (*errors per 1000 words*), i.e. le nombre d'UP déviantes divisé par le nombre total de mots multiplié par 1000 (qui s'apparente aussi à la mesure traditionnellement choisie pour étudier la précision lexicale). Nous avons, pour ce calcul, opté pour un dénominateur plus grand que celui généralement adopté pour des phénomènes fréquents, en l'occurrence 1000 mots au lieu de 100 mots, puisque notre objet d'étude ne porte que sur les UP construites avec les deux verbes *prendre* et *donner*.

Troisièmement, nous avons intégré à ces mesures traditionnelles de *précision*, strictement quantitatives, la dimension qualitative de la *gravité* de la déviance, sur le modèle du calcul de précision proposé par Foster & Wigglesworth (2007) pour mesurer la compétence syntaxique, en *propositions* (en anglais *clauses*), des apprenants d'une L2. Ces chercheuses suggèrent un nouvel indice de précision qui permette de rendre compte du niveau de gravité d'une déviance dans le calcul de l'indice. La mesure acquiert ainsi une dimension qualitative, plus proche des conditions réelles de réception de la

production de l'apprenant en situation de communication. Foster & Wigglesworth distinguent trois niveaux de gravité de déviance et attribuent un indice à chaque niveau :

Instead of dividing these [clauses] simply into *error-free* or *not error-free*, we propose to score clauses which are not error free by three levels of error:

- No error: An accurately constructed clause
- Level one: the clause has minor errors (e.g. morphosyntactic) which do not obscure the intended meaning
- Level two: the clause has more serious errors (e.g. word choice or word order) which make the intended meaning harder to recover
- Level three: the clause has errors which make the intended meaning difficult to recover (Foster & Wigglesworth 2007: diapositives 17-22)

Elles attribuent le score 1.0 aux propositions non déviantes, 0.8 aux déviances de niveau 1, 0.5 au niveau 2 et 0.1 au niveau 3. Il s'agit ensuite d'obtenir un indice global de gravité des déviances par texte en divisant la somme des indices de gravité par le nombre total de *propositions (clauses)* par texte. Plus le score est élevé, plus l'apprenant s'exprimerait avec *précision*. Prenant les auteures au mot quand elles concluent leur présentation en affirmant que « the new measure (...) can be calibrated differently if desired » (Foster & Wigglesworth, 2007: diapositive 33), nous avons adapté cette mesure globale à notre objet d'étude: la somme totale des indices de gravité des UP déviantes est divisée par le nombre total de mots, puis multipliée par 1000 (*accuracy per 1000 words*). Rappelons en outre que nous ne nous sommes pas contentée d'adapter cette mesure au phénomène phraséologique, mais que nous l'avons précisée en proposant un calcul de l'indice de *gravité* des déviances qui tient compte de plusieurs facteurs : le *type* de déviance en cause, la *localisation* de cette déviance, le *nombre* de déviances par UP, le degré de *collocabilité* (en français L1) et la *fréquence* (en français L1) de l'UP déviante. Nous avons ainsi obtenu un degré de gravité global pour chaque UP étudiée (cf. variable *GravDév16Global*, sous IV.3.3.6., p. 222). À l'instar de Foster & Wigglesworth (2007), nous avons ensuite attribué aux UP un indice pour chaque niveau de gravité de déviance, sur la base de la catégorisation obtenue dans la variable *GravDév16Global* :

- UP non déviante (indice = 1.0)

(161) Et pourtant, il convient de **prendre** une distance intellectuelle par rapport à ces cris du « vox populi » et de se poser des questions, qui montreront que du point de vue éthique et moral, la peine de mort n'est guère défendable (FL2-NL, <TNL216>, occ. 1037)

- Niveau 1 de déviance (indice = 0.8) : degré faible de gravité

(128) Le plan national de passage à l'Euro regroupe toutes les décisions et recommandations **|prises|**, au niveau national, pour assurer dans les meilleures conditions de la monnaie unique en chaque pays (FL2-NL, <TAN142>, occ. 231)

→ Reconstruction native : *Le plan national de passage à l'Euro regroupe toutes les décisions prises et les recommandations faites*

- Niveau 2 de déviance (indice = 0.5) : degré moyen de gravité

(68) La littérature en général, a comme première tâche de rendre les lecteurs conscients de notre monde et de ses problèmes. (...) La première fonction de la littérature est en voie de disparition, et la deuxième, celle du divertissement, *|prend| le premier plan. (FL2-NL, <TNL103>, occ. 950)

→ Reconstruction native : *La première fonction de la littérature est en voie de disparition, et la deuxième, celle du divertissement, se place au/ revient au/ passe au/ occupe le premier plan*

- Niveau 3 de déviance (indice = 0.1) : degré élevé de gravité

(200) Les égyptologues de leur part, ne veulent pas accepter cette théorie, qui semble vraisemblable, parce qu'ils n'admettent pas d'autres chercheurs dans le domaine où ils sont les autorités, et en plus, en intégrant cette nouvelle théorie, ils devraient admettre qu'ils ont, en **donnant** *|leur| explication plutôt |lénigmatique|, fait une erreur de 8500 ans, ce qui leur donnait une perte de visage. (FL2-NL, <TNL120>, occ. 1101)

→ Reconstruction native : *ils devraient admettre qu'ils ont, en **donnant** une explication plutôt fumeuse/vague, fait une erreur de 8500 ans*

Après avoir effectué cette transposition des résultats obtenus précédemment en indices de gravité, nous avons divisé la somme totale de ces indices de gravité par le nombre total de mots, pour multiplier ensuite le résultat par 1000 (précision « améliorée » pour 1000 mots). Rappelons que l'avantage principal de cette méthode est d'intégrer à la mesure quantitative de *précision* la dimension qualitative de *gravité* de la déviance (cf. II.2.2.4. et IV.3.3.). Le Tableau (81) (ci-dessous) présente les résultats auxquels nous sommes arrivés en appliquant à nos données phraséologiques chacune des trois mesures de précision expliquées plus haut : le pourcentage d'UP non déviantes, la proportion d'UP déviantes pour 1000 mots et l'indice de précision « amélioré » pour 1000 mots (précision quantitative et qualitative).

Mesures de précision	FL2-AN	FL2-NL	FL1
% UP non déviantes	(61/86) * 100 70,93%	(89/121) * 100 73,55%	(276/296) * 100 93,24%
UP déviantes pour 1000 mots	25/55 0.45	32/75 0.43	20/185 0.11
UP non déviantes	61	89	276
UP déviante Niveau 1	7	3	7
UP déviante Niveau 2	17	28	12
UP déviante Niveau 3	1	1	1
Indice UP non déviantes (n x 1.0)	61.0	89.0	276.0
Indice Niveau 1 (n x 0.8)	5.6	2.4	5.6
Indice Niveau 2 (n x 0.5)	8.5	14.0	6.0
Indice Niveau 3 (n x 0.1)	0.1	0.1	0.1
Total des indices bruts	75.2	105.5	287.7
Divisé par le nombre de mots * 1000	/55	/75	/185
Précision « améliorée » pour 1000 mots	1.37	1.41	1.56

Tableau 81. Analyse de précision phraséologique dans les textes argumentatifs non natifs et natifs

Les trois mesures traduisent une compétence phraséologique plus proche de la compétence native, en termes de précision, pour les LNN néerlandophones que pour les LNN anglophones. Par ailleurs, la mesure de précision « améliorée » de Foster &

Wigglesworth (2007) ne semble pas jouer ici un rôle crucial, puisque les résultats vont dans le même sens que les autres mesures de précision. Néanmoins, la comparaison des résultats obtenus par ce calcul de précision « améliorée » avec les résultats obtenus à partir du pourcentage d'UP déviantes et non déviantes, mais cette fois sans tenir compte du degré de gravité, a donné lieu à des résultats plus nuancés (cf. Tableau 82, ci-dessous).

Total d'UP non déviantes	61	89	276
Divisé par le nombre de mots * 1000	/55	/75	/185
Précision pour 1000 mots	1.11	1.87	1.49
Précision « améliorée » pour 1000 mots	1.37	1.41	1.56

Tableau 82. Comparaison de la mesure de précision pour 1000 mots et de la mesure de précision « améliorée » pour 1000 mots, pour les textes argumentatifs non natifs et natifs

Alors que la compétence phraséologique s'améliore pour les LNN anglophones (1.11 → 1.37) et les LNN (1.49 → 1.56), elle diminue pour les LNN néerlandophones (1.87 → 1.41). Ce constat laisse à penser que les LNN néerlandophones ont une meilleure compétence phraséologique générale que les LNN anglophones, mais que ces derniers commettent plus d'UP légèrement déviantes que les néerlandophones qui produisent des déviances phraséologiques plus graves. Ces résultats, relatifs aux textes académiques de type argumentatif, sont similaires à ceux que nous avons obtenus pour l'ensemble des textes académiques (cf. Tableau 42, en IV.3.3.6, p. 223), qui révélaient une proportion plus importante d'UP moyennement déviantes chez les LNN néerlandophones (86,8% des UP déviantes) que chez les LNN anglophones (74,7% des UP déviantes), allant de pair avec une proportion d'UP légèrement déviantes plus importante chez les LNN anglophones (18,7%) que chez les néerlandophones (10,5%).¹¹¹ L'influence du degré de gravité sur la mesure de précision ne paraît pourtant pas suffisamment importante pour inverser les niveaux de compétence phraséologique entre les deux groupes d'apprenants non natifs. Cette absence d'une réelle influence du degré de gravité sur le taux global de précision réside probablement dans le fait que, dans la variété avancée du français L2 que nous étudions, la proportion d'UP déviantes est minime comparée à la proportion d'UP non déviantes.

Après avoir mesuré la compétence phraséologique des LNN avancés du français L2 par le biais de mesures de *précision*, nous abordons dans le point suivant la

¹¹¹ Dans l'ensemble des textes académiques, les UP gravement déviantes se sont révélées être très rares dans les deux corpus non natifs, avec 1 occ. (sur 38 UP déviantes) dans le corpus néerlandophone et 6 occ. (sur 91 UP déviantes) dans le corpus anglophone. Dans l'ensemble des textes argumentatifs, les UP gravement déviantes concernent une seule occ. dans chaque corpus (L2 et L1). L'impact de celles-ci sur le taux global de précision est donc minime quand on mesure le taux de précision « améliorée » des productions non natives.

compétence phraséologique de ces apprenants non natifs par le biais de mesures de *complexité*, adaptées au phénomène phraséologique, à savoir la *complexité structurelle* (cf. VI.2.1.) et la *diversité lexicale* (cf. VI.2.2.) des UP.

VI.2. Analyse de complexité phraséologique

La *complexité* langagière peut être envisagée soit du point de vue de la *complexité linguistique* interne au système de la langue cible (LC) (i.e. *linguistic complexity*), soit du point de vue de la *difficulté* éprouvée par l'apprenant en situation de production langagière (*psycholinguistic complexity*).

Complexity in language can be assessed on different levels. Language can be complex in and by itself (linguistic complexity), or language can be difficult to process by a language user (psycholinguistic complexity; see Pienemann, 1998). When complexity is considered as a dimension of productive language proficiency, it can be measured by looking at the amount of *complex structures* produced (grammatical or syntactic complexity), or at the amount of *rare/different words* produced (lexical complexity or richness). (Bulté, 2007: 34)

Par ailleurs, comme le note Bulté (2007), la *complexité* est souvent assimilée, de manière réductrice, à celle de *difficulté*. Elle est aussi parfois définie comme caractérisant ce qui est acquis tardivement¹¹² ou ce qui est moins fréquent en langue. Pour notre part, nous considérons que le phénomène phraséologique, qui porte dans notre étude sur des séquences construites avec des verbes à haute fréquence, se caractérise à la fois par sa *complexité linguistique* et sa *complexité psycholinguistique*. Premièrement, en termes de *complexité linguistique*, nous avons vu que le figement était en soi un phénomène *complexe*, puisque les frontières entre les différentes catégories d'unités phraséologiques (UP) (quasi-figées, semi-figées et figées) semblent résister à toute délimitation univoque et que des contraintes relevant des domaines à la fois syntaxique, sémantique et lexical, semblent peser simultanément, mais pas de manière systématique, sur ces UP (cf. la *polyfactorialité* et la *scalarité* du figement, en I.1., p. 22, en I.1.4.2., pp. 46-47 et en I.1.4.4., p. 55). Deuxièmement, les UP étudiées ici appartiennent presque exclusivement à la sphère du semi-figement (voire du quasi-figement) : elles se situent entre le champ de la combinatoire libre et celui de la phraséologie, partageant des traits (syntaxiques, sémantiques et lexicaux) issus de l'un et l'autre domaine. Troisièmement, les verbes à

¹¹² Le raisonnement est alors circulaire, puisque la complexité, quand elle est définie sur la base du critère de l'*élaboration* de l'IL (en anglais *elaborated language*), est par définition considérée comme un trait tardif en acquisition d'une L2.

haute fréquence que nous étudions sont des verbes hautement polysémiques, autrement dit, sémantiquement complexes. Cette *complexité linguistique* rend le phénomène phraséologique particulièrement *difficile* à appréhender pour le chercheur ou l'enseignant (Bardovi-Harlig, 2002) et *difficile* à comprendre et/ou à produire par l'apprenant avancé d'une L2 (cf. III.2.1., p. 144) en situation de production langagière (*complexité psycholinguistique*).

Rappelons en outre que nous avons opté pour la définition générale qu'Ellis & Barkhuizen (2005: 139) ont donnée de la *complexité*, à savoir « the extent to which learners produce elaborated language » (cf. II.3.3). Selon cette définition, nous émettons l'hypothèse que la production d'UP dans une interlangue (IL) est l'indice d'un degré d'élaboration élevé de la production langagière de l'apprenant (cf. Hypothèse 1, ci-dessous).

Hypothèse 1

Le phénomène phraséologique engendrerait des difficultés pour les apprenants d'une L2 jusqu'à un niveau avancé d'acquisition. Ces difficultés seraient liées entre autres à la complexité linguistique qui caractérise les UP semi-figées, puisque celles-ci se rapprochent de la combinatoire libre tout en possédant déjà des caractéristiques phraséologiques (contraintes syntaxiques, sémantiques et/ou lexicales). L'écart entre la compétence phraséologique interlangagière des apprenants avancés et la compétence phraséologique native serait ainsi mesurable en termes de complexité structurelle et lexicale.

VI.2.1. Complexité structurelle

Afin d'élaborer une mesure de la complexité structurelle des IL qui soit applicable au phénomène phraséologique, nous nous sommes appuyée sur l'analyse structurelle du syntagme nominal (SN) des unités phraséologiques (UP) opérée en V.1.2. Notre hypothèse de départ est que la complexité structurelle des UP, pour un apprenant du français L2, se traduit notamment par les contraintes pesant sur la structure du SN de l'UP (cf. V.1.2. et V.3.2.2.), en l'occurrence par des contraintes sur le déterminant (« zéro » ou possessif) (Ex. 201 et 202) ou sur la structure complexe du SN, i.e en *N de N* (Ex. 203), avec proposition subordonnée complément du nom (*N + P Sub*) (Ex. 204) ou avec préposition (*Prép + SN*) (Ex. 205).

Exemple d'UP sur laquelle pèse une contrainte sur le déterminant « zéro » :

(201) Elles peuvent s'enrichir davantage mutuellement et **donner naissance** à des biens communs très intéressants. (FL1, <TFRB312>, occ. 798)

Exemple d'UP sur laquelle pèse une contrainte sur le déterminant possessif :

(202) Toutefois il lui faut également s'en affranchir dans une certaine mesure pour **prendre son envol**. (FL1, <TFRB159>, occ. 712)

Exemple d'UP sur laquelle pèse une contrainte sur la structure en *N de N* :

(203) En effet, si quelqu'un se distingue en mal, le commun des mortels le rejettera, et s'il se fait remarquer par ses qualités, il l'admira. **Prenons** l'exemple d'un étudiant qui doit rédiger un texte argumentatif. (FL1, <TFRB015>, occ. 622)

Exemple d'UP sur laquelle pèse une contrainte sur la structure propositionnelle complément du nom :

(204) Mais un adolescent qui se rebellerait entièrement contre les normes qui lui sont « imposées » peut petit-à-petit se réfugier dans des artifices comme la drogue, ce qui ne lui **donnerait** par ailleurs seulement l'impression d'être différent. (FL1, <TFRB161>, occ. 860)

Exemple d'UP sur laquelle pèse une contrainte sur le syntagme prépositionnel :

(205) Aujourd'hui le touriste d'agence est totalement **pris en charge** et passe souvent quinze jours de farniente. (FL1, <TFRB257>, occ. 763)

Par contraste, la catégorie de structuration du SN avec un déterminant quelconque (Ex. 206) a été considérée comme potentiellement plus flexible que les structures précitées.¹¹³

Exemple d'UP avec déterminant quelconque :

(206) (cela implique que des mesures adéquates soient prises en vue de poursuivre, dans le sens du progrès, l'harmonisation des conditions existantes dans les différents Etats membres) (FL1, <TFRB294>, occ. 588)

Pour mesurer le degré de complexité structurelle des UP, nous avons comparé les fréquences relatives (pour 55.000 mots) des structures potentiellement contraintes (vs les structures avec un déterminant quelconque) entre les trois corpus de textes argumentatifs (L2 et L1). Nous avons ainsi constaté que les anglophones produisaient moins d'UP complexes (48,8%) que les néerlandophones (53,7%), qui en produisaient eux-mêmes moins que les natifs (63,9%).

	FL2-AN 55.000 mots	FL2-NL 75.000 mots	FL1 185.000 mots
UP complexes	42 48,8%	47,66 [65] 53,7%	56,19 [189] 63,9%
UP avec dét. quelconque	44 51,2%	41,06 [56] 46,3%	31,81 [107] 36,1%
Total	86 100%	88,73 [121] 100%	88 [296] 100%

Tableau 83. Analyse de complexité structurelle phraséologique – Pourcentage d'UP à structure potentiellement contrainte (vs flexible) dans les textes argumentatifs non natifs et natifs

Les UP produites par les LNN sont donc structurellement moins complexes, i.e. moins contraintes, que celles produites par le LN dans des textes académiques de type argumentatif. Autrement dit, les LNN semblent, d'une part, ne pas avoir encore atteint le niveau de compétence phraséologique native en termes de complexité de la structuration des UP et, d'autre part, les LNN néerlandophones semblent faire preuve d'une meilleure

¹¹³ Nous rappelons que le degré de figement d'une UP repose sur des contraintes linguistiques relevant (non exclusivement) de la syntaxe, de la sémantique et/ou du lexique, et que ces contraintes sont souvent corrélées. Le principe selon lequel les UP avec un déterminant quelconque ont plus de chances d'être flexibles que des UP construites, entre autres, avec un déterminant « zéro » ou en *N de N*, ne remet donc aucunement en cause l'appartenance des UP avec un déterminant quelconque au champ phraséologique.

compétence phraséologique que les LNN anglophones à cet égard. Ces résultats sont par ailleurs confirmés quand on examine la proportion des structures nominales potentiellement plus contraintes dans les UP construites avec le verbe *prendre*, par comparaison avec les structures incluant le verbe *donner* (cf. Tableau 84, ci-dessous) : les UP construites avec *prendre* et avec *donner* sont plus complexes structurellement en français L1 (avec respectivement 65,8% de structures contraintes pour *prendre* et 60% pour *donner*) qu'en français L2. Les LNN anglophones semblent avoir une moins bonne compétence phraséologique que les LNN néerlandophones sur le plan de la complexité structurelle des UP, puisqu'ils utilisent une proportion moins importante de structures complexes du SN pour les deux verbes.

	Prendre			Donner		
	FL2-AN 55.000 mots	FL2-NL 75.000 mots	FL1 185.000 mots	FL2-AN 55.000 mots	FL2-NL 75.000 mots	FL1 185.000 mots
UP complexes	27 51,9%	22 [30] 57,7%	38,35 [129] 65,8%	15 44,1%	25,66 [35] 51,7%	17,84 [60] 60%
UP avec dét. quelconque	25 48,1%	16,13 [22] 42,3%	19,92 [67] 34,2%	19 55,9%	24,93 [34] 49,3%	11,89 [40] 40%
Total	52 100%	38,13 [52] 100%	58,27 [196] 100%	34 100%	50,6 [69] 100%	29,73 [100] 100%

Tableau 84. Analyse de complexité structurelle pour les UP avec *prendre* vs *donner* – Pourcentage d'UP à structure potentiellement contrainte (vs flexible) dans les textes argumentatifs non natifs et natifs

Examinant le Tableau (84) de plus près, nous constatons que les résultats montrent une tendance des UP construites avec *prendre*, dans les corpus non natifs et natifs, à présenter une structure syntaxique du constituant nominal plus complexe, par rapport aux UP construites avec *donner* qui comptent un pourcentage moins élevé d'UP potentiellement plus complexes structurellement en français L2 et L1.

VI.2.2. Diversité lexicale

La *complexité lexicale* des UP construites avec des verbes à haute fréquence réside dans la capacité de ces verbes fortement polysémiques à engendrer un grand nombre de combinaisons lexico-grammaticales. Notons que, selon la définition des UP comme étant des unités polylexicales dont la combinatoire lexicale est contrainte, plus une UP serait *diversifiée* ou *riche* lexicalement, plus elle aurait tendance à se rapprocher d'une combinatoire libre (vs la restriction lexicale sur les possibilités de substitution paradigmatisque qui caractérise les UP). Autrement dit encore, plus une UP est *complexe lexicale*ment, moins elle est *figée* : le degré de *complexité lexicale* d'une UP est inversement proportionnel à son degré de *figement lexical*. Il s'agira donc ici de mesurer

et de comparer la *diversité lexicale* au sein des UP construites avec les verbes *prendre* et *donner* en L2 et en L1. Pour ce faire, nous avons adapté au phénomène phraséologique le calcul de l'indice du *type-token ratio* (TTR) revu par Guiraud (IG). Pour rappel, cet indice de *diversité polylexicale* est obtenu en divisant la somme d'UP *types* (*prendre/donner* + N *types*) par la racine carrée du nombre d'UP produites dans chaque corpus (*prendre/donner* + N *occurrences*) (cf. II.3.3., p. 116). Un écart entre un indice supérieur en L1 et inférieur en L2 indique que les UP produites par les LNN sont moins diversifiées lexicalement que celles produites par les LN. D'après les résultats obtenus¹¹⁴, le verbe *donner* se caractérise en L1 par une plus grande diversité de combinaisons polylexicales (IG = 5.5), que le verbe *prendre* (IG = 3.93). Ce constat semble aller dans le sens d'une plus grande polysémie du verbe *donner*, par opposition au verbe *prendre*, en français L1. Nous avons par ailleurs observé un écart entre la compétence lexicale phraséologique des LNN et celle des LN (cf. Tableau 85, ci-dessous): les résultats montrent un indice de diversité polylexicale inférieur chez les LNN anglophones (IG = 4.29) et néerlandophones (IG = 4.57) par rapport aux natifs (IG = 5.5) pour le verbe *donner*, ainsi que chez les anglophones pour le verbe *prendre* (IG = 3.74) par rapport aux natifs (IG = 3.93). Par contre, il semblerait que les UP soient plus diversifiées lexicalement chez les LNN néerlandophones (IG = 4.44) que chez les natifs (IG = 3.93) pour le verbe *prendre*.

	FL2-AN 55.000 mots	FL2-NL 75.000 mots	FL1 185.000 mots
Prendre	IG = 3.74	IG = 4.44	IG = 3.93
Donner	IG = 4.29	IG = 4.57	IG = 5.5

Tableau 85. Analyse de diversité polylexicale – Indice de Guiraud (IG) pour les UP construites avec *prendre* et *donner* dans les textes argumentatifs des corpus non natifs et natifs

Une mesure de la diversité polylexicale des UP en anglais argumentatif L1 et en néerlandais L1 pourrait nous indiquer si l'hypothèse du transfert de la langue source (LS) vers la langue cible (LC) du potentiel combinatoire des verbes à haute fréquence entre ici en jeu. Nous pouvons en effet supposer, sur la base de ces résultats, que le verbe *nemen* est plus complexe lexicalement, i.e. accepte un plus grand nombre de combinaisons avec des lexèmes nominaux *types* différents, que le verbe *prendre* en français argumentatif L1, ce qui donnerait lieu à un transfert, par le LNN néerlandophone, du potentiel combinatoire des UP construites avec *nemen* dans la LS, sur les UP qu'il produit avec le verbe *prendre* dans son interlangue.

Pour conclure, nous rappelons que nous avons formulé l'hypothèse (cf. Hypothèse 2, ci-dessous) qu'un lexique phraséologique moins riche chez les LNN irait de pair avec

¹¹⁴ Notons qu'en effectuant ce type de mesure, nous incluons les UP déviantes dans les résultats.

un suremploi par les LNN d'UP *types*, que nous avons dénommées des *doudous polylexicaux* (cf. en anglais les concepts de *lexical* et *phrasal teddy bears*, en II.4.2.3., pp. 130-131).

Hypothèse 2

La compétence phraséologique des apprenants d'une L2, concernant les UP construites avec des verbes à haute fréquence, se caractériserait paradoxalement par un écart en termes de (a) suremploi et/ou de (b) sous-emploi par rapport à l'usage natif.

a) Les apprenants avancés non natifs auraient tendance à recourir à des *doudous polylexicaux*, ce qui se traduirait à la fois par un suremploi de certaines UP *types* et par une diminution de la richesse lexicale des productions phraséologiques des apprenants en L2.

[b) Les constructions délexicales verbo-nominales seraient sous-employées par les apprenants avancés.]

C'est ce que nous nous proposons d'étudier dans le point suivant, en procédant à une analyse de *structures collocationnelles* reposant sur la co-sélection de structures et d'unités lexicales particulières (cf. la définition des *combinaisons syntagmatiques récurrentes*, en I.2.1.).

VI.2.3. Structures collocationnelles

Selon la première partie de l'Hypothèse 2 (cf. *supra*), la diversité polylexicale moins importante des productions non natives irait de pair avec un suremploi de certaines structures lexico-grammaticales, servant de *doudous polylexicaux* aux apprenants avancés du français L2. Tenant compte de la difficulté liée aux possibilités combinatoires (structurelles et lexicales) des UP, on peut en effet raisonnablement supposer que le LNN va soit suremployer un certain type d'UP (*doudous polylexicaux*), soit éviter de produire des UP (sous-emploi).¹¹⁵ Nous avons abordé ici le phénomène de suremploi/ sous-emploi des UP en adoptant une perspective *lexico-grammaticale*, i.e. en examinant la manière dont les unités *lexicales* étudiées entraînent en combinaison avec certaines *structures collocationnelles*. La fréquence peu élevée d'occurrences d'une même UP *type* nous a par ailleurs poussée à formuler nos résultats, non pas en termes de suremploi/ sous-emploi (ce

¹¹⁵ Notre hypothèse est ici que les écarts de fréquence observés peuvent être le résultat d'une prise de conscience par le LNN de la complexité linguistique propre au phénomène phraséologique, qui tend alors à contourner cet obstacle communicationnel soit en l'évitant (sous-emploi), soit en ayant davantage recours à des UP dont il maîtrise l'emploi (suremploi de *doudous polylexicaux*). Nous tenons à souligner que le suremploi et le sous-emploi des UP peuvent également être le fruit de mécanismes de transfert (positif ou négatif) qui sont activés, consciemment ou non, par le LNN en situation de production langagière. Il serait intéressant par la suite de mettre en rapport ces deux causes potentielles de suremploi/sous-emploi par l'apprenant avancé d'une L2, qui interagissent souvent simultanément en situation langagière (cf. II.1.4.2., pp. 85-86). Rappelons aussi que le recours à ces mécanismes est influencé par des facteurs individuels, tels que la motivation (intégrative *vs* instrumentale, cf. II.1.4.2., p. 89 et II.4.2.3., pp. 128-129) ou l'attitude de l'apprenant (cf. la notion de *risk taker*, en II.4.2.3., pp. 127-128).

qui impliquerait que l'écart de fréquence entre les LNN et les LN puisse être vérifié statistiquement), mais en termes d'écart de fréquence plus ou moins marqué entre la fréquence relative des occurrences d'UP *types* en L2 et en L1.

Dans un premier temps, nous avons examiné la fréquence relative (FqR) des UP (incluant les UP déviantes) dans les trois corpus, en ne retenant que les UP *types* dont la fréquence absolue (FqA) était de minimum 3 occurrences dans au moins un des trois corpus. Nous mettons en évidence ci-dessous les résultats obtenus, en ne présentant que les cas les plus représentatifs regroupés par structure syntaxique. Les structures syntaxiques concernées (ou *patrons*), qui se sont révélées les plus intéressantes de ce point de vue, sont les structures que nous avons considérées comme étant plus *complexes* (cf. *supra*) et pour lesquelles nous avons constaté un suremploi ou un sous-emploi par les LNN, dans les textes argumentatifs (cf. V.3.2.2.): (i) les structures avec déterminant « zéro », qui sont sous-employées par les LNN et en particulier par les anglophones quand elles sont construites avec le verbe *prendre* ; (ii) les structures en *N de N* avec le verbe *prendre*, qui sont sous-employées par les LNN anglophones et suremployées par les LNN néerlandophones ; (iii) les structures avec une proposition subordonnée complément du nom, qui sont suremployées par les LNN néerlandophones ; (iv) les structures prépositionnelles, qui sont sous-employées par les LNN néerlandophones. Pour chaque catégorie structurelle, nous mentionnons d'abord les UP qui semblent faire l'objet d'un évitement par les LNN, puis celles qui semblent être utilisées plus fréquemment en L2 qu'en L1 et, enfin, dans certains cas, les UP congruentes en L2 et L1, i.e. dont la fréquence d'emploi est similaire en L1 et en L2.

- Structures avec déterminant « zéro » :

- UP qui sont utilisées plus fréquemment en français L1, mais sont absentes en L2 : *prendre conscience de* (Ex. 16) (FqR = 4,46), *donner envie* (Ex. 207) (FqR = 0,89), *prendre exemple sur* (Ex. 208) (FqR = 1,19).

(16) La famille est un concept qui traverse les âges et dont chaque génération **prend conscience** très tôt dans sa vie (FL1, <TFRB280>, occ. 782)

(207) Chaque être humain est unique et c' est ce caractère unique qui nous distingue des autres et qui justement fait notre valeur et **donne envie** aux autres de s' intéresser à nous. (FL1, <TFRB128>, occ. 851)

(208) Un autre élément non négligeable est celui de la mode. La plupart des gens préfèrent s'habiller tendance, et avoir un physique de rêve, **prennant exemple** sur les mannequins. (FL1, <TFRB065>, occ. 656)

- UP qui sont utilisées plus fréquemment en français L2, mais moins ou pas du tout en français L1 : *donner lieu* (Ex. 209) (FqR en FL1 = 0,89 ; FL2-AN = 1 ; **FL2-NL = 2,93**), *donner naissance* (Ex. 201) (FqR en FL1 = 1,78 ; **FL2-AN = 4** ;

FL2-NL = 1,46). Notons que les UP construites avec le verbe *donner* et le nom *naissance* incluent ici les cas de prédicat d'événement/ achèvement (Ex. 201), ainsi que les cas de prédicats d'activité physique au sens de 'mettre au monde', 'accoucher' (Ex. 210). Une distinction entre les structures collocationnelles appartenant à l'une ou l'autre catégorie sémantique est opérée dans les pages qui suivent.

(209) Un meurtre reste donc toujours un meurtre. Un acte inhumain qui ne peut-être justifié par la loi. C'est une loi qui ne sert pas pour faire fonctionner le système judiciaire et qui **donnerait lieu** à des abus. La peine de mort est donc un mal à éviter. (FL2-NL, <TNL190>, occ. 1209)

(201) Elles peuvent s'enrichir davantage mutuellement et **donner naissance** à des biens communs très intéressants. (FL1, <TFRB312>, occ. 798)

(210) Par exemple, si une femme **a donné naissance** à un enfant, n'est-elle pas donc beaucoup plus capable de s'en occuper au début ? (FL2-AN, <TAN131>, occ. 401)

- UP congruentes en français L1 et L2 : *prendre part* (Ex. 211, 212 et 26) (FqR en FL1 = 1,19 ; FL2-AN = 1 ; FL2-NL = 0,73), *prendre position* (Ex. 213, 214) (FqR en FL1 = 1,49 ; FL2-AN = 0 ; FL2-NL = 2).

Exemples de l'UP *prendre part*:

(211) L'Irlande refuse d'entrer dans la guerre, la France et l'Angleterre sont sur le terrain, la Belgique refuse d'y **prendre part**, mais s'occupe plutôt de la logistique, de la défense, du déminage et des soins aux victimes, si elle devait y prendre une part plus active ce serait vraisemblablement en tant que membre de l'Otan pour venir en aide à la Turquie. (FL1, <TFRB309>, occ. 557)

(212) même dans une société démocratique où le seul façon pour **prendre part** à ce qui se passe est d'électer des politiciens (FL2-AN, <TAN027>, occ. 65)

(26) Il existe cependant des moyens légitimes pour sortir de la pauvreté. Les jeunes qui sont très souvent des chômeurs pourraient chercher un emploi et ainsi **prendre part** à la société et améliorer leur sort. (FL2-AN, <TNL085>, occ. 964)

Exemples de l'UP *prendre position*:

(213) Aujourd'hui, dans la course du temps, il est nécessaire de s'accrocher, de **prendre position** et aussi de répondre à toute question. Ainsi, nous aborderons tous l'année 1992 en pleine conscience. (FL1, <TFRB328>, occ. 608)

(214) Pourtant, il y a toujours beaucoup de défenseurs de la peine de mort. C'est donc la responsabilité des gouvernements de **prendre position** en examinant les avantages et les désavantages de la peine de mort. (FL2-NL, <TNL202>, occ. 1033)

- Structures en *N de N*, avec le verbe *prendre* :

- UP qui sont utilisées plus fréquemment en français L1, mais moins ou pas du tout en L2 : *prendre l'exemple de N* (Ex. 215) (FqR en FL1 = 8,62 ; FL2-AN = 0 ; FL2-NL = 0,73).

(215) **Prenons l'exemple des sorciers ou des magiciens au Moyen-Age.** (FL1, <TFRB078>, occ. 661)

- UP qui sont utilisées plus fréquemment en français L2, mais moins ou pas du tout en français L1 : *prendre Dét cas de N* (Ex. 216) (FqR en FL1 = 1,78 ; FL2-AN = 1 ; **FL2-NL = 4,4**), *prendre Dét moment de réflexion* (Ex. 133, 217 et 218) (FqR en FL1 = 0 ; FL2-AN = 0 ; **FL2-NL = 2,2**).

Exemples de l'UP *prendre Dét cas de N*:

(216) **Prenons** d'abord le cas des romans. De nos jours la majorité de la population ne lit plus. (FL2-NL, <TNL057>, occ. 934)

Occurrences de l'UP *prendre Dét moment de réflexion*:

(133) Dans le cas d'une erreur judiciaire enfin, les conséquences sont catastrophiques. Espérons que tous ceux qui favorisent avec la peine de mort **prennent un *|moment| de réflexion** et se rendent compte du caractère inhumain de cette décision. (FL2-NL, <TNL152>, occ. 1016)

→ Reconstruction native : **prennent un temps de réflexion**

(217) En plus, condamner le criminel à la mort est un acte irréversible qui devient problématique dans le cas d'innocence. Espérons que ceux qui appuient la peine de mort **prennent un *|moment| de réflexion** pour qu'ils ne se transforment pas eux-mêmes en assassin. (FL2-NL, <TNL181>, occ. 1025)

→ Reconstruction native : **prennent un temps de réflexion**

(218) Espérons que tous ceux qui visent à sympathiser avec les défenseurs de la peine de mort **prendront d'abord un *|moment| de réflexion** et qu'ils adopteront une position plus critique envers ces arguments. (FL2-NL, <TNL128>, occ. 1040)

→ Reconstruction native : **prendront d'abord un temps de réflexion**

Cette dernière UP, *prendre Dét *moment de réflexion*, est déviante formellement dans les 3 occurrences observées (FqA = 3 ; FqR = 2,2), mais elle est le fait de LNN différents et ne semble donc pas reposer sur des facteurs individuels. La déviance semble être plutôt de l'ordre du *mélange collocationnel* (cf. Howarth, en II.4.2.3., pp. 134-135) entre l'UP *prendre un temps/*moment de réflexion* et les UP avec structure propositionnelle *prendre/se donner le temps de la réflexion* et *prendre un moment/*temps pour réfléchir*. Le *mélange collocationnel* serait donc dû à la fois à un amalgame lexical et structurel. Cette déviance collocationnelle *prendre Dét *moment de réflexion* pourrait aussi être liée à un mécanisme de transfert de la LS de l'apprenant. D'une part, les structures collocationnelles *een moment/tijd (om) na te denken nemen* ('prendre un moment pour réfléchir') semblent coexister en néerlandais. D'autre part, l'expression très courante *zonder een moment na te denken* ('sans prendre le temps de réfléchir') possède un équivalent français qui est construit avec le nom *temps*, mais n'est pas attesté avec *moment* (*sans prendre le temps/*moment de réfléchir*), à l'inverse du néerlandais qui accepte le nom *moment*, mais pas le nom *tijd* dans cette construction en particulier (*zonder een moment/*tijd na te denken*). L'influence potentielle de l'anglais comme L3 n'est pas non plus à exclure, puisqu'en anglais ce type de

collocation se construit avec le nom *moment* (*to take a moment to think*), mais pas avec le nom *time* (*to take a *time to think*).

- Structures avec proposition subordonnée complément du nom :

- UP qui sont utilisées plus fréquemment en français L1, mais pas du tout en L2 : *prendre Dét temps PSub* (Ex. 45) (FqR en FL1 = 0,89 ; FL2-AN = 0 ; FL2-NL = 0).

(45) Par conséquent, même si de profonds problèmes existent, le pays membre se doit d' être activement conscient, et de **prendre le temps de réfléchir** posément à différentes questions comme la perte d' identité ou l' enrichissement. (FL1, <TFRB328>, occ. 609)

- UP qui sont utilisées plus fréquemment en français L2, mais moins ou pas du tout en français L1 : *prendre Dét décision PSub* (Ex. 219) (FqR en FL1 = 0,3 ; **FL2-AN = 3** ; FL2-NL = 0), *donner Dét impression PSub* (Ex. 220) (FqR en FL1 = 0,89 ; FL2-AN = 0 ; **FL2-NL = 4,4**), *donner Dét occasion PSub* (Ex. 221) (FqR en FL1 = 0,89 ; FL2-AN = 1 ; **FL2-NL = 1,46**), *donner Dét possibilité PSub* (Ex. 222) (FqR en FL1 = 1,78 ; **FL2-AN = 3** ; FL2-NL = 1,46).

(219) Suite à la deuxième guerre mondiale les Européens **ont pris** la décision de se libérer de la domination du dollar. (FL2-AN, <TAN153>, occ. 247)

(220) En effet, appliquer cette peine **donne** l'impression à de nombreux citoyens que l'assassinat est la solution par excellence de résoudre les problèmes qui règnent dans la société. (FL2-NL, <TNL201>, occ. 1214)

(221) L'université remplit dans ces cours-ci sa tâche de formation mais elle provoque le résultat inverse ; elle ne rend pas les étudiants plus critiques et elle ne les **donne** pas l'occasion de réfléchir eux-mêmes. (FL2-NL, <TNL132>, occ. 1112)

(222) Globalement, l'euro est une aventure formidable qui peut harmoniser Europe, pas seulement sur le plan fiscal et économique, mais également sur le plan politique et social. Elle nous **donnera** la possibilité d'assurer la stabilité des prix en Europe, de rééquilibrer les disparités géographiques, de créer une homogénéité partout en ce qui concerne le taux de chômage et la qualité et la pérennité de vie, d'encourager la croissance et de nous rendre finalement stable, solide et compétitif. (FL2-AN, <TAN145>, occ. 531)

- Structures prépositionnelles :

- UP qui sont utilisées plus fréquemment en français L1, mais pas du tout en L2 : *prendre + en charge* (Ex. 223) (FqR = 0,89), *prendre + comme exemple* (Ex. 224) (FqR = 1,19), *prendre + pour exemple* (Ex. 225) (FqR = 0,89), *prendre + pour modèle* (Ex. 226) (FqR = 1,49).

(223) On crée des organismes qui **prennent en charge** les trois pouvoirs exécutif, législatif et judiciaire. (FL1, <TFRB302>, occ. 552)

(224) Or, l' Allemand ne pense pas comme le Français. Que l' on **prenne** comme exemple les appréciations différentes au sujet de la guerre dans le Golfe persique. (FL1, <TFRB320>, occ. 604)

(225) Dès lors, deux choix possibles : le [Réf. = l'autre] **prendre** pour exemple ou se forger son propre mode de pensée. (FL1, <TFRB054>, occ. 649)

(226) Un exemple simple est celui des adolescents : combien **prennent** pour modèle une star et vont jusqu' à recopier parfaitement faits et gestes de leur idole ? (FL1, <TFRB198>, occ. 729)

- UP qui sont utilisées plus fréquemment en français L2, mais moins en français L1 : *prendre + en compte* (Ex. 227) (FqR en FL1 = 2,68 ; **FL2-AN = 4** ; FL2-NL = 0), *prendre + par exemple* (Ex. 228) (FqR en FL1 = 0,59 ; **FL2-AN = 3** ; FL2-NL = 0).

(227) Vu le fait que les bourses, les institutions financières et les PME soient en première ligne aux effets de l'euro, il faut **prendre** en compte leurs avis et leurs actions vers cette monnaie unique. (FL2-AN, <TAN152>, occ. 245)

(228) **Prenons** par exemple la Bosnie. (FL2-AN, <TAN131>, occ. 128)

- UP congruentes en français L1 et L2 : *prendre + en considération* (Ex. 229, 230 et 231) (FqR en FL1 = 1,78 ; FL2-AN = 3 ; FL2-NL = 2,2).

(229) Si le genre humain se targue de ses progrès scientifiques et techniques, la littérature de science-fiction, elle, le met en garde. Elle **prend** tous les risques en considération et les lui présente. (FL1, <TFRB246>, occ. 751)

(230) Tous ces facteurs sont d'importance égales et doivent **être pris** en considération. (FL2-AN, <TAN147>, occ. 233)

(231) Des ces simples définitions, on peut déduire que la peine de mort n'est pas une sanction légitime, puisqu'elle ne **prend** pas en considération l'aspect correctif de la punition (FL2-NL, <TNL175>, occ. 1023)

Parmi tous ces cas, les UP construites avec le verbe *prendre* et le nom *exemple* semblent les plus représentatives de ce qui différencie les productions non natives des productions natives en termes d'écart de fréquence. D'une part, les LN utilisent ces deux cooccurrents (verbe + nom) dans des structures plus diversifiées que les LNN : en L1, nous retrouvons les structures prépositionnelles *prendre + comme/en/par/pour exemple*, la structure avec déterminant « zéro » *prendre exemple (sur)*, la structure en *N de N (prendre Dét exemple de N)* et la structure avec déterminant quelconque ; en L2, nous n'avons qu'une seule occurrence de la structure en *N de N* dans le corpus néerlandophone et trois occurrences de la structure prépositionnelle avec *par* dans le corpus anglophone. D'autre part, les UP construites avec *prendre Dét cas de N* sont employées plus fréquemment par les LNN néerlandophones, par comparaison avec les UP types *prendre Dét exemple de N* qui sont employées moins fréquemment par les LNN tant néerlandophones qu'anglophones. Les structures avec le nom *cas* seraient des *structures collocationnelles* préférées par les LNN néerlandophones, alors que les LNN anglophones auraient pour leur part tendance à éviter de manière générale l'utilisation d'UP à fonction d'exemplification, pourtant très courantes dans les textes argumentatifs en français L1.

Dans un second temps, nous avons mis le doigt sur certaines *structures collocationnelles* qui favorisaient certaines relations sémantiques (cf. la notion de *prosodie sémantique*, sous I.2.4., p. 63). Nous rappelons que les *combinaisons syntagmatiques récurrentes* (CSR) se définissent en fonction du critère de co-sélection récurrente entre certaines structures phraséologiques et des formes lexicales ou des catégories sémantiques particulières (cf. la notion de *patrons*, de *collostructions* et de *cadres collocationnels*, sous I.2.6.). Cet examen nous a permis de mettre au jour des similitudes sémantiques entre des termes entrant dans un même patron syntaxique (cf. I.2.6., pp. 65-66). Pour réaliser cet examen, nous avons mis en corrélation les variables paramétriques portant sur la structure syntaxique du SN des UP (*SyntaxSNUP*), sur la nature du constituant lexical nominal (*Lemme*) et la catégorie sémantique de l'UP (*Sém1ClassGénR2*). Nous mentionnons ci-dessous les cas les plus représentatifs en les classant, comme ci-dessus, par catégorie structurelle.

- Structures avec déterminant « zéro » :

En français L1, 55,6% des UP avec dét. « zéro » ont un contenu prédicatif d'activité mentale (30 occ.) et 29,9% expriment un événement (14 occ.). Il est intéressant de noter que la tendance au sous-emploi des UP avec déterminant « zéro », par les LNN anglophones et néerlandophones, s'accompagne d'une production non déviante de ce type de structure collocationnelle : les LNN utilisent moins fréquemment ces structures collocationnelles par rapport aux LN, mais ils les emploient avec une *précision* native, puisqu'aucune de ces structures collocationnelles, à contenu prédicatif d'activité mentale ou d'événement, n'est déviante. Observant de plus près les sous-catégories sémantiques, nous notons tout d'abord que la fréquence élevée des UP à valeur cognitive (i.e. d'activité mentale) est liée au nombre d'occurrences de l'UP type *prendre conscience* qui est couramment utilisée en L1 (15 occ.) et pas du tout en L2.

Activité mentale	FL2-AN	FL2-NL	FL1	Total
<i>Prendre connaissance</i>	1	0	0	[1]
<i>Prendre conscience</i>	0	0	4,46 [15]	[15]
<i>Prendre exemple (sur)</i>	0	0	1,19 [4]	[4]
<i>Prendre part</i>	1	0,73 [1]	1,19 [4]	[6]
<i>Prendre position</i>	0	1,47 [2]	1,49 [5]	[7]
<i>Donner raison</i>	0	0,73 [1]	0	[1]
<i>Donner sens</i>	0	0	0,3 [1]	[1]
<i>Donner tort</i>	0	0	0,3 [1]	[1]
Total	2	2,93 [4]	[30]	[36]

Tableau 86. Structures collocationnelles à contenu prédicatif d'activité mentale, avec déterminant « zéro », dans les textes argumentatifs des corpus non natifs et natifs

Ensuite, nous notons que la majorité des événements sont des *achèvements* (22 occ. sur 23 occ.) – i.e. des prédicats temporellement bornés ayant une orientation téléique (Ex. 232) –, qui sont construits avec le verbe *donner* (17 occ. sur 22).

(232) Le 25 mars 1957, une étape supplémentaire est franchie: le traité de Rome qui **donne naissance** à la CEE. (FL1, <TFRB304>, occ. 814)

Ces UP d'achèvement construites avec le verbe *donner*, surlignées en gris clair dans le Tableau (87) (ci-dessous), sont plus fréquentes en français L1 (10 occ. sur 14 UP d'achèvement) et dans le corpus néerlandophone (6 occ., i.e. la totalité des UP d'achèvement), que dans le corpus anglophone (1 occ. sur 2 UP d'achèvement).

Événement / Achèvement	FL2-AN	FL2-NL	FL1	Total
<i>Prendre fin</i>	1	0	0	[1]
<i>Prendre forme</i>	0	0	0,59 [2]	[2]
<i>Donner jour</i>	0	0	[1]	[1]
<i>Donner lieu</i>	1	2,93 [4]	0,89 [3]	[8]
<i>Donner naissance</i>	0	1,47 [2]	1,78 [6]	[8]
<i>Prendre place</i>	0	0	0,3 [1]	[1]
<i>Prendre racine</i>	0	0	0,3 [1]	[1]
Total	2	4,4 [6]	4,16 [14]	[22]

Tableau 87. Structures collocationnelles à contenu prédicatif d'événement/achèvement, avec déterminant « zéro », dans les textes argumentatifs des corpus non natifs et natifs

- Structures en *N de N* :

En français argumentatif L1, 79,1% des UP en *N de N* ont un contenu prédicatif d'activité mentale (38 occ.), contre 61,9% des UP dans le corpus néerlandophone (13 occ.) et 37,5% dans le corpus anglophone (3 occ.). Les résultats les plus saillants sont ceux qui concernent les structures avec le nom *exemple*, mentionnées plus haut : les LN utilisent ce cadre collocationnel principalement avec le verbe *prendre* en cooccurrence avec les noms *exemple* (29 occ. sur 38 occ.), contrairement aux non-natifs qui tendent à éviter ce type de structures collocationnelles. En outre, les néerlandophones utilisent plus fréquemment les UP du type *prendre Dét cas de N* (6 occ.) que les anglophones (1 occ.).

Activité mentale	FL2-AN	FL2-NL	FL1	Total
<i>Donner Dét aperçu de N</i>	1	0	0	[1]
<i>Prendre Dét cas de N</i>	1	4,4 [6]	1,78 [6]	[13]
<i>Donner Dét chance de N</i>	0	0	0,3 [1]	[1]
<i>Prendre Dét chemin de N</i> [FL2-AN : 1 occ. déviante (*)]	1	0	0	[1]
<i>Donner Dét conséquence de N</i> [FL2-NL : 1 occ. déviante (*)]	0	0,73 [1]	0	[1]
<i>Prendre Dét droit de N</i> [FL2-NL : 1 occ. déviante (*)]	0	0,73 [1]	0	[1]
<i>Prendre Dét exemple de N</i>	0	0,73 [1]	8,62 [29]	[30]
<i>Donner Dét exemple de N</i> [FL2-NL : 2 occ. déviante (¹)] [FL1 : 1 occ. déviante (²)]	0	1,47 [2]	0,59 [2]	[4]
<i>Donner Dét possibilité de N</i> [FL2-NL : 1 occ. déviante (*)]	0	0,73 [1]	0	[1]
<i>Donner Dét résumé de N</i>	0	0,73 [1]	0	[1]
Total	3	9,53 [13]	11,3 [38]	[54]

Tableau 88. Structures collocationnelles à contenu prédicatif d'activité mentale, en N de N, dans les textes argumentatifs des corpus non natifs et natifs

Concernant la proportion d'UP déviantes, nous constatons que les LNN anglophones produisent une UP déviante (Ex. 233) sur les 3 occurrences observées d'UP d'activité mentale dont la structure du SN est en *N de N*.

(233) chacun a des avantages et des désavantages. Pourquoi pas *|**prendre le chemin de milieu**? (FL2-AN, <TAN030>, occ. 66)

→ Reconstruction native : Pourquoi pas **opter pour une solution/voie intermédiaire** ?

Dans le corpus néerlandophone, 5 UP d'activité mentale en *N de N* sont déviantes, sur 13 occurrences : 4 occurrences avec le verbe *donner* (Ex. 234, 127, 235 et 236) et 1 occurrence avec le verbe *prendre* (Ex. 237).

UP déviantes avec le verbe *donner* dans le corpus néerlandophone (activité mentale et structure en *N de N*) :

(234) dans la conclusion je *|**donne**| les conséquences d'une éducation défailante et j'essaye d'adopter une opinion plus optimiste (voir conclusion) (FL2-NL, <TNL077>, occ. 1072)

→ Reconstruction native : dans la conclusion, je **tire/ m'interroge sur les conséquences d'une éducation défailante**

(127) Le monde romanesque se perd. On peut y insister |en **donnant l'exemple de** Montaigne.

(FL2-NL, <TNL067>, occ. 1067)

→ Reconstruction native : On peut y insister en **prenant l'exemple de Montaigne**.

(235) Sur ce plan il est bien convenable de |**donner l'exemple des mesures prises par le gouvernement flamand pour diminuer le chômage**|. (FL2-NL, <TNL093>, occ. 1080)

→ Reconstruction native : Sur ce plan il est bien convenable de **prendre l'exemple des mesures prises par le gouvernement flamand pour diminuer le chômage**.

(236) Quoique ceci puisse exiger un effort mental surhumain d'eux, la victime ou ses proches doivent **donner** au criminel la possibilité de *|réparation| des fautes commises et de *|répénir|. (FL2-NL, <TNL175>, occ. 1199)

→ Reconstruction native : la victime ou ses proches doivent **donner** au criminel la possibilité de [P Infinitive:] réparer les fautes commises et de se repentir.

UP déviante avec le verbe *prendre* dans le corpus néerlandophone (activité mentale et structure en *N de N*) :

(237) Mais, en appliquant la peine capitale, le gouvernement ne commet-il pas un crime égal au crime commis par le malfaiteur, à savoir celui de tuer quelqu'un. En effet, le gouvernement *|**prend**| le droit de vie de la personne condamnée, un droit qu'ont tous les hommes vivants, y compris les criminels. (FL2-NL, <TNL201>, occ. 1032)

→ Reconstruction native : *En effet, le gouvernement **exerce/ s'arroge/ use d'/ disposer d'un droit de vie (et/ou de mort) sur la personne condamnée***

Une seule déviance native (Ex. 123), portant sur l'UP *donner Dét exemple de N*, a été observée parmi les 38 UP d'activité mentale à structure en *N de N* qui sont produites par les LN. Nous avons explicité les raisons qui nous ont poussée à considérer cette UP comme étant déviante sous V.1.1.3. (p. 236).

(123) Si tout le monde imite une seule personne, il faut donc faire tout ce que celle-ci dit, comme elle le dit. Ionesco nous **donne un ?|merveilleux| exemple d'un fait similaire**. Dans sa pièce intitulée « Rhinocéros », il présente la montée du nazisme et de ses idées auxquelles tout le monde adhère. (FL1, <TFRB093>, occ. 845)

→ Reconstruction native : *Ionesco nous **donne un bel exemple d'un fait similaire***

Par ailleurs, il est intéressant de noter, quand on examine la répartition de ces UP dans les sous-catégories de prédicats d'activité mentale, que la majorité de ces structures en *N de N* remplissent une fonction d'organisation conceptuelle ou langagière (cf. les valeurs 17 à 22 de la variable *Sém4Prédicatif* de la liste paramétrique – Annexe 3), avec 37 occ. sur 38 occ. chez les LN, 11 occ. sur 13 occ. chez les LNN néerlandophones et 1 occ. sur 3 occ. chez les LNN anglophones (*Prendre Dét cas de N*). Ce patron est donc essentiel pour illustrer ou exemplifier le propos d'un texte, en particulier dans le cas de textes académiques à visée argumentative. Allant plus loin, nous constatons que 20 occ. sur les 29 occ. en français L1 sont à l'impératif pluriel, avec une seule occ. à la 2^e pers. (Ex. 238) et les 19 autres à la 1^e pers. du pluriel (*Prenons Dét exemple de N*) (Ex. 239).

(238) **Prenez l' exemple d' un miroir, où vous voyez le reflet d' une personne.** (FL1, <TFRB117>, occ. 691)

(239) **Prenons l' exemple de l' intégration** : de nombreuses études ont démontré que si la consommation de cigarettes chez les jeunes et ce de plus en plus tôt est en augmentation croissante c' est le plus souvent pour des raisons d' integration. (FL1, <TFRB190>, occ. 726)

Les autres cas de *prendre Dét exemple* sont à l'infinif (6 occ.) (Ex. 240) ou entrent dans la structure conditionnelle *si on/nous prend/prenons Dét exemple de N* (3 occ.) (Ex. 241, 242 et 243).

Exemple de l'UP *prendre Dét exemple*, à l'infinif :

(240) On peut **prendre** également l' exemple des parents qui choisissent des prénoms qui sont fort répandu pour leurs enfants. (FL1, <TFRB095>, occ. 672)

Occurrences de l'UP *si on/nous prend/prenons Dét exemple de N*:

(241) Si nous **prenons l' exemple de jeunes à l' école**, certains sont expulsés d' un groupe juste parce qu' il ne s' habille pas de la même façon qu' eux. (FL1, <TFRB130>, occ. 699)

(242) Si on **prend l'exemple de notre société occidentale**, on ne peut nier que nos valeurs principales concernent en partie tout ce qui touche à l'économie. (FL1, <TFRB144>, occ. 704)

(243) Si nous **prenons l'exemple d'Hitler qui a voulu faire des hommes parfait**, nous pouvons constater que toutes les personnes qui ne se conformaient pas à ses règles étaient abattues de sang-froid, et donc ici la peur est installée et obligée à être comme tous le monde, à ressembler à autrui pour vivre. (FL1, <TFRB231>, occ. 745)

Le patron collocationnel le plus fréquent en L1 est donc celui-ci : *prenons Dét exemple de N* (voir l'Ex. 150, *supra*).

- Structures avec proposition subordonnée complément du nom :

En français argumentatif L1, 56% des UP avec un complément propositionnel ont un contenu prédicatif d'activité mentale (28 occ.), pour une proportion moindre dans le corpus néerlandophone (37,5%, 14 occ.), mais supérieure dans le corpus anglophone (88,9%, 8 occ.). Nous présentons, dans le tableau ci-dessous, la répartition des occurrences (FqA et FqR) d'UP d'activités mentales au sein des corpus argumentatifs L2 et L1.

Activité mentale	FL2-AN	FL2-NL	FL1	Total
<i>Donner Dét chance PSub</i>	0	0	0,3 [1]	[1]
<i>Prendre Dét décision PSub</i> [FL2-AN : 1 occ. déviante (*)]	3	0	0,3 [1]	[4]
<i>Donner Dét droit PSub</i>	0	0,73 [1]	0	[1]
<i>Donner Dét moyen PSub</i>	0	0	0,59 [2]	[2]
<i>Donner Dét occasion PSub</i>	1	1,47 [2]	0,89 [3]	[6]
<i>Donner Dét opportunité PSub</i>	0	0	0,3 [1]	[1]
<i>Donner Dét possibilité PSub</i>	3	1,47 [2]	1,78 [6]	[11]
<i>Prendre Dét possibilité PSub</i> [FL2-AN : 1 occ. déviante (?)]	1	0	0	[1]
<i>Donner Dét volonté PSub</i>	0	0,73 [1]	0	[1]
Total	8	4,4 [6]	4,16 [14]	[28]

Tableau 89. Structures collocationnelles à contenu prédicatif d'activité mentale, avec proposition subordonnée complément du nom, dans les textes argumentatifs des corpus non natifs et natifs

Il ressort de ce tableau que les sous-catégories sémantiques de prédicat d'activité mentale concernées sont soit des prédicats exprimant une opportunité, surlignés en gris clair dans le Tableau (89) ci-dessus (*Donner Dét chance/ droit/ moyen/ occasion/ opportunité/ possibilité + PSub*), soit des prédicats conatifs (*Prendre Dét décision + PSub* – cf. Ex. 112, 219, 244 et 245, ci-dessous) ou volitifs (*Donner Dét volonté + PSub* – cf. l'Ex. 63, ci-dessous).

Occurrences de l'UP *Prendre Dét décision + PSub*:

(112) D'un côté, au moment du lancement de l'euro *lune décision a été prise *que la banque centrale devrait être indépendante lorsque, de l'autre côté, aux États-Unis, la monnaie est beaucoup plus solide parce que la banque centrale américaine a un Etat derrière sa structure. (FL2-AN, <TAN151>, occ. 241)

→ Reconstruction native : D'un côté, au moment du lancement de l'euro la décision a été prise de *[P Infinitive:] rendre la banque centrale indépendante*

(219) Suite à la deuxième guerre mondiale les Européens **ont pris** la décision de se libérer de la domination du dollar. (FL2-AN, <TAN153>, occ. 247)

(244) La décision prise par l'UE de mettre en place une monnaie unique n'était pas fondée uniquement sur des considérations économiques. (FL2-AN, <TAN157>, occ. 250)

(245) Par ailleurs, si aucune femme n' **avait pris** la décision de s' affirmer en tant que telle et de réclamer ses droits, quel statut aurait-on aujourd'hui dans nos sociétés occidentales ? (FL1, <TFRB217>, occ. 736)

Occurrence de l'UP *Donner Dét volonté* + PSub:

(63) En revanche, nous ne pouvons pas non plus négliger les immenses avantages qu'apportent des systèmes universitaires à une société telle que la nôtre. Apprendre aux jeunes d'être critique et de se rendre compte du monde dans lequel ils vivent. (...) Leur **donner** la volonté de se consacrer entièrement à une branche qui porte tout leur intérêt. (FL2-NL, <TNL122>, occ. 1104)

Notons toutefois que les structures sélectionnées par le mot *décision* sont le plus souvent des structures sans proposition subordonnée, i.e. avec un modifieur quelconque (cf. Tableau 90, ci-dessous), en L1 (Ex. 246) (9 occ.) comme en L2, avec respectivement 7 occurrences dans le corpus anglophone (Ex. 247), dont une occurrence déviante (Ex. 248), et 3 occurrences non déviantes dans le corpus néerlandophone (Ex. 249).

Exemple de l'UP *prendre Dét décision* avec modifieur quelconque, en français L1:

(246) A l' âge adulte, l' homme est confronté à **prendre** de grandes décisions et à faire de plus en plus de choix. (FL1, <TFRB041>, occ. 638)

Exemple de l'UP *prendre Dét décision* avec modifieur quelconque, dans le corpus anglophone de français L2:
(247) c'est à lui de **prendre** les décisions les plus importantes au niveau familial. (FL2-AN, <TAN130>, occ. 127)

UP déviante de *prendre Dét décision* avec modifieur quelconque, dans le corpus anglophone de français L2:
(248) d'innombrables *[décisions, plus petites], moins importantes qui ont été prises avant (FL2-AN, <TAN014>, occ. 58)

→ Reconstruction native : d'innombrables décisions, moindres/ mineures, moins importantes qui ont été prises avant

Exemple de l'UP *prendre Dét décision* avec modifieur quelconque, dans le corpus néerlandophone de français L2:

(249) Cependant, nous pouvons nous poser le question de savoir si les membres d'un jury sont capables de **prendre** une décision rationnelle et objective concernant la mort ou la vie d'un concitoyen. (FL2-NL, <TNL212>, occ.1035)

En français L1, 13 occurrences sur 14 occurrences (surlignées en gris clair dans le Tableau 89, *supra*) appartiennent à la sous-catégorie des *opportunités* (cf. valeur 21 de la variable *Sem4Prédicatif* de la liste paramétrique – Annexe 3). Ces dernières peuvent être construites, en français L2, soit avec proposition subordonnée (Ex. 250) (cf. Tableau 89, *supra* – 5 occ. par les anglophones et 3 occ. par les néerlandophones), soit avec un modifieur quelconque (Ex. 251) (cf. Tableau 90, ci-dessous).

Activité mentale	FL2-AN	FL2-NL	FL1	Total
<i>Prendre Dét décision (Modifieur)</i> [FL2-AN : 1 occ. déviante (*)]	7	2,2 [3]	2,68 [9]	[19]
<i>Donner Dét chance (Modifieur)</i>	0	2,2 [3]	0	[3]
<i>Donner Dét choix (Modifieur)</i> [FL2-AN : 3 occ. déviantes (3)]	3	0	0	[3]
<i>Donner Dét occasion (Modifieur)</i> [FL2-AN : 1 occ. déviante (*)]	2	0	0	[2]

Tableau 90. Structures collocationnelles avec un modifieur quelconque, construites avec les noms *décision*, *chance*, *choix* et *occasion*, dans les textes argumentatifs des corpus non natifs et natifs

Exemple d'UP d'opportunité construite avec une proposition subordonnée complément du nom, en français L2 :

(250) Il est important de **donner** aux enfants la possibilité d'évoluer à leurs propres rythmes. (FL2-AN, <TAN227>, occ. 458)

Exemple d'UP d'opportunité construite avec un modifieur quelconque, en français L2 :

(251) La peine de mort est intolérable, non pas parce qu'il devrait être interdit de faire à un autre ce qu'on aimerait pas qu'on fasse à soi-même, mais parce qu'il serait très dommage de ne pas profiter de l'occasion pour **donner** une deuxième chance au coupable. (FL2-NL, <TNL173>, occ. 1198)

Contrairement aux LNN, les LN emploient presque exclusivement des constructions avec proposition subordonnée pour cette sous-catégorie sémantique d'UP: 13 UP d'opportunité sont construites avec une proposition subordonnée complément du nom (Ex. 252), contre une seule UP en *N de N* (Ex. 253).

Exemple d'UP d'opportunité construite avec une proposition subordonnée complément du nom, en français L1 :

(252) Dans le domaine social, le projet de 1992 **donnera** la possibilité aux travailleurs d'être employer dans n'importe quel pays de la Communauté. (FL1, <TFRB289>, occ. 807)

Occurrence de l'UP d'opportunité en *N de N*, en français L1 :

(253) Quand l'Europe parviendra à entretenir en son sein assez de cultures fortes et variées pour se **donner** les chances d'une seconde Renaissance... alors la nymphe, enlevée par Zeus tonnant, enfantera un monde sans discrimination, ouvert, juste et incroyablement riche d'identités ethniques diverses... L'Europe... (FL1, <TFRB308>, occ. 794)

Notons également que, hormis l'UP déviante avec le verbe *prendre* + *possibilité* (Ex. 106, ci-dessous), les constructions avec un prédicat exprimant l'opportunité sont toutes construites avec le verbe *donner* en L1 et en L2, quelle que soit la structure de leur SN.

(106) On a le pouvoir de choisir sa vie, donc on doit la mener comme on veut. Si on a la possibilité de réussir on doit la ?|prendre par les deux mains|, mais on doit se rendre compte qu'il y a d'autres personnes et on doit aussi respecter la manière dont ils vivent leur vie. (FL2-AN, <TAN031>, occ. 67)

→ Reconstruction native : Si on a la possibilité/ l'opportunité de réussir on doit la saisir sans hésiter/ Si on a la possibilité de réussir on doit sauter à pieds joints sur cette occasion

Suite à ces observations, il nous a paru intéressant de vérifier si la différence entre les UP exprimant l'opportunité en L1 (qui présentent une nette majorité de constructions avec une proposition subordonnée) vs L2 (qui présentent une proportion équivalente de SN à structure avec proposition et sans proposition subordonnée) n'était pas due à des emplois déviants des structures avec modifieur quelconque par les LNN. Dans le Tableau (91) (ci-dessous), nous présentons les occurrences d'UP d'activité mentale exprimant

l'opportunité en L2 et en L1, qui sont contruites soit avec une proposition subordonnée complément du nom, soit avec un modifieur quelconque.

Activité mentale/ Opportunité			
SN avec PSub	FL2-AN	FL2-NL	FL1
<i>Donner Dét chance PSub</i>	0	0	0,3 [1]
<i>Donner Dét droit PSub</i>	0	0,73 [1]	0
<i>Donner Dét moyen PSub</i>	0	0	0,59 [2]
<i>Donner Dét occasion PSub</i>	1	1,47 [2]	0,89 [3]
<i>Donner Dét opportunité PSub</i>	0	0	0,3 [1]
<i>Donner Dét possibilité PSub</i>	3	1,47 [2]	1,78 [6]
<i>Prendre Dét possibilité PSub</i> [FL2-AN : 1 occ. déviante (?)]	1	0	0
Sous-total	5	3,67 [5]	3,87 [13]
SN avec modifieur quelconque	FL2-AN	FL2-NL	FL1
<i>Donner Dét chance (Modifieur)</i>	0	2,2 [3]	0
<i>Donner Dét choix (Modifieur)</i> [FL2-AN : 3 occ. déviantes (?)]	3	0	0
<i>Donner Dét occasion (Modifieur)</i> [FL2-AN : 1 occ. déviante (*)]	2	0	0
Sous-total	5	2,2 [3]	0
Total	10	5,87 [8]	3,87 [13]

Tableau 91. Structures collocationnelles à contenu prédicatif d'activité mentale d'opportunité, avec ou sans proposition subordonnée complément du nom, dans les textes argumentatifs des corpus non natifs et natifs

Parmi les UP d'opportunité avec modifieur quelconque, les 3 occurrences de *donner* + *chance* du corpus néerlandophone (Ex. 254, 255 et 251) se sont avérées n'être pas déviantes.

Occurrences des UP construites avec *donner* + *chance*, avec ou sans proposition subordonnée complément du nom :

(254) Mais que faire quand l'assistance psychologique n'a aucun résultat, quand il n'y a pas de changement dans la conduite des malfaiteurs ? Faut-il prendre encore plus de patience ? Faut-il les **donner une nouvelle chance** encore ? (FL2-NL, <TNL086>, occ. 1114)

(255) Une société plus avancée, par contre, devrait **donner une nouvelle chance** aux délinquants. (FL2-NL, <TNL160>, occ. 1191)

(251) La peine de mort est intolérable, non pas parce qu'il devrait être interdit de faire à un autre ce qu'on aimerait pas qu'on fasse à soi-même, mais parce qu'il serait très dommage de ne pas profiter de l'occasion pour **donner une deuxième chance** au coupable. (FL2-NL, <TNL173>, occ. 1198)

A l'inverse, les 3 occurrences de *donner* + *choix* produites par les anglophones sont toutes trois contextuellement déviantes, avec une déviance portant soit sur toute l'UP (V + N) (Ex. 256 et 257), soit sur le verbe (Ex. 258).

(256) Philes nous ?**donne la choix** entre deux manières de vivre totalement contraires. A l'un côté, on a la description d'une système où les civils peuvent partager aux affaires politiques, économiques, etc. (...) A l'autre côté, le citoyen contemporain n'est pas accomodé pour une vie en pleine nature. (FL2-AN, <TAN027>, occ. 327)

→ Reconstruction native : *Philes nous soumet/ fait part de/ expose/ présente (le choix entre) deux manières de vivre totalement contraires*

(257) Il n'est pas plus qu'un animal : il vit pour survivre. Ainsi ?|le choix donné| n'existe pas. Une vie sans histoire est une vie sans but (et en plus un vie sans histoires n'est pas égale à une vie au milieu des oiseaux). L'homme ne peut pas renoncer à ce qu'il a commencé. (FL2-AN, <TAN029>, occ. 328)

→ Reconstruction native : *Ainsi il est impossible de choisir/ aucun choix ne s'offre à lui/ tout est décidé d'avance/ etc.*

(258) Pour avoir de l'égalité entre les sexes, est-ce que ce n' est pas nécessaire d' avoir de l' uniformité ? La seule façon dont une femme peut se concurrencer avec un homme dans le monde du travail est en étant capable de jouer les mêmes rôles que lui. (...) Donc, supprimer les différences veut dire ?|donner le choix|. (FL2-AN, <TAN131>, occ. 402)

→ Reconstruction native : *Donc, supprimer les différences veut dire **laisser le choix***

Ensuite, sur les 3 occurrences de *donner + occasion* produites par les apprenants anglophones (cf. l'Ex. 99 avec proposition subordonnée et les Ex. 259 et 260 avec modifieur quelconque), une des deux UP catégorisées dans les structures du SN avec modifieur quelconque est déviante formellement. La déviance porte en l'occurrence sur la structure du complément du nom, puisque ce dernier est manquant (complément « zéro ») (Ex. 99).

(99) c'est également vrai à dire que l'homme peut montrer, quand on lui *|*Cpl Zéro*| **donne** l'occasion, qu'il est assez capable de s'occuper des enfants que la femme (FL2-AN, <TAN130>, occ. 398)

→ Reconstruction native : *quand on lui **en donne** l'occasion*

(259) L'éducation entraîne l'esprit. Elle peut changer nos façons de penser. Elle peut ouvrir des chemins dans la vie, nous **donner** une occasion qu'on aurait jamais eu. (FL2-AN, <TAN165>, occ. 408)

(260) Sans ces types d'attributs un Etat se sent fragile ; l'euro **donne** aux pays d'Europe l'occasion de travailler ensemble et d'être une vraie puissance sur la scène internationale. (FL2-AN, <TAN140>, occ. 525)

Sur l'ensemble des UP avec *donner + possibilité* en L2, qui sont toutes construites avec une proposition subordonnée (3 occ. dans le corpus anglophone et 3 occ. dans le corpus néerlandophone), une occurrence produite par les apprenants néerlandophones est déviante formellement et porte sur le complément du nom (Ex. 236).

(236) Quoique ceci puisse exiger un effort mental surhumain d'eux, la victime ou ses proches doivent **donner** au criminel la possibilité de *|réparation| des fautes commises et de *|repentir|. (FL2-NL, <TNL175>, occ. 1199)

→ Reconstruction native : *la victime ou ses proches doivent **donner** au criminel la possibilité de [P Infinitive:] réparer les fautes commises et de se repentir*

En bref, excepté les 3 UP avec le nom *choix* produites par les apprenants anglophones qui sont toutes déviantes (cf. Ex. 256, 257 et 258, *supra*), les déviations portant sur les UP d'activité mentale d'opportunité, dont la structure du SN se caractérise par la présence d'un complément du nom propositionnel ou d'un modifieur quelconque (cf. Tableau 91, ci-dessus), sont peu nombreuses (5 occ. en français L2). Ces UP déviantes sont en outre uniquement le fait des apprenants anglophones.

- Structures prépositionnelles :

Les structures prépositionnelles du SN des UP présentent principalement, en français argumentatif L1, des UP d'activités mentales : 59,6% des UP prépositionnelles (*Prép + SN*) ont un contenu prédicatif d'activité mentale (28 occ.) en L1, pour une proportion légèrement inférieure dans le corpus néerlandophone (50%, 4 occ.), mais supérieure dans le corpus anglophone (78,6%, 11 occ.). Le sous-emploi des structures prépositionnelles (cf. V.1.2., pp. 242-243 et V.3.2.2, pp. 273-274) et des prédicats d'activité mentale (cf. V.2.2., p. 259), que nous avons mis au jour pour les LNN néerlandophones, semble donc être confirmé quand on ne considère que les UP prédictives d'activité mentale à structure prépositionnelle. C'est ainsi que dans le Tableau (92) (ci-dessous), nous répertorions les UP avec SN prépositionnel dont le contenu prédicatif est une activité mentale, en L2 et en L1.

Activité mentale	FL2-AN	FL2-NL	FL1	Total
<i>Prendre + dans Dét champ</i> [FL1 : 1 occ. déviante (*)]	0	0	0,3 [1]	[1]
<i>Prendre + en compte</i>	4	0	2,68 [9]	[13]
<i>Prendre + en considération</i>	3	2,2 [3]	1,78 [6]	[12]
<i>Prendre + dans Dét ensemble</i>	0	0	0,3 [1]	[1]
<i>Prendre + comme exemple</i> [FL1 : 1 occ. déviante (*)]	0	0	1,19 [4]	[4]
<i>Prendre + en exemple</i> [FL1 : 1 occ. déviante (')]]	0	0	0,3 [1]	[1]
<i>Prendre + par exemple</i>	3	0	0,59 [2]	[5]
<i>Prendre + pour exemple</i>	0	0	0,3 [1]	[1]
<i>Prendre + à la légère</i>	1	0	0	[1]
<i>Prendre + au pied de la lettre</i>	0	0	0,3 [1]	[1]
<i>Prendre + dans Dét sens</i> [FL2-NL : 1 occ. déviante (')]]	0	0,73 [1]	0,59 [2]	[3]
Total	11	2,93 [4]	8,32 [28]	[43]

Tableau 92. Structures collocationnelles prépositionnelles à contenu prédicatif d'activité mentale, dans les textes argumentatifs des corpus non natifs et natifs

Examinant les UP *types* qui sont reprises dans ce tableau, il est intéressant de noter que les UP prépositionnelles à contenu d'activité mentale sont toutes sans exception, en L2 et en L1, construites avec le verbe *prendre*. Mis à part les cas des UP avec *prendre + exemple* dont nous avons déjà discuté précédemment – qui se construisent avec un plus grand nombre de types de prépositions en L1 (*comme, par, pour*) (Ex. 261, 262 et 263) qu'en L2 (*par*) (Ex. 264) – nous mettons ici le doigt sur les UP construites avec le verbe *prendre* et le SN prépositionnel *en compte* (surlignées en gris clair dans le Tableau 92), fréquentes en L1 (Ex. 265) (9 occ.), mais qui semblent être évitées par les LNN néerlandophones qui n'en utilisent aucune.

Exemple d'UP construite avec *prendre + comme exemple*, en français L1 :

(261) **Prenons** comme exemple, un adolescent qui veut essayer de s'introduire dans une bande de jeunes. (FL1, <TFRB119>, occ. 692)

Exemple d'UP construite avec *prendre + par exemple*, en français L1 :

(262) Ainsi ce besoin de se distinguer ressurgit à chaque période de notre histoire. **Prenez** l'Antiquité par exemple, des héros spectaculaires et grandioses ont marqué l'histoire de leur empreinte. (FL1, <TFRB011>, occ. 619)

Exemple d'UP construite avec *prendre + pour exemple*, en français L1 :

(263) J'en **prends** pour exemple La planète des singes, où le protagoniste ne comprend qu'il se trouve sur la planète bleue que quand il aperçoit la statue de la liberté. (FL1, <TFRB246>, occ. 750)

Exemple d'UP construite avec *prendre + par exemple*, en français L2 :

(264) **prenons** par exemple l'Irlande du nord ou le Moyen Orient (FL2-AN, <TAN131>, occ. 129)

Exemple d'UP construite avec *prendre + en compte*, en français L1 :

(265) Mais c' est justement ce jeu d' esprit qu' il faut **prendre en compte** car il développe en nous la prudence et la critique envers les idées perçues non seulement dans le film mais aussi celles de la vie de tous les jours. Le cinéma ne nous apporte pas que des notions nouvelles mais il en requiert aussi. (FL1, <TFRB126>, occ. 698)

Concernant les UP prépositionnelles exprimant une activité mentale qui sont déviantes, une seule UP déviante est produite par un apprenat non natif (Ex. 132).

UP déviante construite avec *prendre + dans Dét sens*, en français L2 :

(132) Car les adultes, eux non plus, ne font le sport pour le sport. " Le sport est l'art par lequel l'homme se libère de soi-même " a dit Giraud. **Prise** dans ?l|el sens négatif, il n'y a rien à redire à cette phrase. (FL2-NL, <TNL107>, occ. 954)

→ Reconstruction native : **Prise** dans un/son sens négatif, il n'y a rien à redire à cette phrase

Les autres UP déviantes sont produites par les LN. L'UP construite avec *prendre + dans Dét champ* (Ex. 266) est une déviance formelle, puisque cette UP n'est attestée sous cette forme ni dans les corpus de référence de français L1 (*Le Soir* 1997 et *Valibel*), ni dans la base de données lexicales *Les Voisins De Le Monde*, ni dans la liste du LADL. Par contre, l'UP *prendre + dans Dét sens*, qui correspond au sens activé par le contexte, est attestée dans les corpus de français L1 (FqR *Le Soir* = 0,37 et FqR *Valibel* = 2,9), dans la liste des expressions figées du LADL (expression correspondante : *prendre dans un certain sens*) et dans *Les Voisins De Le Monde* avec un indice d'IM élevé (IM = 4,008).

UP déviante construite avec *prendre + dans Dét champ*, en français L1 :

(266) Se distinguer des autres est un rêve en chacun de nous et que la société ou l'on vit, et notre manière de penser, favorise l'envie de le réaliser. Il faut **prendre** cette phrase dans un *|champ| plus large, en y voyant un moyen pour la paix, une unification de ce que l'on a, au dépend de pensée personnelle. (FL1, <TFRB126>, occ. 644)

→ Reconstruction native : Il faut **prendre** cette phrase dans un sens plus large

L'UP construite avec *prendre + comme exemple* (Ex. 267) a également été considérée comme étant déviante du point de vue de la forme, puisque la construction *prendre + comme exemple du contraire*, n'est pas attestée dans les corpus de référence de français L1 (*Le Soir* 1997 et *Valibel*), ni dans les ressources collocationnelles de français L1 dont nous disposons, i.e. dans la liste des expressions verbales figées du LADL et la base

distributionnelle *Les Voisins De Le Monde*.¹¹⁶ La reconstruction native¹¹⁷ repose ici principalement sur notre intuition, car aucun cooccurrent syntaxique n'a émergé de l'examen des cooccurrents syntaxiques potentiels sur le site *Les Voisins De Le Monde*. Malgré les réserves que nous émettons sur la pertinence du recours au moteur de recherche *Google* dans le cadre d'études sur corpus, signalons tout de même que nous avons trouvé, via *Google.fr*, trois occurrences de la séquence « prendre comme contre-exemple(s) ». Ces occurrences, bien que peu nombreuses, confirment le caractère attesté dans l'usage L1 de la reconstruction native proposée.

UP déviante construite avec *prendre + comme exemple*, en français L1 :
 (267) Il est néanmoins simpliste de croire que la "Belgique" n'existe plus dans les consciences. **Prendre comme *exemple du contraire**, l'atonium, l'exposition de 1958, le Congo Belge et l'intérêt qui les entourent toujours serait tomber dans les travers des clichés faciles. (FL1, <TFRB304>, occ. 595)
 → Reconstruction native : **Prendre comme contre-exemples** l'atonium, (...)

La troisième UP prépositionnelle déviante en français L1 est construite avec *prendre + en exemple* (Ex. 268). La déviance est ici quantitative, puisque cette UP est attestée formellement dans le sens d' 'exemplifier/ illustrer un propos' en français L1 (FqR *Valibel* = 1,32) et montre un indice d'association lexicale relativement élevé (IM = 3,507), mais n'est pas préférée par les natifs qui utilisent plus fréquemment l'UP *prendre + comme exemple* (FqR *Le Soir* = 0,3 et FqR *Valibel* = 3,68) se caractérisant par un indice d'IM plus élevé (IM = 5,769).

UP déviante construite avec *prendre + en exemple*, en français L1 :
 (268) De plus, les bons films proposés ne sont pas légions. Et enfin, il est évident qu'il n' y a pas que des illettrés qu' y intéressent il suffit de **prendre l'en exemple des personnages tels que Godart, Renoir ou Buñuel**. (FL1, <TFRB121>, occ. 695)
 → Reconstruction native : il suffit de **prendre comme exemple des personnages tels que Godart, Renoir ou Buñuel**

En conclusion, nous n'avons pas pu mettre ici en exergue des *doudous polylexicaux* à proprement parler, principalement à cause de la taille restreinte de nos corpus argumentatifs d'apprenants du français L2. Par contre, nous avons mis en évidence certaines *structures collocationnelles* qui ont tendance à être plus employées ou évitées par les LNN.

¹¹⁶ La SVN prépositionnelle *prendre + comme exemple* est quant à elle attestée en français L1 (FqR *Le Soir* = 0,3 et FqR *Valibel* = 3,68), avec un modifieur différent de celui de l'Ex. 267 (i.e. *du contraire*), et montre un indice statistique d'association lexicale élevé (IM = 5,769). Nous l'avons donc considérée comme appartenant au champ phraséologique.

¹¹⁷ L'emploi du terme reconstruction « native » peut paraître incongru aux yeux du lecteur, étant donné que l'UP déviante est ici elle-même le fait d'un LN. Rappelons néanmoins que nous avons traité les données des corpus L2 et L1 de manière identique, sans opérer *a priori* de distinction entre les données produites par des LNN de celles produites par les LN. L'analyse collocationnelle et l'analyse de déviance ont donc été effectuées en appliquant exactement la même méthode et les mêmes critères d'analyse à l'ensemble des données, qu'elles proviennent de corpus de français L2 ou L1.

VI.3. Conclusion intermédiaire

Nous avons tout d'abord déterminé, dans ce chapitre, le taux de précision (en VI.1.) et le degré de complexité structurelle et lexicale (en VI.2.1. et VI.2.2.) des interlangues (IL) des apprenants avancés anglophones *vs* néerlandophones, en les comparant avec la complexité et la précision propres aux locuteurs natifs (LN), dans des textes académiques à visée argumentative. Nous avons ensuite mis en exergue des traits phraséologiques, en l'occurrence des structures collocationnelles (en VI.2.3.) qui sont propres à ce type de textes. En résumé, nous avons d'abord mesuré le niveau global de compétence phraséologique des groupes d'apprenants étudiés, en fonction de leur langue source (LS), puis nous avons mis en exergue les structures lexico-grammaticales spécifiques à chacun de ces groupes.

Premièrement, par le biais de trois mesures de précision adaptées au phénomène phraséologique, nous avons constaté que les LNN néerlandophones étaient plus proches de la compétence phraséologique native que les LNN anglophones dans leur production d'UP construites avec les deux verbes *prendre* et *donner* [Seconde partie de l'Hypothèse 4 vérifiée]. La mesure de précision « améliorée » pour 1000 mots (à partir de Foster & Wigglesworth, 2007) a en outre montré que les LNN néerlandophones commettaient plus d'UP gravement déviantes que les anglophones, mais que ce facteur qualitatif de gravité des déviances n'influençait pas de manière cruciale le niveau général de compétence phraséologique des LNN.

Deuxièmement, par le biais d'une mesure de complexité structurelle du SN des UP, élaborée dans le cadre de ce travail, nous avons montré que les LNN néerlandophones faisaient preuve d'une meilleure compétence phraséologique que les LNN anglophones à cet égard [Partie de l'Hypothèse 1 sur la complexité structurelle vérifiée]. Nous avons en outre constaté que les UP construites avec *prendre* avaient tendance à être structurellement plus complexes que les UP construites avec *donner* en français L2 et L1. Ce constat plaide en faveur de la complexité linguistique (ici, structurelle) comme cause possible du sous-emploi des UP avec *prendre vs donner* par les apprenants néerlandophones du français L2 dans les textes argumentatifs. Les LNN anglophones, pour leur part, ne seraient pas influencés par la complexité structurelle du verbe *prendre*. Nous pourrions donc penser (i) que la structure du SN des UP construites avec *take* en anglais L1 est aussi complexe que les structures avec *prendre* en français L1, ce qui favoriserait le transfert (conscient ou non) des structures du SN des UP avec *take*

en L2 par les LNN anglophones, et/ou (ii) que les LNN anglophones se caractérisent par une attitude favorisant la *complexité* par opposition à la *précision* (cf. la variable individuelle de l'*attitude* et la notion de *risktaker*, en II.4.2.3., pp. 127-128), contrairement aux LNN néerlandophones qui tendent à éviter de produire des UP construites avec *prendre*. Nous pouvons raisonnablement supposer que le mécanisme d'évitement activé par ces derniers serait quant à lui dû (i) à une prise de conscience de la complexité structurelle du verbe *prendre* (*vs* une complexité structurelle moindre pour *nemen*), (ii) à une prise de conscience de la complexité du phénomène phraséologique en général (*vs* combinatoire libre) et/ou (iii) à une prise de risque globalement moins marquée dans l'attitude des apprenants, qui favorise la *précision* par opposition à la *complexité* en situation de production d'une L2.

À cette mesure de *complexité structurelle*, nous avons ajouté une mesure de *diversité* ou de *richesse lexicale* des UP construites avec *prendre vs donner* en français L2 et L1 [Partie de l'Hypothèse 1 sur la complexité lexicale vérifiée]. D'après les résultats obtenus pour le français L1, les UP construites avec le verbe *donner* sont plus riches lexicalement que celles construites avec le verbe *prendre*, puisque *donner* se combine avec un plus grand nombre d'unités lexicales *types* que le verbe *prendre* (cf. la mesure du *type-token ratio* pondérée par le biais de l'indice de Guiraud). Cette différence entre la combinatoire lexicale des deux verbes a également été observée dans le corpus argumentatif anglophone. Par contre, les UP construites avec le verbe *prendre* se sont avérées plus diversifiées lexicalement chez les LNN néerlandophones que chez les natifs : les UP construites avec *prendre*, dans le corpus argumentatif néerlandophone, se combinent avec un nombre plus important d'unités lexicales nominales *types* que dans le corpus natif. Nous pouvons raisonnablement supposer, sur la base de ces résultats, que le verbe *nemen* posséderait une combinatoire lexicale plus libre que le verbe *prendre* en français argumentatif L1, ce qui favoriserait le transfert du potentiel combinatoire (plus étendu) des UP construites avec *nemen* par le LNN néerlandophone, sur les UP qu'il produit avec le verbe *prendre* dans l'interlangue. Ce constat semble aller dans le sens d'une différence marquée, tant sur le plan structurel que lexical, entre les deux verbes correspondants en français (*prendre*) et en néerlandais (*nemen*), laissant supposer que le premier aurait un potentiel phraséologique plus important que le second : en termes de complexité structurelle, les structures du SN des UP avec *prendre* seraient plus contraintes (donc plus complexes) que celles du verbe *nemen* ; en termes de possibilités combinatoires lexicales, le potentiel combinatoire des UP avec *prendre* serait plus limité

que celui de *nemen* en néerlandais. Ces deux facteurs, qui plaident en faveur d'un mécanisme d'évitement mis en œuvre (consciemment ou non) par le LNN pour contourner la difficulté liée au potentiel phraséologique élevé du verbe *prendre* (vs *nemen* en néerlandais), semblent expliquer en partie les causes du sous-emploi des UP avec *prendre* chez les LNN néerlandophones. Ces hypothèses pourraient ultérieurement faire l'objet d'études contrastives plus spécifiques, incluant un examen détaillé des emplois phraséologiques des verbes correspondants dans les langues sources, afin d'expliquer les causes de ces différences d'emploi dans la production d'UP construites avec les deux verbes à haute fréquence *prendre* et *donner* en français L2.

Troisièmement, nous avons mis en exergue (sous VI.2.3.) des tendances dénotant un écart de fréquence dans la production de certaines *structures collocationnelles* entre, d'une part, les LNN anglophones et néerlandophones et, d'autre part, entre les LNN et les LN. Nous avons ainsi examiné la manière dont les UP *types* étudiées et/ou les *catégories sémantiques* d'UP les plus fréquentes entraînent en combinaison avec certaines *structures collocationnelles* particulières. Dans le tableau de synthèse que nous présentons ci-dessous (cf. Tableau 93), nous répertorions les catégories sémantiques d'UP et les UP types qui sont utilisées moins fréquemment en L2 qu'en L1 (1^e ligne du Tableau 93) et, inversement, celles qui sont plus fréquemment utilisées en L2 qu'en L1 (2nde ligne du Tableau 93)¹¹⁸, en les catégorisant par type de structure du SN des UP.

¹¹⁸ Il resterait à vérifier si ces écarts de fréquence sont statistiquement significatifs et interprétables en termes de sous-emploi ou suremploi, par le biais d'une analyse portant sur de plus vastes corpus argumentatifs.

	FL2-AN	FL2-NL
UP moins fréquentes en français L2 (vs L1)	Structures collocationnelles avec dét. « zéro »	
	<ul style="list-style-type: none"> • Catégories sémantiques : UP d'activité mentale (> absence de l'UP <i>prendre conscience de</i>) • UP types : <i>prendre conscience de, donner envie, prendre exemple sur</i> 	
	<ul style="list-style-type: none"> • UP d'événement/ achèvement avec <i>donner</i> (par ex., en L1, <i>donner jour/ lieu/ naissance</i>) 	/
	Structures collocationnelles en <i>N de N</i>	
	<ul style="list-style-type: none"> • Catégories sémantiques : UP d'activité mentale • UP type (absente): <i>prendre Dét exemple de N</i> (en particulier, de <i>prenons Dét exemple de N</i>) 	<ul style="list-style-type: none"> • UP type (moins fréquente): <i>prendre Dét exemple de N</i> (en particulier, <i>prenons Dét exemple de N</i>)
	Structures collocationnelles avec PSub compl. du nom	
	<ul style="list-style-type: none"> • UP type : <i>prendre Dét temps PSub</i> 	/
	/	<ul style="list-style-type: none"> • Catégorie sémantique : UP d'activité mentale
	Structures collocationnelles prépositionnelles	
	<ul style="list-style-type: none"> • UP avec <i>prendre + Prép exemple</i> • UP types : <i>prendre + en charge, prendre + comme exemple, prendre + pour exemple, prendre + pour modèle</i> 	/
/	<ul style="list-style-type: none"> • Catégorie sémantique : UP d'activité mentale (sous-emploi) • UP type : <i>prendre + en compte</i> 	
UP plus fréquentes en français L2 (vs L1)	Structures collocationnelles avec dét. « zéro »	
	<ul style="list-style-type: none"> • UP type : <i>donner naissance</i> (incluant les prédicats d'événement/ achèvement et les prédicats d'activité physique au sens de 'mettre au monde', 'accoucher') 	<ul style="list-style-type: none"> • UP type : <i>donner lieu</i>
	Structures collocationnelles en <i>N de N</i>	
	/	<ul style="list-style-type: none"> • UP types : <i>prendre Dét cas de N, prendre Dét moment de réflexion</i>
	Structures collocationnelles avec Psub compl. du nom	
	<ul style="list-style-type: none"> • Catégories sémantiques : UP d'activité mentale, en particulier d'opportunité (par ex., <i>donner Dét possibilité + PSub</i>) • UP types : <i>prendre Dét décision PSub</i> (vs la structure collocationnelle avec un modifieur quelconque, plus fréquente en L1), <i>donner Dét possibilité PSub</i> 	<ul style="list-style-type: none"> • UP types : <i>donner Dét impression PSub, donner Dét occasion PSub</i>
	Structures collocationnelles prépositionnelles	
<ul style="list-style-type: none"> • Catégories sémantiques : UP d'activité mentale • UP types : <i>prendre + en compte, prendre + par exemple</i> 	/	

Tableau 93. Structures collocationnelles dans les corpus argumentatifs non natifs, présentant un écart de fréquence par rapport à leur fréquence d'emploi dans le corpus natif

En terme de mésusage, certaines UP déviantes ont été mises en avant dans le point précédent, en fonction de la structure du SN de l'UP concernée. Nous mettons ici en exergue ici les quelques cas d'UP déviantes qui se démarquent le plus dans les productions des LNN :

- Les UP avec déterminant « zéro », à contenu prédicatif d'activité mentale ou d'événement, se caractérisent par l'absence de déviations en français L2, ce qui nous amène à penser que, malgré que les LNN sous-emploient les UP avec déterminant « zéro », ils le font avec une *précision* maximale quand ils utilisent des UP appartenant aux catégories sémantiques les plus fréquentes en français argumentatif.
- Parmi les UP dont la structure du SN est en *N de N*, nous n'avons observé chez les LNN anglophones qu'une seule UP d'activité mentale déviante (sur 3 occ.). Les LNN néerlandophones produisent quant à eux une proportion plus importante d'UP déviantes (à structure en *N de N* et à prédicat d'activité mentale), avec 5 UP déviantes sur 13 UP, dont 4 construites avec le verbe *donner* (cf. p. 301): *donner + conséquence* (cf. Ex. 234 : déviation formelle sur le verbe), *exemple* (cf. Ex. 127 et 235 : déviations quantitatives sur le verbe) et *possibilité* (cf. Ex. 236 : déviation formelle sur le complément du nom). Nous avons également noté, pour ce type de structure collocationnelle *en N de N*, les 3 occurrences déviantes de l'UP *prendre Dét *moment de réflexion* dans le corpus néerlandophone.
- Parmi les UP construites avec une proposition subordonnée complément du nom, dont le contenu sémantique est une activité mentale, seules deux UP sont déviantes. Ces UP déviantes sont produites par les LNN anglophones (sur 8 occ.): *prendre + décision* (cf. Ex. 112 : déviation formelle multiple sur le déterminant et la structure du complément du nom) et *prendre + possibilité par les deux mains* (cf. Ex. 106 : déviation contextuelle sur toute l'UP). Considérant uniquement les UP d'opportunité, nous avons constaté que ces UP étaient rarement déviantes avec une structure propositionnelle complément du nom, alors que 4 UP d'opportunité sur 5, dans le corpus anglophone, sont déviantes avec un modifieur quelconque (3 occ. de *donner + choix* et 1 occ. de *donner + occasion*).
- Les structures prépositionnelles sont quant à elles toutes construites avec le verbe *prendre* et sont rarement déviantes en L2 : une seule UP est déviante contextuellement dans le corpus néerlandophone (sur 4 occ.). En outre, les LNN anglophones utilisent une proportion plus importante d'UP prépositionnelles que les LN, dont aucune n'est déviante. Comme c'était le cas des structures avec déterminant « zéro » pour les LNN en général, nous pouvons dire que les productions des LNN anglophones se caractérisent par une *précision* maximale dans l'emploi de ces structures collocationnelles. Notons que nous avons par ailleurs relevé trois déviations (sur 28

occ.) dans le corpus de français L1, dont deux portant sur des UP construites avec le mot *exemple* (cf. Ex. 267, *prendre + comme exemple du contraire* : déviance formelle sur le complément du nom ; Ex. 268, *prendre + en exemple* : déviance quantitative sur la préposition).

Rappelons que le nombre d'occurrences phraséologiques dans les corpus argumentatifs étant restreint, il n'est pas possible de tirer des conclusions généralisables à partir des résultats présentés plus haut. Nous pensons néanmoins que ces résultats fournissent une base solide de travail pour l'étude des UP construites avec des verbes à haute fréquence en français L2. Nous proposons en outre, dans le dernier chapitre de ce travail (en VII.3.), quelques pistes de réflexion sur les applications didactiques envisageables à partir de l'étude descriptive que nous avons faite du phénomène phraséologique en français L2.

Chapitre VII : Synthèse et perspectives

Ce chapitre conclusif a pour objectifs de présenter une synthèse des résultats obtenus tout au long de la présente étude (cf. VII.1.) et de proposer quelques pistes explicatives (cf. VII.2.) et didactiques (cf. VII.3.) qui ont émergé de ces résultats.

VII.1. Synthèse des résultats

VII.1.1. Etude pilote sur le degré de figement (français L1)

VII.1.2. Analyse contrastive interlangagière sur corpus d'apprenants informatisés (français académique L2 vs L1)

VII.1.2.1. Analyse collocationnelle (français académique L2 vs L1)

VII.1.2.2. Analyse de déviance (français académique L2 vs L1)

VII.1.2.3. Analyse de précision et de complexité phraséologiques (français argumentatif L2 vs L1)

VII.1.2.4. Analyse de structures collocationnelles (français argumentatif L2 vs L1)

VII.2. Perspectives explicatives

VII.2.1. Transfert des verbes nucléaires

VII.2.2. Complexité linguistique du phénomène phraséologique dans la LC

VII.2.3. Facteurs extralinguistiques

VII.3. Perspectives didactiques

VII.1. Synthèse des résultats

Afin de donner une vision globale des résultats qui ont émergé des différentes phases de notre recherche, nous nous proposons d'opérer une synthèse de ces résultats en suivant l'ordre des étapes majeures qui ont balisé notre travail: (i) l'étude pilote, en français L1, sur le degré de figement de 200 séquences verbo-nominales (SVN) construites avec le verbe à haute fréquence *prendre* (analyse sur corpus en français L1, sous I.1.4.) ; (ii) l'étude collocationnelle et l'analyse de déviance des 906 SVN construites avec les verbes à haute fréquence *prendre* et *donner*, dont 703 unités phraséologiques (UP), produites par des apprenants avancés (anglophones et néerlandophones) du français académique L2 et par des locuteurs francophones natifs (analyse contrastive interlangagière sur corpus informatisés – cf. Chapitres IV et V) ; (iv) l'analyse de complexité (sous VI.1.) et de précision (sous VI.2.1. et VI.2.2.) des 503 UP extraites des textes académiques à visée argumentative en français L2 (207 UP) et L1 (296 UP) ; (v) l'analyse des structures collocationnelles les plus représentatives des interlangues (IL) avancées en français argumentatif L2 (sous VI.2.3.), parmi les 207 UP avec les verbes *prendre* et *donner* qui sont produites par les apprenants non natifs du français L2.

VII.1.1. Etude pilote sur le degré de figement (français L1)

Cette étude paramétrique et statistique a été menée sur 200 SVN construites avec le verbe à haute fréquence *prendre*, dans des textes journalistiques et littéraires de français L1 (400.000 mots). Les résultats de cette étude ont corroboré, d'une part, l'hypothèse de la gradation du phénomène de figement et, d'autre part, la corrélation forte entre les contraintes syntaxique, sémantique et lexicale qui sont en jeu dans le processus de figement. C'est ainsi que l'occurrence de la séquence *prendre le taureau par les cornes* se définit par un haut degré de figement global (87%), alors que *prendre acte de N* se définit par un degré moyen de figement (56%) et *prendre un fruit* par un degré faible de figement (16%). L'étude a par ailleurs montré qu'à une même séquence *type* pouvaient correspondre des occurrences caractérisées par un degré de figement variable, allant par exemple de 39% à 71% pour les occurrences construites avec le verbe *prendre* et le nom *forme*. En français L1, les UP construites avec le verbe *prendre* se caractérisent donc par leur dimension graduelle et polyfactorielle, ce qui empêche de poser les frontières du phénomène phraséologique de manière fixe, claire et univoque, tant au niveau supérieur de figement (entre les UP semi-figées et les UP très figées) qu'au niveau inférieur du figement (entre les UP semi-figées ou quasi-figées et les séquences libres).

VII.1.2. Analyse contrastive interlangagière sur corpus d'apprenants informatisés (français académique L2 vs L1)

Cette étude s'est concentrée sur l'aspect contrastif de l'acquisition des UP construites avec les deux verbes à haute fréquence *prendre* et *donner*, dans des textes académiques d'apprenants avancés du français L2. Certaines caractéristiques propres au français académique L2, ainsi que les difficultés spécifiques aux apprenants avancés anglophones vs néerlandophones dans leur production d'unités phraséologiques, ont ainsi été mises en évidence.

VII.1.2.1. Analyse collocationnelle (français académique L2 vs L1)

Une analyse collocationnelle des 906 SVN extraites des corpus de français L2 et L1 a été effectuée, afin de déterminer le statut phraséologique ou libre de ces séquences. Cette analyse s'est basée sur trois types de ressources disponibles pour le français L1: (i) une

vaste base de données phraséologiques établie pour les expressions verbales figées en français dans le cadre du lexique-grammaire de M. Gross (1975, 1988) ; (ii) une base distributionnelle de données lexicales constituée à partir du quotidien français *Le Monde* (*Les Voisins De Le Monde*, cf. Fabre & Bourigault, 2006), qui calcule notamment l'indice d'association lexicale de cooccurrents syntaxiques (prédicat + argument) ; (iii) des corpus intégraux, journalistique (*Le Soir*) et oral (*Valibel*, cf. Francard *et al.*, 2002), en français L1, permettant de vérifier le caractère attesté ou non des UP dans l'usage courant. Cette méthode a permis d'appréhender de manière la plus exacte et objective possible l'usage que font les francophones natifs des UP. Elle a par ailleurs l'avantage de combiner deux approches complémentaires en phraséologie : (i) l'approche fonctionnelle, qui repose sur des critères linguistiques de figement (syntaxiques, sémantiques et lexicaux), hérités de la linguistique théorique ; (ii) l'approche statistique, qui repose sur des mesures d'association lexicale entre termes (e.a. par le calcul d'un indice d'information mutuelle), héritée de la linguistique de corpus.

En étudiant l'influence de variables externes (langue source, type de textes, medium oral/écrit) et internes (structure formelle du syntagme nominal, catégorie sémantique de la séquence) sur la production d'unités, un sous-emploi des unités phraséologiques construites avec le verbe *prendre* et un suremploi des unités phraséologiques construites avec *donner* a été constaté dans l'ensemble des textes non natifs ($p < *0.005$). Ces résultats se sont ensuite vus nuancés, surtout concernant le corpus anglophone, par un examen plus approfondi de l'impact du type de textes sur la production d'UP: dans les textes strictement argumentatifs, appartenant à la catégorie des textes académiques, les locuteurs non natifs (LNN) anglophones tendent à se rapprocher de la compétence native, puisqu'ils emploient les UP construites avec le verbe *prendre* et *donner* dans une proportion comparable à celle des locuteurs natifs (LN).

Sur le plan structurel, il ressort une tendance significative ($p = *0.000$) des LNN anglophones du français académique L2 à sous-employer certaines structures dans les UP qu'ils produisent – i.e. les structures sans déterminant (i.e. avec déterminant « zéro »), les structures en *N de N* et les structures prépositionnelles –, et à en suremployer d'autres – i.e. les structures avec un déterminant et un modifieur quelconques, ainsi que les structures nominales avec proposition subordonnée. Ce sous-emploi et ce suremploi seraient en fait dus à une différence d'emploi des UP construites avec le verbe *prendre*, à l'inverse des UP avec *donner* qui montrent une construction syntaxique du syntagme nominal (SN) similaire à celle des natifs. Cette différence entre LNN anglophones et LN,

très significative dans les textes académiques, reste relativement similaire quand on ne considère que les textes argumentatifs (503 UP), à l'exception des structures du SN avec proposition subordonnée (suremployées dans les textes académiques) et les structures prépositionnelles (sous-employées dans les textes académiques) qui montrent, dans les textes argumentatifs, une fréquence d'emploi proche de celle des LN.

La différence d'emploi des UP par les LNN néerlandophones en fonction de la structure syntaxique du SN s'est également avérée significative, bien que de manière moins marquée ($p < *0.05$) par rapport aux apprenants anglophones. Cette différence entre LNN néerlandophones et LN, légèrement significative dans les textes académiques (703 UP), est encore atténuée quand on compare entre eux les textes exclusivement argumentatifs (503 UP): dans les textes néerlandophones strictement argumentatifs, la répartition des types de structures du SN des UP reste sensiblement la même que dans l'ensemble des textes académiques, avec néanmoins une légère réduction de l'écart de fréquence entre les LNN et les LN. Seul le suremploi des structures avec proposition subordonnée complément du nom est davantage marqué dans les textes néerlandophones argumentatifs que dans l'ensemble des textes académiques.

De manière générale, les UP avec *prendre* montrent par ailleurs, dans l'ensemble des corpus académiques non natifs anglophones et néerlandophones, surtout des structures marquées en *N de N* ou prépositionnelles (*Prép + SN*). Les UP avec *donner* présentent plutôt des structures marquées avec proposition subordonnée complément du nom (*N + P Sub*) ou avec déterminant « zéro ».

Sur le plan sémantique, l'ensemble des données phraséologiques (703 UP) est constitué d'une majorité d'emplois prédicatifs, avec principalement des activités mentales, puis des états et, enfin, des événements. Deux tendances ressortent de cette étude : il y a un suremploi, par les néerlandophones, des UP d'événement et un sous-emploi d'UP d'activité mentale ; les textes narratifs, qui ne représentent qu'1/5 du corpus anglophone, comptent majoritairement des emplois prédicatifs d'état.

Enfin, nous avons vu que le code langagier oral (vs écrit), à l'inverse du type de textes, n'influçait pas de manière importante la production d'UP construites avec les verbes *prendre* et *donner* par les apprenants non natifs : les LNN et les LN ont recours, les uns comme les autres, à des UP spécifiques au médium oral dans leurs productions écrites, avec une plus haute fréquence de ces emplois par les LN, puis par les LNN anglophones et, enfin, par les LNN néerlandophones.

VII.1.2.2. Analyse de déviance (français académique L2 vs L1)

Une analyse de déviance portant sur l'ensemble des UP (703 occ.) a été réalisée. Cette analyse de déviance a permis de décrire les déviances en les catégorisant selon: (i) le *type* de déviance (formelle, contextuelle ou quantitative) ; (ii) la *localisation* de la déviance (verbe+nom, verbe, nom, déterminant ou modifieur) ; (iii) le degré de *gravité* de la déviance, tenant compte du *type* de déviance en cause, de la *localisation* de cette déviance, du *nombre* de déviances par unité phraséologique, du degré de *collocabilité* (en français L1) et de la *fréquence* (en français L1) de l'unité phraséologique déviante.

Les LN ne commettent que peu de déviances (6,8% d'UP déviantes), par comparaison avec les apprenants qui produisent 33,7% de déviances dans le corpus anglophone et 27,7% dans le corpus néerlandophone. En termes de *type* de déviances, il ressort que 40% des déviances natives sont quantitatives – i.e. attestées formellement dans l'usage L1 et appropriées en contexte, mais non conventionnelles (non préférées par les natifs) –, alors que la majorité des déviances produites par les apprenants du français L2 sont formelles (non attestées dans l'usage natif).

En termes de *localisation* de la déviance, les déviances portent principalement sur l'élément verbal, pour plus de 25% des cas pour chacun des trois corpus. Dans les corpus d'apprenants du français L2, les déviances les plus importantes sont secondairement les déviances portant sur le nom, avec respectivement 11% de UP déviantes dans le corpus anglophone (10 occ.) et 16% dans le corpus néerlandophone (6 occ.). Les néerlandophones produisent par ailleurs 11% des déviances (4 occ.) portant sur le complément du nom (i.e. le modifieur adjectival ou propositionnel).

En termes de degré de *gravité* de la déviance, il ressort que les UP déviantes les plus graves ne concernent que de rares cas (moins de 10%) dans les corpus L2 et L1. Les UP faiblement déviantes concernent quant à elles 35% des UP déviantes chez les natifs, 18,7% des UP déviantes chez les anglophones et 10,5% chez les néerlandophones. La proportion d'UP déviantes la plus importante dans l'ensemble des corpus concerne les UP dont le degré de gravité est moyen, représentant 60% des UP déviantes chez les natifs, 74,7% des UP déviantes chez les anglophones et 86,8% des UP déviantes chez les néerlandophones.

VII.1.2.3. Analyse de précision et de complexité phraséologiques (français argumentatif L2 vs L1)

En mesurant le taux de précision des productions non natives et natives, nous avons constaté que les LNN néerlandophones étaient plus proches de la compétence phraséologique native que les LNN anglophones dans leur production d'UP construites avec les deux verbes *prendre* et *donner*. Une mesure de précision prenant en compte le degré de gravité des UP déviantes a en outre montré que les LNN néerlandophones commettaient plus d'UP gravement déviantes que les anglophones. Ce facteur qualitatif de gravité ne s'est toutefois pas avéré suffisant pour influencer le niveau général de compétence phraséologique des LNN : les productions phraséologiques des LNN néerlandophones, dans les textes académiques à visée argumentative, sont plus précises que celles des LNN anglophones.

Par le biais d'une mesure de complexité structurelle du SN des UP, il s'est avéré que les LNN néerlandophones faisaient preuve d'une meilleure compétence phraséologique que les LNN anglophones à cet égard : les productions phraséologiques des LNN néerlandophones, dans les textes académiques à visée argumentative, sont plus complexes structurellement que celles des LNN anglophones. En termes de diversité lexicale, les UP produites par les LNN anglophones sont lexicalement moins riches, i.e. plus contraintes lexicalement, que les UP produites par les LNN néerlandophones, pour les UP construites avec *prendre* et avec *donner*. En outre, les UP construites avec le verbe *donner*, dans les corpus anglophone et natif, sont plus riches lexicalement et donc plus libres en termes de combinatoire lexicale que les UP construites avec *prendre*, dont le potentiel combinatoire est plus restreint. Chez les LNN néerlandophones, les UP construites avec *prendre*, bien que légèrement moins riches lexicalement que celles avec *donner*, sont plus diversifiées lexicalement et donc moins contraintes lexicalement que chez les natifs.

VII.1.2.4. Analyse de structures collocationnelles (français argumentatif L2 vs L1)

Dans les textes argumentatifs, nous avons mis le doigt sur quelques *structures collocationnelles* soit plus fréquentes, soit ayant tendance à être évitées en français L2, qui tendent à caractériser, d'une part, les deux interlangues (IL) avancées du français argumentatif L2 l'une par rapport à l'autre et, d'autre part, les deux IL avancées du français argumentatif L2 par opposition au français argumentatif L1.

L'IL avancée anglophone se caractérise par une tendance à éviter les UP d'événement (achèvement) avec les verbe *donner* dont la structure est un déterminant « zéro », mais à utiliser plus fréquemment que les LN des UP avec une proposition subordonnée complément du nom, qui ont un contenu prédicatif d'activité mentale exprimant l'opportunité (par ex., *donner Dét possibilité + PSub*) ou qui sont construites avec le verbe *prendre* et le nom *décision PSub* (par contraste avec les LN qui emploient davantage d'UP avec *prendre* et *décision* avec un modifieur quelconque). Les LNN anglophones emploient également plus fréquemment des UP prépositionnelles d'activité mentale que les LN, notamment *prendre + en compte* et *prendre + par exemple*. En termes de mésusage, considérant les UP d'opportunité avec un modifieur quelconque, nous avons en outre observé un certain nombre d'UP déviantes (*donner + choix* et *donner + occasion*) dans les textes argumentatifs anglophones. Par contraste, les LNN anglophones, qui utilisent une proportion plus importante d'UP prépositionnelles que les LN, les utilisent toutes de manière correcte.

L'IL avancée néerlandophone se caractérise quant à elle notamment par une tendance au sous-emploi des structures prépositionnelles d'activité mentale, telle que l'UP type *prendre + en compte*. À l'inverse, les LNN néerlandophones emploient plus fréquemment que les LN les UP types avec déterminant « zéro » *donner lieu*, en *N de N* *prendre Dét cas de N* et *prendre Dét moment de réflexion*, avec proposition subordonnée *donner Dét impression PSub* et *donner Dét occasion PSub*. En termes de mésusage, les LNN néerlandophones produisent une proportion importante d'UP d'activité mentale en *N de N* déviantes, dont une majorité avec le verbe *donner* (*donner + conséquence/ exemple/ possibilité*). Nous avons également noté que les 3 occurrences de l'UP *prendre Dét *moment de réflexion* étaient déviantes.

Les IL avancées, anglophones et néerlandophones, du français argumentatif L2 (vs L1) se caractérisent toutes les deux par une tendance à éviter l'UP type avec déterminant « zéro » *prendre conscience de*, la structure collocationnelle *prendre Dét exemple de N* (en particulier la forme impérative *prenons Dét exemple de N*), les UP types *prendre Dét temps de PSub* et les UP à structure prépositionnelle (*prendre + en charge*, *prendre + pour/comme exemple* et *prendre + pour modèle*). En termes de mésusage, les LNN, qui sous-emploient les UP avec déterminant « zéro », les utilisent néanmoins avec une *précision* maximale quand elles véhiculent un sens d'activité mentale ou d'événement (aucune de ces UP n'est déviante en L2). De même, les structures prépositionnelles, qui sont toutes construites avec le verbe *prendre*, sont rarement déviantes en L2.

Ces résultats, concernant les textes argumentatifs d'apprenants du français L2, portent sur un nombre limité d'occurrences et ne sont donc pas généralisables. Néanmoins, ils donnent un aperçu de ce qui tend à distinguer les IL avancées du français L2, en termes de fréquence d'emploi phraséologique par rapport à l'usage natif. Ces résultats ont en outre l'avantage de mettre en avant des structures collocationnelles qui ont été repérées dans des productions d'apprenants authentiques, reflétant de manière concrète et empirique l'état de la compétence phraséologique de ces apprenants du français L2 à un moment donné de leur itinéraire acquisitionnel.

En conclusion, en termes de production d'UP verbo-nominales construites avec les verbes *prendre* et *donner* dans les textes académiques de français L2, les LNN anglophones semblent se caractériser par un niveau de compétence phraséologique inférieur à celui des LNN néerlandophones. Les LNN anglophones produisent en effet une proportion plus importante d'UP déviantes que les néerlandophones. De plus, la différence entre la structuration du SN des UP en L2 et L1 est statistiquement plus significative pour les LNN anglophones que pour les LNN néerlandophones. Deux traits relativisent toutefois ce point de vue : le degré de gravité des UP déviantes produites par les LNN anglophones est inférieur à celui des UP produites par les LNN néerlandophones ; les LNN anglophones utilisent des catégories sémantiques d'UP dans une proportion proche de celle des LN, ce qui n'est pas le cas des LNN néerlandophones.

Dans les textes à visée argumentative, ce sont également les LNN anglophones qui se caractérisent par une compétence phraséologique moins avancée par rapport aux néerlandophones. Cette différence est surtout visible quand on mesure le degré de complexité phraséologique et le taux de précision phraséologique des productions non natives anglophones *vs* néerlandophones. Néanmoins, notons que la différence, qui est significative concernant le suremploi des UP avec *donner* et le sous-emploi des UP avec *prendre* dans l'ensemble des textes académiques en L2, ne l'est plus quand on considère les productions des LNN anglophones dans les textes strictement argumentatifs.

Mais comment peut-on expliquer les traits phraséologiques qui différencient ces deux variétés avancées du français L2, anglophone et néerlandophone ? Nous proposons ci-dessous quelques pistes de réflexion qui ont émergé au fil de nos recherches et qui nous paraissent les plus susceptibles d'expliquer (du moins en partie) les résultats que nous venons de synthétiser.

VII.2. Perspectives explicatives

Nous présentons ici quelques axes de recherche qui pourraient étayer les observations faites ci-dessus en leur fournissant une base explicative.

VII.2.1. Transfert des verbes nucléaires

Chez les apprenants d'une L2, le transfert joue un rôle important dans l'emploi des verbes *nucléaires*, i.e. des verbes de base (fréquents dans l'usage) qui tendent à avoir des équivalents sémantiques dans d'autres langues (Viberg, 2002b), comme c'est le cas des verbes de possession *prendre* et *donner* en français, qui ont pour équivalents *take* et *give* en anglais et *nemen* et *geven* en néerlandais. Comme nous l'avons vu précédemment, les *verbes nucléaires* se distinguent néanmoins d'une langue à l'autre par de légères différences sémantiques et des différences dans leur potentiel combinatoire au niveau lexical, ce qui peut donner lieu au suremploi ou au sous-emploi des verbes de la langue cible (LC) par l'apprenant avancé (Altenberg & Granger, 2001 ; Viberg, 2002a, 2002b).

Selon cette perspective, le suremploi des UP construites avec le verbe *donner* par les LNN, dans les textes académiques, pourrait être lié à une extension du potentiel polysémique et de la combinatoire lexicale des verbes *give* et *geven* : ces verbes seraient plus polysémiques et entreraient en combinaison avec un plus grand nombre d'unités lexicales nominales *types* différentes, en anglais et en néerlandais, qu'en français. Cette hypothèse est néanmoins mise à mal quand on observe les résultats de l'analyse de *diversité lexicale* des UP dans les textes argumentatifs : les UP avec le verbe *donner*, en français argumentatif, sont moins riches lexicalement en L2 qu'en L1, ce qui semble contredire à première vue l'hypothèse de l'extension du potentiel combinatoire – supposé plus important – des verbes de la langue source (LS) sur les formes verbales de la LC – dont le potentiel combinatoire est supposé être plus restreint. Notons par ailleurs que le suremploi du verbe *donner*, dans les textes argumentatifs, ne concerne pas le corpus d'apprenants anglophones, ces derniers se rapprochant de l'usage natif en termes de fréquence d'emploi des UP avec *prendre* et avec *donner*.

Quant au sous-emploi des UP avec le verbe *prendre* dans l'ensemble des textes académiques, il pourrait être lié à une différence de potentiel phraséologique, en termes de combinatoire lexicale, entre le verbe de la LS et celui de la LC : *prendre* se définirait ainsi par un potentiel combinatoire plus étendu, en français, que *take* en anglais et *nemen*

en néerlandais, qui seraient plus spécifiques. Selon cette perspective, les LNN produiraient des UP construites avec le verbe *prendre* uniquement quand une forme équivalente, avec *take* ou *nemen*, existe dans la LS. Cette hypothèse ne vaut de nouveau pas pour les textes argumentatifs anglophones, dans lesquels les UP construites avec les deux verbes *prendre* et *donner* ne font l'objet d'aucun suremploi ou sous-emploi. Elle ne semble pas non plus s'appliquer aux textes argumentatifs d'apprenants néerlandophones, qui sous-emploient les UP avec le verbe *prendre* et suremployent les UP avec *donner*, mais qui montrent un emploi des UP construites avec *prendre* plus diversifié lexicalement que chez les LN. Ces observations sèment donc le doute quant au possible transfert du potentiel sémantique et phraséologique, d'une part, de *take/nemen* sur le verbe *prendre* et, d'autre part, de *give/geven* sur le verbe *donner*. Rappelons que nous avons émis précédemment plusieurs hypothèses (e.a. celle du transfert) pouvant expliquer la différence de *complexité structurelle* et *lexicale*, d'une part, entre les UP construites avec *prendre* vs *donner* et, d'autre part, entre les emplois non natifs vs natifs, dans les textes spécifiquement argumentatifs. Nous ne reviendrons pas ici sur ces hypothèses, mais renvoyons le lecteur à cette partie de notre travail (cf. en VI.3., pp. 311-313) où est abordée la question du rôle que certains facteurs d'ordre plus psychologique (e.a. l'attitude et la motivation de l'apprenant) sont susceptibles de jouer dans la production d'UP en L2.

Le transfert peut également jouer un rôle important dans la production de *structures collocationnelles*, telles que les UP avec proposition subordonnée complément du nom exprimant l'opportunité (par ex., *donner Dét possibilité + PSub*) ou incluant le nom *décision*, qui sont fréquentes dans le corpus argumentatif anglophone (par opposition au corpus de français L1). Les LNN anglophones emploient également plus fréquemment des UP prépositionnelles d'activité mentale que les LN (*prendre + en compte ; prendre + par exemple*). Ce type de structures collocationnelles seraient-elles plus spécifiques à l'anglais qu'au français argumentatif L1 ? Un examen du comportement phraséologique des verbes *take* et *give* en anglais argumentatif L1 devrait pouvoir apporter des éléments de réponse à cette question.

VII.2.2. Complexité linguistique du phénomène phraséologique dans la LC

Rappelons que la *complexité* des UP construites avec les verbes à haute fréquence *prendre* et *donner* résiderait notamment dans le potentiel plus ou moins phraséologique de ces verbes en français L1. L'hypothèse, selon cette perspective « native », est que les apprenants avancés tendraient à éviter certaines structures phraséologiques plus complexes et/ou à suremployer les structures phraséologiques plus simples. Dans le cas de l'évitement, on peut raisonnablement supposer qu'il y aurait une prise de conscience par l'apprenant de la difficulté du phénomène phraséologique, du point de vue de la combinatoire lexico-grammaticale de certaines UP. En français académique et argumentatif L2, nous avons ainsi vu que les LNN anglophones sous-employaient les UP avec déterminant « zéro » et les structures en *N de N*, considérées comme des structures contraintes (plus *complexes*), alors qu'ils suremployaient les structures avec un déterminant et un modifieur quelconques, moins *complexes*. Le verbe *prendre*, qui est structurellement plus complexe et moins riche lexicalement que le verbe *donner* en L1, est quant à lui sous-employé dans les textes académiques anglophones et néerlandophones. La *complexité* lexicale irait ici de pair avec une plus grande liberté combinatoire, autrement dit, avec une restriction sur les paradigmes synonymiques moins forte. Les UP moins riches lexicalement seraient donc aussi les UP les plus « phraséologiques », ce qui expliquerait que les UP avec le verbe *prendre* (moins riches lexicalement et plus complexes structurellement en L1) seraient sous-employées par les apprenants anglophones et néerlandophones en français académique L2. De nouveau, cette hypothèse ne vaut pas pour les textes argumentatifs anglophones, qui présentent une proportion équivalente d'UP avec *prendre* et *donner* par rapport au corpus natif.

VII.2.3. Facteurs extralinguistiques

Enfin, certains facteurs extralinguistiques sont à prendre en compte pour expliquer l'écart entre les productions non natives anglophones vs néerlandophones. Premièrement, le degré de xénité plus important (au niveau géographique et sociopolitique) des apprenants anglophones par rapport à celui des néerlandophones de Belgique est un des facteurs qui nous semble expliquer de manière la plus directe la différence de niveau de compétence phraséologique entre les deux IL avancées du français L2 : les néerlandophones de Belgique, dont la compétence phraséologique générale semble plus proche de celle des

francophones natifs, sont plus susceptibles d'être en contact avec la langue française que les anglophones du Royaume-Uni, qui montrent une compétence phraséologique moindre. Deuxièmement, il serait intéressant de savoir si la motivation de l'apprenant est plutôt *intégrative* ou *instrumentale*. Le premier type de motivation implique en effet que l'apprenant ait pour objectif la maîtrise native de la LC, ce qui favoriserait un niveau plus avancé de compétence phraséologique, alors que le second type de motivation implique la limitation par l'apprenant de son objectif acquisitionnel à la réussite de l'acte de communication langagière. Selon ce facteur, nous pouvons supposer que les LNN néerlandophones se caractériseraient par une motivation *intégrative*, contrairement aux anglophones qui se caractériseraient par une motivation *instrumentale*. Troisièmement, l'attitude de ces apprenants non natifs serait intéressante à étudier, pour déterminer dans quelle mesure ceux-ci osent prendre des risques pour s'exprimer dans la L2, sachant qu'une prise de risques plus grande favoriserait des productions plus *complexes* mais plus déviantes, contrairement aux apprenants qui privilégient la *précision* et produiraient donc à la fois moins d'UP, puisque ces unités sont *complexes*, et moins de déviances. Cette dernière hypothèse pourrait en partie expliquer la production plus importante et plus complexe (structurellement et lexicalement) d'UP par les LNN néerlandophones qui, bien que produisant moins d'UP déviantes que les LNN anglophones, montrent un degré de gravité des UP déviantes plus élevé que ces derniers.

A partir des écarts de fréquence décelés entre les corpus non natifs et les corpus natifs, des perspectives didactiques peuvent être envisagées.

VII.3. Perspectives didactiques

Although collocations have received increasing attention in language teaching in recent years (Granger 1998c:159; Howarth 1998a:30), we are still far from the development of a coherent methodology and even further from a wide-spread and systematic treatment of collocations in language teaching materials and syllabi (...). (Nesselhauf, 2005: 3)

Suite aux observations qui ont émané de notre recherche, nous pensons que certains critères sont à prendre en compte pour l'élaboration d'une *norme pédagogique* (NP) destinée aux apprenants avancés du français académique L2, qui visent à acquérir ou améliorer leur compétence phraséologique en situation de production langagière. Les résultats auxquels nous sommes arrivés dans cette étude pourraient en effet s'intégrer dans le processus d'élaboration d'une NP qui tienne compte du niveau avancé

d'acquisition des apprenants du français L2, de leur langue source (LS) respective, du type de texte académique concerné (par ex. le texte argumentatif) et de la problématique de l'apprentissage des UP comme unités linguistiques contraintes sur les plans syntaxique (*fixité*), sémantique (*non-compositionnalité* et *aréférenciation*), lexical (*restrictions paradigmatiques*) et/ou structurel (*structures collocationnelles*). Rappelons que le concept de NP vise à prendre en compte les besoins pédagogiques des apprenants en fonction de leur niveau d'acquisition interlangagier. À un stade avancé d'apprentissage, la nécessité de complexifier les NP s'impose pour que l'apprenant puisse réduire la distance qui le sépare de la compétence native. Le phénomène phraséologique étant considéré comme un trait tardif d'acquisition d'une L2, cette complexification des NP passerait par la prise de conscience de la complexité du phénomène phraséologique dans la LC (par opposition au fonctionnement des UP dans la LS).

Tout d'abord, l'enseignant comme l'apprenant devraient idéalement avoir une idée de la fréquence des UP et du caractère plus ou moins conventionnel de celles-ci, afin de focaliser l'enseignement sur les UP que l'apprenant est le plus susceptible de rencontrer dans l'usage courant (selon le critère de fréquence) et de produire de manière la plus conventionnelle (selon le critère de collocabilité) en situation de communication langagière. Les UP déviantes *quantitatives*, dont la fréquence d'emploi et/ou dont l'indice d'association lexicale est inférieur à celui des natifs, sont très représentatives de cette étrangeté qui ressort de certaines productions d'apprenants non natifs avancés, acceptables formellement et appropriées contextuellement, mais qui donnent une impression que la maîtrise de la LC n'est pas encore véritablement *native*.

Ensuite, il semble indispensable de savoir quelles sont les UP conventionnelles spécifiques au type de texte utilisé et quelle est la fonction que l'on attribue à ces UP. Par exemple, nous avons constaté, dans les corpus argumentatifs non natifs, que la structure collocationnelle *prenons Dét exemple de N* était peu employée, contrairement à l'usage fréquent qui en est fait dans les textes de français L1. Les apprenants avancés du français L2 qui veulent écrire un texte argumentatif devraient intégrer cette structure collocationnelle dans leur répertoire polylexical, en ayant conscience que la forme la plus courante en français argumentatif L1 est à la 1^e pers. pl. de l'impératif, que la structure sélectionnée est une structure en *N de N*, que la fonction de cette UP est une fonction d'organisation langagière visant à illustrer un propos qui vient d'être énoncé et qu'elle est le plus souvent placée en position initiale de phrase. L'apprenant devra également savoir

qu'il existe une UP équivalente dont la fonction est identique, *prendre Dét cas de N*, mais qui est moins courante en français L1 que la précédente.

Cet exemple d'application didactique imaginée à partir des résultats de notre recherche pourrait être étendu à un ensemble plus vaste d'UP, par exemple aux UP qui sont construites avec d'autres verbes nucléaires (*faire, aller, venir, dire, voir, savoir*, etc.) ou qui appartiennent à d'autres catégories discursives (par ex. nom + adjectif/adverbe). Il serait également intéressant de proposer des pistes didactiques qui distingueraient les emplois phraséologiques spécifiques à différents types de textes académiques (argumentatif, texte de synthèse, persuasif, comparatif, etc.), ce qui mettrait à la disposition des apprenants avancés des outils appropriés aux besoins de la situation de communication langagière dans laquelle ils se trouvent.

Après avoir opéré un tour d'horizon rétrospectif de l'ensemble des résultats qui ont été mis en avant dans ce travail et avoir fait un saut prospectif vers quelques pistes explicatives et didactiques, nous cédon maintenant la place à la conclusion générale qui donne à la présente recherche sa finitude matérielle et son cadre temporellement borné.

Conclusion

Nous avons formulé, dans l'introduction générale de ce travail, la question suivante :

Les unités phraséologiques : un phénomène linguistique complexe à décrire et à acquérir?

Au vu des résultats obtenus, nous pouvons désormais affirmer « sans complexe » que les unités phraséologiques (UP) sont bien des phénomènes linguistiques *complexes*, tant du point de vue de leurs caractéristiques linguistiques, polyfactorielles et graduelles, que du point de vue de la difficulté que leur apprentissage sous-tend jusqu'à un stade avancé d'acquisition du français L2. Nous pourrions ajouter que la *complexité* intrinsèque à ce phénomène linguistique oblige le chercheur à développer une méthodologie elle-même *complexe*, qui lui permette d'appréhender les unités phraséologiques de la manière la plus objective, exhaustive et rigoureuse possible.

Outre l'intérêt méthodologique de cette étude, nous espérons avoir montré qu'il était nécessaire que des études descriptives du phénomène phraséologique, reposant sur l'examen de données authentiques de corpus, soient entreprises pour le français L2. Ces études descriptives constituent en effet une étape indispensable à l'élaboration d'applications didactiques dans le domaine de la phraséologie, visant à améliorer les techniques d'enseignement, mais aussi à favoriser la création ou l'adaptation du matériel didactique en fonction des besoins communicationnels réels des apprenants avancés du français L2.

Dans son étude sur la collocation en anglais académique L2, Nesselhauf (2005: 8) a souligné la nécessité de déterminer (i) quelles collocations sont les plus problématiques pour certains groupes d'apprenants, (ii) quelles sont les déviances produites et (iii) pourquoi ces dernières sont produites. Par le biais de cette recherche doctorale, nous avons tenté de répondre aux deux premières questions (quelles collocations et quelles déviances ?), ouvrant ainsi la voie à des perspectives explicatives qui répondraient à la troisième question (pourquoi ?). Une quatrième question (comment les enseigner ?) ouvrirait quant à elle la voie aux perspectives didactiques qui pourraient être imaginées dans le futur, notamment pour le français académique L2. D'autres questions ont certainement leur place ici, que nous laissons à d'autres chercheurs le soin d'imaginer.

Bibliographie

- Altenberg, B. (1998) « On the phraseology of spoken English: The evidence of recurrent word-combinations ». In A. P. Cowie (éd.), *Phraseology: Theory, Analysis, and Applications*, Oxford: Oxford University Press, 101-122.
- & Granger, S. (2001) « The grammatical and lexical patterning of *make* in native and non-native student writing ». *Applied Linguistics* 22/2, 173-194.
- Andersen, H. L. (2007) « Marqueurs discursifs propositionnels ». In G. Dostie (éd.), *Les marqueurs discursifs, Langue française* 154, 13-28.
- Anscombe, J.-Cl. (2003) « Les Proverbes sont-ils des expressions figées? ». *Cahiers de Lexicologie* 82/1, 159-173.
- Baarda, D. B., De Goede, M. P. M. & Van Dijkum, C.J. (2004) *Introduction to Statistics with SPSS – A Guide to the Processing, Analysing and Reporting of (Research) Data*. Groningen/ Houten (The Netherlands): Wolters-Noordhoff.
- Balibar-Mrabti, A. & Vaguer, C. (2005) « Présentation. Le semi-figement ». *LINX* 53, 7-16.
- Bardovi-Harlig, K. & Gass, S. M. (2002) « Introduction ». In S. M. Gass, K. Bardovi-Harlig, S. S. Magnan & J. Walz (éds), *Pedagogical Norms for Second and Foreign Language Learning and Teaching*, Amsterdam/ Philadelphia: Benjamins, 1-12.
- Bardovi-Harlig, K. (2002) « A new starting point? Investigating formulaic use in future expression ». *Studies in Second Language Acquisition* 24/2, 189-198.
- Bartning, I. (1997) « L'apprenant dit avancé et son acquisition d'une langue étrangère – Tour d'horizon et esquisse d'une caractérisation de la variété avancée ». *AILE* 9, 9-50.
- & Kirchmeyer, N. (2003) « Le développement de la compétence textuelle à travers les stades acquisitionnels en français L2 ». *AILE* 19, 9-39.
- & Schlyter, S. (2004) « Itinéraires acquisitionnels et stades de développement en français L2 ». *French Language Studies* 14, 281-299.
- Ben Amor, T. (2003) « Polylexicalité, polysémie et jeu de mots ». In S. Mejri (éd.), *Polysémie et polylexicalité, Syntaxe et Sémantique* 5, 207-222.
- Binon, J., Verlinde, S. & Selva, T. (2001) « Lexicographie pédagogique et enseignement/apprentissage du vocabulaire en français langue étrangère ou seconde (FLES). Un mariage parfait ». *Cahiers de Lexicologie* 78-1, 41-63.
- Blanche-Benveniste, C., Bilger, M., Rouget, C. & Van Den Eynde, K. (1990) *Le français parlé. Etudes grammaticales*. Paris : CNRS.
- Blumenthal, P. (2005) « Profil combinatoire des mots : analyse contrastive ». In C. Bolly, J. R. Klein & B. Lamiroy (éds), *La Phraséologie dans tous ses états. Actes du colloque « Phraséologie 2005 » (Louvain-la-Neuve, 13-15 octobre 2005)*. Louvain-la-Neuve: CILL 31/2-4, 131-148.
- (2006) « De la logique des mots à l'analyse de la synonymie ». In P. Blumenthal & F. Hausmann (éds), *Collocations, corpus, dictionnaires, Langue française* 150, 14-31.
- Blumenthal, P. & Hausmann, F. (éds) (2006) *Collocations, corpus, dictionnaires. Langue française* 150.
- Bolly, C. (2005) « Au seuil du figement. Séquences (semi) figées avec le verbe à haute fréquence *prendre* en FL1 et FL2 ». In C. Bolly, J. R. Klein & B. Lamiroy (éds), *La Phraséologie dans tous ses états. Actes du colloque « Phraséologie 2005 » (Louvain-la-Neuve, 13-15 octobre 2005)*. Louvain-la-Neuve: CILL 31/2-4, 149-167.
- (à paraître) « The acquisition of phraseological units by advanced learners of French as an L2: High-frequency verbs and learner corpora ». In E. Labeau & F. Myles (éds), *The Advanced Learner Variety: The Case of French*. Bern-Berlin: Peter Lang.
- Bourigault, D., Fabre, C., Frérot, C., Jacques, M.-P. & Ozdowska, S. (2005) « Syntex, analyseur syntaxique de corpus » (*TALN 2005*, Dourdan, 6-10 juin 2005). Disponible à la page [<http://w3.univ-tlse2.fr:8880/erss/index.jsp?perso=bourigault&subURL=TALN05-bourigault-Syntex.pdf>]

- Branca-Rosoff, S. (2007) « Les normes du français en situation publique ». In G. Siouffi & A. Steuckardt (éds), *Les linguistes et la norme. Aspects normatifs du discours linguistique*, Berne: Peter Lang, 21-48.
- Bulté, B. (2007) « Measure for measure: Evaluating quantitative measures of written language proficiency ». In J. Eyckmans, A. Housen, S. Janssens, M. Pierrard, H. Stengers, S. Van Daele, Kuiken, F. et I. Vedder (éds), *Complexity, Accuracy and Fluency in Second Language Use, Learning and Teaching (Book of Abstracts of the Symposium on Complexity, Accuracy and Fluency in Second Language Use, Learning and Teaching, Brussels, Belgium, March 29-30 2007)*, 34-36.
- Cacciari, C. & Glucksberg, S. (1995) « Imaging Idiomatic Expressions: Literal or Figurative Meanings? ». In M. Everaert *et al.* (éds), *Idioms: structural and psychological perspectives*, Hillsdale (New Jersey): Lawrence Erlbaum Associates, 43-56.
- Cassart, A., Dagneaux, E., Granger, S., Husquet, C., Verhulst N. and Watrin, P. (éds) (2002) *Free Text – Deliverable 15 – Three Error-Coded Corpora*. Louvain-la-Neuve: Université catholique de Louvain/ Centre for English Corpus Linguistics.
- Colson, J.-P. (2000) « Les locutions verbales françaises et allemandes dans le discours journalistique : pistes de recherche, fausses pistes, pistes brouillées ». In G. Gréciano (éd.), *Micro- et macrolexèmes et leur figement discursif*, Paris-Louvain : Peeters, 173-189.
- Corder, S. P. (1967, trad. 1980) « Que signifient les erreurs chez les apprenants ? ». *Langages* 57, 9-16.
- _____ (1980) « Dialectes idiosyncrasiques et analyse d'erreurs ». *Langages* 57, 17-28.
- Cowie, A. P. (éd.) (1998) *Phraseology: Theory, Analysis, and Applications*, Oxford: Oxford University Press.
- Cuq, J.-P. (2003) *Dictionnaire de didactique du français, langue étrangère et seconde*. Paris : CLE International.
- _____ & Gruca, I. (2002) *Cours de didactique du français langue étrangère et seconde*. Grenoble : Presses Universitaires de Grenoble.
- De Bot, K., Lowie, W. & Verspoor, M. (2005) *Second Language Acquisition. An advanced resource book*. London & New York: Routledge Applied Linguistics.
- De Cock, S. (2003) *Recurrent Sequences of Words in Native Speaker and Advanced Learner Spoken and Written English: a corpus-driven approach* (Thèse de doctorat non publiée). Louvain-la-Neuve: Université catholique de Louvain.
- _____ (2004) « Preferred sequences of words in NS and NNS speech ». *Belgian Journal of English Language and Literatures*, New Series 2, 225-246.
- _____ & Granger, S. (2004) « High Frequency Words: the Bête Noire of Lexicographers and Learners Alike. A close look at the verb “make” in five monolingual learners dictionaries of English ». In G. Williams & S. Vessier (éds), *Proceedings of the Eleventh EURALEX International Congress*, Université de Bretagne-Sud: Lorient, 233-243.
- Degand, L. & Bestgen, Y. (2004) « Connecteurs et analyses de corpus : de l'analyse manuelle à l'analyse automatisée ». In S. Porhiel & D. Klingler (éds), *L'Unité texte (Actes du colloque 'Regards croisés sur l'Unité texte/Conjoint Perspectives on Text', Chypre, 18, 19, 20 mars 2004)*, 49-73.
- Di Stefano, G. & McGillivray, P.-G. (éds) (1984) *La locution (Actes du Colloque international, université McGill, Montréal, 15-16 octobre 1984)*. Montréal : CERES.
- Dörnyei, Z., Durow, V. & Zarhan, K. (2004) « Individual differences and their effects on formulaic sequence acquisition ». In N. Schmitt (éd.), *Formulaic Sequences. Acquisition, Processing and Use*, Amsterdam/ Philadelphia: Benjamins, 87-106.
- Dubois, J., Giacomo-Marcellesi, M. & Gespin, L. (2001) *Dictionnaire de linguistique*. Paris : Larousse-Bordas/ HER.
- Ellis, R. (1994) *The Study of Second Language Acquisition*. Oxford: Oxford University Press, 47-72.
- _____ & Barkhuizen (2005) *Analysing learner language*. Oxford: Oxford University Press.
- Ellis, N. (2002) « Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition ». *Studies in Second Language Acquisition* 24/2, 143-188.

- _____ (2006) « Selective attention and transfer phenomena in L2 acquisition: contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning ». *Applied Linguistics* 72/2, 164-194.
- Erman, B. & Warren, B. (2000) « The idiom principle and the open choice principle ». *Text* 20/1, 29-62.
- Everaert, M., Van Der Linden, E.-J., Schenk, A. & Schreuder, R. (éds) (1995) *Idioms: structural and psychological perspectives*. Hillsdale (New Jersey): Lawrence Erlbaum Associates.
- Fabre, C. & Bourigault, C. (2006) « Extraction de relations sémantiques entre noms et verbes au-delà des liens morphologiques ». *TALN 2006*.). Disponible en ligne à la page [<http://w3.univ-tlse2.fr:8880/erss/index.jsp?perso=bourigault&subURL=Fabre-et-bourigaultTALN2006.pdf>]
- Fiala, P., Lafon, P. & Piguet, M.-F. (éds) (1997) *La locution : entre lexicologie, syntaxe et pragmatique*. Paris : Klincksieck / Collection Saint-Cloud.
- Fitzpatrick, T. & Wray, A. (2006) « Breaking up is not so hard to do. Individual differences in L2 memorisation ». *Canadian Modern Language Review* 63/1, 35-58.
- Forsberg, F. (2005) « “Prêt-à-parler”. Les séquences préfabriquées en français parlé L2 et L1 ». In C. Bolly, J. R. Klein & B. Lamiroy (éds), *La Phraséologie dans tous ses états. Actes du colloque « Phraséologie 2005 » (Louvain-la-Neuve, 13-15 octobre 2005)*. Louvain-la-Neuve: CILL 31/2-4, 183-195.
- _____ (2006) *Le langage préfabriqué en français parlé L2. Etude acquisitionnelle et comparative*. Thèse de doctorat non publiée. Stockholm (Suède) : Forskningsrapporter/ Cahiers de la Recherche 34.
- Foster, P., Tonkyn, A. & Wigglesworth, G. (2000) « Measuring spoken language: a unit for all reasons ». *Applied Linguistics* 23/3, 354-375.
- Foster, P. & Wigglesworth, G. (2007) « Capturing accuracy in second language performance: Exploring the reliability of global and local measures ». In J. Eyckmans, A. Housen, S. Janssens, M. Pierrard, H. Stengers, S. Van Daele, Kuiken, F. et I. Vedder (éds), *Complexity, Accuracy and Fluency in Second Language Use, Learning and Teaching (Book of Abstracts of the Symposium on Complexity, Accuracy and Fluency in Second Language Use, Learning and Teaching, Brussels, Belgium, March 29-30 2007)*, 6-7.
- Francard, M., Geron, G. & Wilmet, R. (2002) « La banque de données VALIBEL: des ressources textuelles orales pour l'étude du français en Wallonie et à Bruxelles », *PUSCH*, 2002, 71-80.
- Francis, G. (1993) « A Corpus-Driven Approach to Grammar: Principles, Methods and Examples ». In M. Baker, G. Francis & E. Tognini-Bonelli (éds), *Text and Technology: In Honour of John Sinclair*, Philadelphia/ Amsterdam: Benjamins, 137-156.
- François, J. (2003) « La représentation des structures verbo-nominales et du figement verbal dans deux formalismes de grammaire fonctionnelle », *Cahiers de lexicologie* 82, 61-88.
- _____ & Manguin, J.-L. (2006) « *Dispute théologique, discussion oiseuse et conversation téléphonique : les collocations adjectivo-nominales au cœur du débat* », 50-65.
- Galisson, R. (1995) « Les palimpsestes verbaux : des révélateurs culturels remarquables mais peu remarqués ». In M. Martins-Baltar (textes réunis par), *La locution en discours, Cahiers du français contemporain 2*, Paris: Didier érudition, 41-64.
- Gardes-Tamine, J. (2007) « Normes ou canons dans la description grammaticale ? ». In G. Siouffi & A. Steuckardt (éds), *Les linguistes et la norme. Aspects normatifs du discours linguistique*, Berne: Peter Lang, 177-189.
- Gass, S. M., Bardovi-Harlig, K., Magnan, S. S. & Walz, J. (éds) (2002) *Pedagogical Norms for Second and Foreign Language Learning and Teaching*. Amsterdam/ Philadelphia: Benjamins.
- Germain, C & Netten, J. (2004) « La précision et l'aisance en FLE/FL2 : définitions, types et implications pédagogiques », *Marges Linguistiques*. Disponible en ligne à la page [http://www.caslt.org/pdf/06%20Germain_Netten%20Marges%20ling%20Juil%202004.pdf].

- Gibbs, R. W. (1995) « Idiomaticity and Human Cognition ». In M. Everaert *et al.* (éds), *Idioms: structural and psychological perspectives*, Hillsdale (New Jersey): Lawrence Erlbaum Associates, 97-116.
- Gledhill, C. & Frath, P. (2007). « Collocation, phrasème, dénomination : vers une théorie de la créativité phraséologique ». *La Linguistique* 43, 63-88.
- González Rey, I. (2002) *La phraséologie du français*. Toulouse : Presses universitaires du Mirail.
- _____ (éd.) (2007) *Les expressions figées en didactique des langues étrangères*. Cortil-Wodon: E.M.E. & InterCommunications.
- Granfeldt, J., Nugues, P., Persson, E., Persson, L., Kostadinov, F., Ågren, M. & Schlyter, S. (2005) « Direkt Profil : un système d'évaluation de textes d'élèves de français langue étrangère fondé sur les itinéraires d'acquisition ». In *Actes de la conférence Traitement Automatique des Langues Naturelles, TALN & RECITAL 2005, 6-10 juin 2005, Dourdan (France), Tome 1 – Conférences principales*, 113-122. Disponible en ligne à la page [http://www.cs.lth.se/home/Pierre_Nugues/Articles/taln2005/taln2005.pdf]
- Granfeldt, J. & Nugues, P. (2007) « Évaluations des stades de développement en français langue étrangère ». *TALN 2007, Toulouse, 5-8 juin 2007*.
- Granger, C. (2004) *Silence in Second Language Learning*. Clevedon (England): Multilingual Matters/ Second Language Acquisition 6.
- Granger, S. (1996) « From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora ». In K. Ajimer, B. Altenberg & M. Johansson (éds) *Languages in Contrast. Paper from a Symposium on Text-based Cross-linguistic Studies*, Lund: Lund University Press, 37-51.
- _____ (1998) « The computer learner corpus: a versatile new source of data for SLA research ». In S. Granger (éd.), *Learner English on Computer*, London/ New York: Longman, 3-18.
- _____ (2004) « Computer Learner Corpus research: Current status and future prospects ». In U. Connor & T. Upton (éds), *Applied Corpus Linguistics: A Multidimensional Perspective*, Amsterdam/ Atlanta: Rodopi, 123-145.
- _____, Bestgen, Y., Degand, L., Klein, J. R. & Meunier, F. (2002) *Foreign Language Learning: Phraseology and Discourse*. Louvain-la-Neuve: Université catholique de Louvain/ Centre for English Corpus Linguistics. [<http://cecl.fltr.ucl.ac.be/PhraseologyARC/welcome.html>].
- Gréciano, G. (1983) *Signification et dénotation en Allemand – La sémantique des expressions idiomatiques*. Paris-Metz : Klincksieck / Recherches Linguistiques IX.
- _____ (1997a) « La variance du figement ». In G. Kleiber & M. Riegel (éds), *Les formes du sens. Etudes de linguistique française, médiévale et générale offertes à Robert Martin à l'occasion de ses 60 ans*, Louvain-la-Neuve: Duculot/ Champs linguistiques, 149-156.
- _____ (1997b) « La phraséogenèse du discours ». In M. Martins-Baltar (textes réunis par), *La locution : entre langue et usages*, Fontenay Saint-Cloud : ENS Editions, 179-200.
- Gries, S.Th. & Stefanowitsch, A. (2004) « Extending collocation analysis. A corpus-based perspective on 'alternations' ». *International Journal of Corpus Linguistics* 9/1, 97-129.
- Gross, G. (1996) *Les expressions figées en français – Noms composés et autres locutions*. Paris : Ophrys.
- Gross, M. (1975). *Méthodes en syntaxe*. Paris : Hermann.
- _____ (1981). « Les bases empiriques de la notion de prédicat sémantique ». In A. Guillet & C. Leclère (éds), *Formes syntaxiques et prédicats sémantiques*, *Langages* 63, 7-52.
- _____ (1988). « Les limites de la phrase figée ». *Langages* 90, 7-22.
- Grossman & Tutin (2003) « Quelques pistes pour le traitement des collocations ». In Grossman & Tutin (éds), *Les collocations. Analyse et traitement* (Travaux et recherches en linguistique appliquée. Série E – n°1), Amsterdam : De Werelt, 5-21.
- Grunig, B.-N. (1997) « La locution comme défi aux théories linguistiques : une solution d'ordre mémoriel ? ». In M. Martins-Baltar (textes réunis par), *La locution entre langue et usages*. Paris: ENS éditions Fontenay/ Saint-Cloud, Collection Signes, Diffusion Ophrys, 13-18.
- Guilbert, L., Lagane, R. & Niobey, G. (dir.) (1971) *Grand Larousse de la langue française*. Paris : Librairie Larousse.
- Habert, B., Nazarenko, A. & Salem, A. (1997) *Les linguistiques de corpus*. Paris: Armand Colin.

- Halliday, M. (1966). « Lexis as a linguistic level ». In C. Bazell, J. Catford, M. A. K. Halliday & R. Robins, *In Memory of J. R. Firth*, London: Longmans.
- _____ (2004, 1^{er} éd. 1985). *An Introduction to Functional Grammar*. Londres : Arnold.
- Han, Z. (2004) *Fossilization in Adult Second Language Acquisition*. Clevedon (England): Multilingual Matters/ Second Language Acquisition 5.
- Hancock, V. & Kirchmeyer, N. (2005) « Discourse structuring in advanced L2 French: The relative clause ». In J.-M. Dewaele (éd.), *Focus on French as a Foreign Language. Multidisciplinary Approaches*, Clevedon /Buffalo/ Toronto: Multilingual Matters, 17-35.
- Harris, Z.S. (1976) *Notes du cours de syntaxe*. Paris: Le Seuil.
- Hasselgren, A. (1994) « Lexical teddy bears and advanced learners: a study into the ways Norwegian students cope with English vocabulary ». *International Journal of Applied Linguistics* 4/2, 237-260.
- Hausmann, F. J. (1997) « Tout est idiomatique dans les langues ». In M. Martins-Baltar (textes réunis par), *La locution : entre langue et usages*, Fontenay Saint-Cloud : ENS Editions, 277-290.
- Hồ-Đắc, L.-M. (2007) *La position initiale dans l'organisation du discours: une exploration en corpus*. Thèse de doctorat nouveau régime (CNRS), sous la direction de M.-P. Péry-Woodley. Toulouse (France) : Université Toulouse II-le Mirail.
- Hoey, M. (2000) « A world beyond collocation: new perspectives on vocabulary teaching ». In M. Lewis (éd.), *Teaching Collocation. Further Developments in the Lexical Approach*, Hove (England): Language Teaching Publications, 224-243.
- Housen, A. (2002) « Paramètres liés à l'environnement dans l'apprentissage guidé d'une langue seconde », *AILE (Acquisition et Interaction en Langue étrangère)* 16. Disponible en ligne à la page [<http://aile.revues.org/document1379.html>]
- Howarth, P. (1998a) « Phraseology and second language proficiency ». *Applied Linguistics* 19/1, 24-44.
- _____ (1998b) « The phraseology of learners' academic writing ». In A. P. Cowie, *Phraseology. Theory, Analysis, and Applications*, Oxford: Oxford University Press, 161-186.
- Hunston, S. & Francis, G. (2000) *Pattern Grammar. A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam/ Philadelphia: Benjamins.
- Imbs, P. (dir.) (1971) *Trésor de la langue française*. Paris: CNRS.
- Jaglinska, A. (2006) *Idiomatycity in learner language: a study of the use of prefabs in the writing of Polish advanced EFL learners*. Thèse de doctorat non publiée. Université de Marie Curie-Skodowska Lublin.
- James (1998) *Errors in Language Learning and Use: Exploring Error Analysis*. London/ New York: Longman.
- Källkvist, M. (1998) « Lexical infelicity in English: the case of nouns and verbs ». In K. Haastrup & Å. Viberg (éds), *Perspectives on lexical acquisition in a second language, Travaux de l'Institut de Linguistique de Lund* 38, Lund : Lund University Press, 149-174.
- Kennedy, G. (1998) *An Introduction to Corpus Linguistics*. London, New York: Longman/ Studies in Language and Linguistics.
- Klein, J. R. (2007) « Le figement, un concept aussi essentiel que fluent. Réflexions à travers la synchronie et la diatopie », In A. Häcki Buhofer & H. Burger, *Phraseology in Motion II. Theorie und Anwendung* (Actes du colloque Europhras, 26-29 août 2004, Bâle, Suisse), Baltmannsweiler: Schneider Verlag Hohengehren/ Phraseologie und Parömiologie 20, 75-84.
- _____ & Rossari, C. (2003) « Figement et variations en français de Belgique, de France, du Québec et de Suisse ». *Linguisticae Investigationes* 26/2, 203-214.
- Kokochkina, I. (2004) « Quelques aspects sémantiques dans le TAL. Le cas des prédicats d'états » (Actes des VIIèmes RJC ED268 'Langage et Langues', Paris III, 15 mai 2004). Disponible en ligne à la page [http://www.cavi.univ-paris3.fr/ilpga/ED/activites/RJC2004_actes/kokochkina_rjc2004.pdf]
- Kübler, N. (1995) *L'automatisation de la correction d'erreurs syntaxiques – application aux verbes de transfert en anglais pour francophones* (Thèse de doctorat en Linguistique

- Théorique et Formelle – Dir. : M. Gross*). Disponible en ligne à la page [<http://wall.jussieu.fr/~nkubler/>]
- Lamiroy, B. (2003) « Les notions linguistiques de figement et de contrainte » (Actes du colloque rendant hommage à Maurice Gross, Paris, 3 juin 2002). *Linguisticae Investigationes* 26/1, 1-14.
- _____ & Klein, J. R. (2005) « Le problème central du figement est le semi-figement ». *Le Semi-figement, LINX* 53, 135-154.
- Langlotz, A. (2001) « Cognitive principles of idiom variation – Idioms as complex linguistic categories ». *Studi Italiani di Linguistica Teorica e Applicata* 30/2, 289-302.
- Laufer, B. & Nation, P. (1995) « Vocabulary size and use: Lexical richness in L2 written production ». *Applied Linguistics* 16/3, 307-322.
- Lerot, J. (1993) *Précis de linguistique générale*. Paris: Minuit.
- Lewis, Margareta (2005) *The Elusive Formulaic Structure. Assessing the adult intermediate learner's idiomaticity, fluency and proficiency*. Mémoire de phil. lic. (non publié). Département d'anglais. Université de Stockholm.
- Lewis, Michael (éd.) (2000) *Teaching Collocation. Further Developments in the Lexical Approach*. Hove (England): Language Teaching Publications.
- _____ (2000) « Language in the lexical approach ». In M. Lewis (éd.), *Teaching Collocation. Further Developments in the Lexical Approach*, Hove (England): Language Teaching Publications, 126-154.
- Magnan, S. S. & Walz, J. (2002) « Pedagogical norms: Development of the concept and illustration from French ». In S. M. Gass, K. Bardovi-Harlig, S. S. Magnan & J. Walz (éds), *Pedagogical Norms for Second and Foreign Language Learning and Teaching*, Amsterdam/ Philadelphia: Benjamins, 15-40.
- Martin, R. (1997) « Sur les facteurs du figement lexical ». In M. Martins-Baltar (textes réunis par), *La locution : entre langue et usages*, Fontenay Saint-Cloud : ENS Editions, 291-305.
- Martins-Baltar, M. (textes réunis par) (1995) *La locution en discours, Cahiers du français contemporain 2*. Paris : Didier érudition.
- _____ (textes réunis par) (1997) *La locution entre langue et usages*. Paris: ENS éditions Fontenay/ Saint-Cloud, Collection Signes, Diffusion Ophrys.
- Mas, M. & Turco, G. (1991) « Des typologies de textes à l'élaboration d'outils pour la formation des maîtres et d'outils pour la classe ». In J.-P. Bronckart, D. Coste & E. Roulet (coord.), *Textes, Discours et Genres, Etudes de linguistique appliquée* 83, 89-99.
- McEnery, T. & Wilson, A. (1996/ 2001, 2^e éd.) *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Mejri, S. (1997a) *Le figement lexical – Descriptions linguistiques et structuration sémantique*. Tunis : Publications de la faculté des lettres de la Manouba.
- _____ (1997b) « Binarisme, dualité et séquences figées ». In R. Martin (éd.), *Les formes du sens: études de linguistique française, médiévale et générale offertes à Robert Martin à l'occasion de ses 60 ans*, Louvain-la-Neuve: Duculot, 249-256.
- _____ (1998) « Le figement lexical – description linguistique et structuration sémantique (Présentation de thèses) ». *L'information grammaticale* 76, 50-51.
- _____ (coord. par) (2003a) *Le figement lexical, Cahiers de lexicologie* 82-1.
- _____ (2003b) « Le figement lexical ». In S. Mejri (coord.), *Le figement lexical, Cahiers de Lexicologie* 83/1, 23-39.
- _____ (2005) « Figement absolu ou relatif : la notion de degré de figement ». *Le Semi-figement, LINX* 53, 183-196.
- Mel'čuk, I. (1993) « La phraséologie et son rôle dans l'enseignement/ apprentissage d'une langue étrangère ». *Revue de Linguistique appliquée* 92, 82-113.
- _____ (1995) « Phrasemes in Language and Phraseology in Linguistics ». In M. Everaert *et al.* (éds), *Idioms: structural and psychological perspectives*, Hillsdale (New Jersey): Lawrence Erlbaum Associates, 167-232.
- _____ (1998) « Collocations and Lexical Functions ». In A. P. Cowie, *Phraseology. Theory, Analysis, and Applications*, Oxford: Oxford University Press, 23-53.

- _____, Clas, A. & Polguère, A. (1995) *Introduction à la lexicologie explicative et combinatoire*. Louvain : AUPELF-UREF-Duculot/ Champs linguistiques.
- Meunier, F. (1998) « Computer tools for the analysis of learner corpora ». In S. Granger (éd.), *Learner English on Computer*, London/ New York: Longman, 19-37.
- Mini-guide SPSS*, Université de Montréal. Disponible en ligne à la page [http://www.dms.umontreal.ca/~stat/Logiciels/spss/mini-guide/spss.html]
- Moon, R. (1998) *Fixed expressions and idioms in English*. Oxford: Clarendon Press.
- Mortureux, M.-Fr. (2001) *La lexicologie entre langue et discours*. Paris : Armand Colin/ Campus-Linguistique.
- _____. (2003) « Figement lexical et lexicalisation ». *Cahiers de lexicologie* 82/1, 11-22.
- Nesselhauf, N. (2004) « How learner corpus analysis can contribute to language teaching: A study of support verb constructions ». In G. Aston, S. Bernardini & D. Stewart (éds), *Corpora and Language Learners*, Philadelphia/ Amsterdam: Benjamins/ Studies in Corpus Linguistics, 109-124.
- _____. (2005) *Collocations in a Learner Corpus*. Amsterdam/ Philadelphia: Benjamins.
- Pander Maat, H. & Degand, L. (2001) « Scaling causal relations and connectives in terms of Speaker Involvement ». *Cognitive Linguistics* 12, 211-245.
- Partington, A. (1998) *Patterns and meanings – Using corpora for English Language Research and Teaching*. Amsterdam/ Philadelphia: Benjamins/ Studies in Corpus Linguistics.
- Perdue, C. (1977) « Etude d'un corpus de productions écrites (apprentissage de l'anglais par des adultes francophones) ». *Etudes de Linguistique Appliquée* 25, 44-53.
- _____. (1980) « L'analyse des erreurs : un bilan pratique ». *Langages* 57, 87-94.
- Pienemann, M., Johnston, M. & Brindley, G. (1988) « Constructing an acquisition-based procedure for second language assessment ». *Studies in Second Language Acquisition* 10, 217-243.
- Pienemann, M. (1998) *Language Processing and Second Language Development: Processability Theory*. Amsterdam: Benjamins.
- Pit, M. (2003) *How to express yourself with a causal connective – Subjectivity and causal connectives in Dutch, German and French*. Amsterdam/ New York: Rodopi/ USL&C.
- Porquier, R. (1977) « L'analyse des erreurs. Problèmes et perspectives ». *Etudes de Linguistique Appliquée* 25, 23-43.
- _____. & Frauenfelder (1980) « Enseignants et apprenants face à l'erreur - Ou de l'autre côté du miroir ». *Le français dans le monde* 154, 29-36.
- Preacher, K. J. (2001, April). « Calculation for the chi-square test: An interactive calculation tool for chi-square tests of goodness of fit and independence [Computer software] ». Disponible en ligne à la page [http://www.quantpsy.org]
- Pusch, C. D. (2007) « 'Faut dire': variation et sens d'un marqueur parenthétique entre connectivité et (inter)subjectivité. In G. Dostie (éd.), *Les marqueurs discursifs, Langue française* 154, 29-44.
- Py, B. (2000) « La construction interactive de la norme comme pratique et comme représentation ». *AILE (Acquisition et Interaction en Langue étrangère)* 12. Disponible en ligne à la page [http://aile.revues.org/document1464.html]
- Rastier, F. (2002) « Enjeux épistémologiques de la linguistique de corpus » (Texte issu d'une conférence aux deuxièmes Journées de Linguistique de Corpus, Lorient, septembre 2002). Disponible en ligne à la page [http://www.revue-texto.net/Inedits/Rastier_Enjeux/Rastier_Enjeux.html]
- _____, Cavazza, M. & Abeillé, A. (1994) *Sémantique pour l'analyse – De la linguistique à l'informatique*. Paris: Masson.
- _____. (2007) « Conditions d'une linguistique des normes ». In G. Siouffi & A. Steuckardt (éds), *Les linguistes et la norme. Aspects normatifs du discours linguistique*, Berne: Peter Lang, 3-20.
- Raupach, M. (1984) « Formulae in second language speech production ». In H. W. Dechert, D. Möhle & M. Raupach (éds), *Second Language Production*. Tübingen (Allemagne): Gunter Narr, 114-137.

- Recanati, C. & Recanati, F. G. (1999) « La classification de Vendler revue et corrigée ». *Cahiers Chronos* 4, 167-184. Disponible en ligne à la page [<http://hal.archives-ouvertes.fr/hal-00085094>].
- Reichler-Béguelin (1993) M.-J. (1993) « Faits déviants et tri des observables ». In *Le traitement des données linguistiques non standard (Actes des Rencontres Besançon-Neuchâtel, Neuchâtel, 29-30 janvier 1993)*, BULAG-TRANEL 20, 89-109.
- Rey, A. (éd.) (1989) *Le Grand Robert de la langue française – Dictionnaire alphabétique et analogique de la langue française*. Paris : Dictionnaires Le Robert.
- Rey-Debove, J. & Rey, A. (éds) (2003) *Le Nouveau Petit Robert – Dictionnaire alphabétique et analogique de la langue française*. Paris: Dictionnaires Le Robert.
- Roulet, E. , Fillietaz, L. & Grobet, A. (2001) *Un modèle et un instrument d'analyse de l'organisation du discours*. Berne: Peter Lang.
- Schmitt, N. (éd.) (2004) *Formulaic Sequences. Acquisition, Processing and Use*. Amsterdam/ Philadelphia: Benjamins.
- _____ & Carter, R. (2004) « Formulaic sequences in action: An introduction ». In N. Schmitt (éd.), *Formulaic Sequences. Acquisition, Processing and Use*, Amsterdam/ Philadelphia: Benjamins, 1-22.
- _____, Dörnyei, Z., Adolphs, S. & Durow, V. (2004) « Knowledge and acquisition of formulaic sequences: A longitudinal study ». In N. Schmitt (éd.), *Formulaic Sequences. Acquisition, Processing and Use*, Amsterdam/ Philadelphia: Benjamins, 55-86.
- Scott, M. *Oxford Wordsmith Tools 4.0*.
- Sinclair, J. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- _____ (2000) *Lexical grammar. Naujoji Metodologija*, 191-204. Disponible en ligne à la page [<http://donelaitis.vdu.lt/publikacijos/sinclair.pdf>]
- _____ & Renouf A. (1991) « Collocational frameworks in English ». In K. Aijmer & B. Altenberg (éds), *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, Amsterdam / Philadelphia: Benjamins, 128-144.
- Siouffi, G. & Steuckardt, A. (éds) (2007) *Les linguistes et la norme. Aspects normatifs du discours linguistique*, Berne: Peter Lang.
- _____ (2007) « Présentation ». In G. Siouffi & A. Steuckardt (éds), *Les linguistes et la norme. Aspects normatifs du discours linguistique*, Berne: Peter Lang, VII-XXIII.
- Site officiel SPSS* [<http://www.spss.com/fr>]
- Stefanowitsch, A. & Gries, S.Th. (2003) « Collocations: Investigating the interaction of words and constructions ». *International Journal of Corpus Linguistics* 8/2, 209-243.
- Stubbs, M. (1995) « Collocations and semantic profiles: On the cause of the trouble with quantitative studies ». *Functions of Language* 2/1, 23-55.
- Svensson, M. H. (2004) *Critères de figement - L'identification des expressions figées en français contemporain*. Suède, Umeå university: Institutionen för Moderna Språk Publikationer.
- Swales, J. M. (1990) *Genre Analysis. English in academic and research settings*. Cambridge: Cambridge University Press.
- Tognini-Bonelli, E. (2001) *Corpus Linguistics a Work*. Amsterdam/ Philadelphia: Benjamins/ Studies in Corpus Linguistics.
- Towell, R., Hawkins, R. & Bazergui, N. (1996) « The development of fluency in advanced learners of French ». *Applied Linguistics* 17/1, 84-119.
- Tutin, A. & Grossman, F. (2002) « Collocations régulières et irrégulières. Esquisse de typologie du phénomène collocatif ». *Revue française de Linguistique appliquée* 7-1, 7-25.
- Tutin, A. (2005). « Collocations du lexique transdisciplinaire des écrits scientifiques : annotation et extraction des propriétés linguistiques dans la perspective d'une application didactique ». In C. Bolly, J. R. Klein & B. Lamiroy (éds), *La Phraséologie dans tous ses états. Actes du colloque « Phraséologie 2005 » (Louvain-la-Neuve, 13-15 octobre 2005)*. Louvain-la-Neuve: CILL 31/2-4, 247-262.
- Van Daele, S., Housen, A. et Pierrard, M. (2007) « Les mécanismes psycholinguistiques à la base de l'expression et du développement de la fluidité, de la précision et de la complexité en L2 ». Présentation au *Symposium on Complexity, Accuracy an Fluency in Second*

- Language Use, Learning and Teaching, Brussels, Belgium, March 29-30 2007*, diapositives 1-36.
- Vendler, Z. (1957) « Verbs and Times ». In Z. Vendler (1967), *Linguistics in Philosophy*, Ithaca: Cornell University Press, 97-121.
- Viberg, Å. (1998) « Crosslinguistic perspectives on lexical acquisition : the case of language-specific semantic differentiation ». In K. Haastrup & Å. Viberg (éds), *Perspectives on lexical acquisition in a second language, Travaux de l'Institut de Linguistique de Lund 38*, Lund : Lund University Press, 175-207.
- _____ (2002a) « Polysemy and disambiguation cues across languages. The case of Swedish *få* and English *get* ». In B. Altenberg & S. Granger (éds), *Lexis in contrast. Corpus based approaches, Studies in Corpus Linguistics 7*, Amsterdam/ Philadelphia: Benjamins, 119-150.
- _____ (2002b) « Basic Verbs in Second Language Acquisition ». *Revue française de Linguistique Appliquée* VII/2, 51-69.
- Watorek, M. (1998) (éd.) *Structure des lectures des apprenants. AILE (Acquisition et Interaction en Langue étrangère)* 11.
- Wiktorsson, M. (2003) *Learning idiomaticity*. Thèse de doctorat. Département d'anglais. Université de Lund (Suède): *Lund Studies in English* 105.
- Willis, D. (2003) *Rules, Patterns and Words. Grammar and Lexis in English Language Teaching*. Cambridge: Cambridge University Press.
- Wray, A. (2002) *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- _____ (2005) « 'Needs only analysis' in linguistic ontogeny and phylogeny ». In A. Cangelosi (ed.), *Proceedings of the Second International Symposium on the Emergence and Evolution of Linguistic Communication*, Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB), Hatfield, April 2005.

Description des annexes [sur cd-rom, ci-joint]

Annexe 1. Etude pilote sur le degré de figement en français L1 – Résultats

Ce document contient le tableau des résultats de l'analyse paramétrique effectuée pour déterminer le degré de figement (DF) de 200 séquences verbo-nominales (SVN) construites avec le verbe *prendre* en français L1 (400.000 mots de textes journalistiques et de textes littéraires) (cf. I.1.4.). Les 200 SVN étudiées sont classées en quatre catégories en fonction de leur DF global : séquences figées (SF), semi-figées (SSF), quasi-figées (SQF) ou libres (SL). Dans l'annexe, le tableau contient les informations suivantes :

SVN	Fragment textuel	Fixité %	Opacité %	Restriction lexicale %	DF %	Catégorie
-----	------------------	-------------	--------------	---------------------------	---------	-----------

Dans le corps de thèse, nous faisons référence aux résultats de cette étude en indiquant la source des exemples (x) entre parenthèse:

(x) _____ (FL1-pilote, <code du texte source>, occ. y)

Ces résultats sont présentés dans l'annexe par ordre d'apparition des SVN dans la concordance, autrement dit, par ordre chronologique croissant de la numérotation (y) des fragments textuels. La numérotation des occurrences (y) dans le corps de thèse correspond donc à celle des fragments textuels dans l'annexe.

Annexe 2. Les stades d'acquisition en français oral L2

Nous présentons dans ce document un résumé de la catégorisation en stades opérée par Bartning & Schlyter (2004), représentant le parcours du processus acquisitionnel du français L2 à l'oral par des apprenants suédophones.

Annexe 3. Analyse sur corpus d'apprenants en L2 – Liste paramétrique

Dans le tableau de cette annexe, nous listons les variables principales qui ont été opérationnalisées pour effectuer l'analyse contrastive interlangagière des unités phraséologiques produites par les apprenants du français académique L2. Cette liste regroupe les variables linguistiques formelles et sémantiques (cf. V.1. et V.2.), les variables extralinguistiques (cf. V.3.), les variables opérationnalisées pour l'étude collocationnelle (cf. IV.1. et IV.2.) et les variables nécessaires à l'analyse de déviance (cf. IV.1. et IV.3.). Nous avons organisé ces variables en indiquant le nom (*Name*), le titre (*Label*), les valeurs (*Values*) et le type de mesure (*Measure*) de chacune des variables encodées dans les tables du logiciel SPSS. Ces informations sont indispensables pour le traitement quantitatif et statistique des données de corpus étudiées.

Nom de la variable (<i>Name</i>)	Titre de la variable (<i>Label</i>)	Valeurs (<i>Values</i>)	Mesure (<i>Measure</i>)
---------------------------------------	--	------------------------------	------------------------------

Annexe 4. Analyse sur corpus d'apprenants en L2 – Description des corpus

Dans le tableau de cette annexe, nous dressons une liste donnant la description détaillée des textes qui constituent nos corpus d'étude (français L2) et de contrôle (français L1) en français académique. Les informations qui y sont listées concernent la nature de la source textuelle (le code du texte, le nom du corpus source, le code du corpus source, la taille du texte en nombre de mots nombre de mots et le type de texte) et le contexte d'apprentissage (le pays et le nom de l'institution scolaire de l'apprenant).

Code du texte	Corpus source	Code du texte source	Taille du texte	Type de texte et/ou thème	Pays	Institution
---------------	---------------	----------------------	-----------------	---------------------------	------	-------------

Nous décrivons d'abord le corpus d'apprenants anglophones du français L2 (237 textes – 185.000 mots) (pp. 1-18), puis le corpus néerlandophone (221 textes – 90.000 mots) (pp. 19-34) et, enfin, le corpus de contrôle en français L1 (330 textes – 185.000 mots) (pp. 35-55).

Annexe 5. Analyse sur corpus d'apprenants en L2 – Résultats

Ce document contient un tableau récapitulatif des résultats de l'analyse collocationnelle et de déviance effectuée pour l'analyse contrastive interlangagière des 906 séquences verbo-nominales (SVN), dont 703 unités phraséologiques (UP), produites par les apprenants du français académique L2 et L1. Ces résultats sont présentés dans l'annexe en listant d'abord, par ordre alphabétique du constituant nominal (2^e colonne, intitulée *Nom*), les SVN construites avec *prendre*, puis celles avec *donner*, indépendamment du corpus dans lequel ces SVN apparaissent.

La trame de fond utilisée sert à opérer une distinction entre les SVN libres, les UP non déviantes et les UP déviantes :

Unité phraséologique déviante
Unité phraséologique
Séquence libre

Dans ce tableau, nous indiquons quels sont le verbe (*prendre* ou *donner*) et le nom lemmatisé constitutifs de la séquence étudiée, suivis du numéro attribué à l'occurrence de cette séquence.

Le fragment textuel dans lequel apparaît la SVN en contexte est également reproduit, en gardant les règles typographiques identiques à celles qui ont été utilisées dans le corps de la thèse : le caractère gras pour le constituant verbal, le soulignement simple pour les syntagmes nominaux a-prépositionnels et le double soulignement pour les syntagmes nominaux prépositionnels ; les annotations des déviances formelles (*), contextuelles (?) et quantitatives (!); la portée de la déviance délimitée par des barres verticales (|) dans l'unité phraséologique.

Après le fragment textuel, nous présentons, pour chaque occurrence d'UP déviante, les reconstructions natives qui ont été retenues comme étant les plus pertinentes (les parties corrigées par rapport à l'occurrence originale sont en bleu clair), le type de déviance (formelle, contextuelle, quantitative ou mixte), la localisation de la déviance (en termes de catégorie grammaticale) et le degré global de déviance (*Gravité 1* = faible gravité ; *Gravité 2* = gravité moyenne ; *Gravité 3* = gravité importante).

Les deux dernières colonnes du tableau nous donnent des informations sur le corpus (FL2-AN, FL2-NL ou FL1) et sur le texte d'où proviennent les occurrences répertoriées.

Verbe	Nom	Occ.	Fragment textuel	Reconstruction(s) native(s) des déviances	Type de déviance	Lieu de la déviance	Gravité déviance	Corpus	Texte
-------	-----	------	------------------	---	------------------	---------------------	------------------	--------	-------

Notons que la numérotation des occurrences (y) dans les exemples (x) du corps de thèse, correspond à celle des occurrences de l'annexe.

(x) _____ (FL2-AN ou FL2-NL ou FL1, <code du texte source>, occ. y)