

GRAMM - R

ÉTUDES DE LINGUISTIQUE FRANÇAISE

Catherine Bolly

# Phraséologie et collocations

Approche sur corpus  
en français L1 et L2



Répondant au manque de consensus qui règne encore actuellement dans le domaine phraséologique, l'auteure propose de redéfinir ici les notions de phraséologie (ou figement) et de collocation (ou semi-figement), en réunissant les littératures de tradition anglo-saxonne et francophone. Le volet acquisitionnel de l'ouvrage se centre quant à lui sur les productions d'apprenants avancés du français langue étrangère (L2), en particulier sur les déviations au sein des collocations verbonominales, encore relativement peu étudiées en français L2. L'ouvrage présente en outre une méthode originale d'analyse paramétrique sur corpus (en L1 et en L2) qui permet, d'une part, de traduire en termes quantitatifs des données initialement qualitatives et, d'autre part, de combiner l'approche fonctionnelle (via des critères linguistiques) et statistique (via des mesures d'association lexicale) en phraséologie. Cette méthode est décrite avec soin par l'auteure qui guide le lecteur en l'impliquant dans sa démarche exploratoire au fil des pages.

Docteur en langues et lettres (linguistique), **Catherine Bolly** est chercheuse à l'Université catholique de Louvain (chargée de recherches du F.R.S.-FNRS) et membre du centre de recherches Valibel – Discours et variation. Ses domaines d'étude privilégiés sont la phraséologie, la grammaticalisation et les marqueurs du discours. Elle s'intéresse en particulier aux phénomènes linguistiques qui sont polysémiques et/ou fréquents dans l'usage, en français oral et écrit.





# **Phraséologie et collocations**

**Approche sur corpus  
en français L1 et L2**



**P.I.E. Peter Lang**

Bruxelles · Bern · Berlin · Frankfurt am Main · New York · Oxford · Wien



**Catherine BOLLY**

# **Phraséologie et collocations**

**Approche sur corpus  
en français L1 et L2**

« GRAMM-R. Études de linguistique française »

N° 9

Institut Langage et Communication (IL&C)  
Place Blaise Pascal, 1, B-1348 Louvain-la-Neuve (Belgique)

Publié avec le concours de la Fondation Universitaire de Belgique



Toute représentation ou reproduction intégrale ou partielle faite par quelque procédé que ce soit, sans le consentement de l'éditeur ou de ses ayants droit, est illicite. Tous droits réservés.

© P.I.E. PETER LANG S.A.  
Éditions scientifiques internationales  
Bruxelles, 2011  
1 avenue Maurice, B-1050 Bruxelles, Belgique  
www.peterlang.com ; info@peterlang.com

Imprimé en Allemagne

ISSN 2030-2363 (L'édition de l'imprimé)  
ISBN 978-90-5201-748-8 E-ISBN 978-3-0352-6101-1  
D/2010/5678/54

Information bibliographique publiée par « Die Deutsche Nationalbibliothek »  
« Die Deutsche Nationalbibliothek » répertorie cette publication dans la « Deutsche  
Nationalbibliografie » ; les données bibliographiques détaillées sont disponibles sur le site  
<<http://dnb.de>>.

*à Jeannou et Lili*

« La linguistique reste une science non normative parfaitement légitime, dès lors qu'elle ne cherche pas à promouvoir les normes puristes, mais à décrire celles qui prévalent dans une société. »

(Branca-Rosoff, 2007 : 22)

« Le hic, c'est quand la linguistique devient pure mécanique. »



## Table des matières

<b>Remerciements</b> .....	11
<b>Liste des tableaux et figures</b> .....	13
<b>Liste des principales abréviations</b> .....	17
<b>INTRODUCTION. Complexité des unités phraséologiques</b> .....	19
<b>CHAPITRE 1. Définir les unités phraséologiques</b> .....	23
1. Phraséologie au sens strict (approche fonctionnelle) .....	23
2. Phraséologie au sens large (approche statistique) .....	52
<b>CHAPITRE 2. Mesurer le degré de figement</b> .....	65
1. Méthode .....	65
2. Postulats et hypothèse .....	70
3. Opérationnalisation et analyse paramétrique (L1) .....	72
4. Traitement des données et résultats.....	77
<b>CHAPITRE 3. Acquérir les unités phraséologiques</b> .....	83
1. Compétence phraséologique en langue étrangère ou seconde .....	83
2. De l'erreur à la déviance .....	110
<b>CHAPITRE 4. Identifier les déviances phraséologiques</b> .....	133
1. Corpus et données .....	133
2. Objectif de recherche .....	139
3. Opérationnalisation et analyse paramétrique (L2) .....	140
4. Analyse collocationnelle .....	156
5. Analyse de déviances collocationnelles .....	168
6. Synthèse et perspectives.....	204
<b>Conclusion</b> .....	209
<b>Bibliographie</b> .....	213
<b>Index des auteurs</b> .....	227
<b>Index des notions</b> .....	231
<b>ANNEXE 1. Liste paramétrique. Degré de figement en L1.</b> .....	237
<b>ANNEXE 2. Résultats. Degré de figement en L1.</b> .....	243

**ANNEXE 3.** Liste paramétrique. Analyse de déviations  
collocationnelles en L2 .....263

**ANNEXE 4.** Résultats. Analyse collocationnelle et de  
déviations en L2 .....267

## 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 ont permis à cet ouvrage de voir le jour.

Mes remerciements vont tout d'abord au Prof. L. Degand (F.R.S.-FNRS/Université catholique de Louvain), pour la confiance qu'elle me témoigne depuis plusieurs années et qui me pousse à croire que la linguistique se conjugue parfaitement à la première personne du féminin maternel. Je remercie ensuite mon promoteur de thèse, le Prof. J. Klein (Université catholique de Louvain), et les membres du jury de ma thèse (le Prof. I. Bartning, le Prof. C. Fairon, le Prof. S. Granger et le Prof. B. Lamiroy) pour avoir si bien encadré (et recadré quand cela s'est avéré nécessaire) mes recherches doctorales. Je n'oublie pas les personnes qui ont contribué à l'élaboration de mon corpus de recherche, à savoir le Prof. F. Thyron (Université catholique de Louvain) pour son aide à la constitution du corpus *CAFLaM*, les Prof. M. Goyens, L. Melis et Y. Sturbeaux (Katholieke Universiteit Leuven) pour l'extension du corpus *FRIDA*, le *CENTAL* et le Prof. L. Degand (Université catholique de Louvain) pour m'avoir transmis les corpus de presse. Je remercie également les Prof. M. Francard et S. Granger (Université catholique de Louvain) d'avoir accepté de mettre à ma disposition leurs corpus respectifs, *Valibel* et *FRIDA*. Deux relecteurs anonymes ont par ailleurs permis à l'ouvrage de gagner en scientificité et en lisibilité, je les en remercie chaleureusement. Merci enfin à mes proches, famille et amis, pour leur soutien sans faille et leur disponibilité sans fin. Merci surtout à mon compagnon de toujours et de tous les jours, pour sa présence (souvent silencieuse) et sa patience dans mes absences (souvent capricieuses).

Cet ouvrage a été publié avec l'aide de la Fondation Universitaire de Belgique. Le travail dont il rend compte s'inscrit dans le cadre d'une thèse de doctorat menée au sein d'un projet collectif financé par la Communauté française de Belgique (ARC 03/08-301), portant sur l'appropriation des langues étrangères d'un point de vue phraséologique et discursif. L'auteure est chercheuse du Fonds de la Recherche Scientifique (F.R.S.-FNRS) à l'Université catholique de Louvain (Louvain-la-Neuve, Belgique).



## Liste des tableaux et figures

Tableau 1. Critères de figement.....	30
Tableau 2. Critère de polylexicalité. Morphologie et phraséologie.....	33
Figure 1. Typologie phraséologique et classification fonctionnelle des unités phraséologiques .....	40
Tableau 3. Définition fonctionnelle de la collocation.....	48
Tableau 4. Cooccurents syntaxiques pour le lemme <i>amoureux</i> ( <i>Les Voisins De Le Monde</i> ) .....	56
Tableau 5. Séquences récurrentes construites avec <i>prendre</i> (+ nom), par ordre décroissant de fréquence .....	58
Tableau 6. Degré de figement en français L1. Description des corpus .....	66
Tableau 7. Degré de figement de quatre SVN représentatives des seuils fixés entre catégories de figement .....	78
Tableau 8. Séquences semi-figée et quasi-figée montrant un degré de figement proche .....	79
Tableau 9. Variabilité du degré de figement de la SVN type construite avec <i>prendre + forme</i> .....	80
Tableau 10. Itinéraire de développement en français L2.....	91
Tableau 11. Description des corpus non natifs et natifs .....	135
Tableau 12. Distribution des séquences verbo-nominales dans les corpus non natifs et natifs .....	139
Tableau 13. Liste du LADL – Correspondance formelle stricte.....	145
Tableau 14. Liste du LADL – Correspondance désambiguïsée .....	146
Tableau 15. Liste du LADL – Correspondance avec variante formelle .....	147
Tableau 16. Liste du LADL – Correspondance partielle.....	147
Tableau 17. Variable : correspondance avec la liste du LADL .....	148
Tableau 18. Cooccurents syntaxiques pour le nom <i>opportunité</i> .....	149
Tableau 19. Voisins distributionnels pour le nom <i>opportunité</i> .....	149
Tableau 20. Moyenne d’IM des SVN phraséologiques avec <i>prendre</i> et <i>donner</i> .....	151

Tableau 21. Cooccurents syntaxiques et couples concurrents pour <i>donner Dét même possibilité de N à X</i> .....	152
Tableau 22. Variables : fréquences absolues et relatives en L1 .....	155
Tableau 23. Fréquence absolue et relative de <i>donner + explication</i> .....	156
Tableau 24. Cooccurents syntaxiques, couples concurrents et fréquence pour <i>prendre + voie</i> .....	164
Tableau 25. Cooccurents syntaxiques et couples concurrents pour <i>prendre + indépendance</i> .....	165
Tableau 26. Variable : statut phraséologique ou libre des SVN.....	167
Tableau 27. Distribution des SVN phraséologiques vs libres dans les corpus non natifs et natifs .....	167
Tableau 28. Variable : types de déviance formelle, contextuelle, quantitative ou multiple.....	172
Tableau 29. UP déviantes vs non déviantes dans les corpus non natifs et natifs .....	183
Tableau 30. Types de déviances phraséologiques non natives et natives.....	184
Tableau 31. Variable : localisation de la déviance en termes de catégories grammaticales.....	186
Tableau 32. Lieux de déviances phraséologiques non natives et natives.....	188
Tableau 33. Valeurs attribuées aux types de localisation de la déviance .....	189
Tableau 34. Variable recodée pour les déviances portant sur le verbe .....	192
Tableau 35. Variable : degré de gravité de l'UP déviante en fonction du nombre total de déviances par UP .....	192
Tableau 36. Variable : degré de gravité de l'UP déviante en fonction du type de déviance phraséologique.....	193
Tableau 37. Variable : degré de gravité de l'UP déviante en fonction des catégories grammaticales .....	196
Tableau 38. Variable : degré de gravité de l'UP déviante en fonction de l'IM en L1 .....	197
Tableau 39. Variable : degré de gravité de l'UP déviante en fonction de sa fréquence relative en L1 .....	200
Tableau 40. Variable : degré global de gravité de l'UP déviante .....	201

Tableau 41. Distribution des UP déviantes dans les corpus non natifs et natifs, en fonction de leur degré global de gravité.....	201
Tableau 42. Analyse d'une UP déviante à degré de gravité élevé.....	202
Tableau 43. Analyse d'une UP déviante à degré de gravité moyen ....	203
Tableau 44. Analyse d'une UP déviante à degré de gravité faible .....	203



## Liste des principales abréviations

AA	Apprenant(s) avancé(s)
AI	Apprenant(s) ayant un niveau élevé d'instruction
CSR	Combinaison(s) syntagmatique(s) récurrente(s)
CVS	Construction(s) à verbe support
DF	Degré de figement
FL1	Corpus de contrôle en français langue première
FL2-AN	Corpus d'apprenants anglophones (français langue seconde/ étrangère)
FL2-NL	Corpus d'apprenants néerlandophones (français langue seconde/ étrangère)
FqA	Fréquence absolue : nombre effectif d'occurrences
FqR	Fréquence relative : <i>n</i> occurrences par million de mots
IL	Interlangue
IM	Information mutuelle
LC	Langue cible
LE	Langue étrangère
LM	Langue maternelle
LN	Locuteur(s) natif(s)
LNN	Locuteur(s) non natif(s)
LS	Langue source
L1	Langue première
L2	Langue seconde ou étrangère
M de mots	Million de mots
NP	Norme(s) pédagogique(s)
SF	Séquence(s) figée(s)
SL	Séquence(s) libre(s)
SP	Séquence(s) préfabriquée(s)
SPIDI	Séquence(s) préfabriquée(s) idiosyncrasique(s) intérimaire(s)
SQF	Séquence(s) quasi-figée(s)
SSF	Séquence(s) semi-figée(s)
SVN	Séquence(s) verbo-nominale(s)
UP	Unité(s) phraséologique(s)
VHF	Verbe(s) à haute fréquence



## INTRODUCTION

### Complexité des unités phraséologiques

Bien qu'omniprésentes dans l'usage courant (par ex. : *avoir lieu, tirer la langue, fou rire, t'sais, à moins que*) et dans les langues de spécialité (par ex. : *rendre un arrêt* (juridique), *chambre stérile* (médical)), les unités phraséologiques ont longtemps été considérées comme des phénomènes linguistiques atypiques gravitant à la périphérie de la syntaxe. Les linguistes eux-mêmes n'ont pas encore fini de chercher à expliquer le fonctionnement de ces unités situées à l'interface du lexique et de la grammaire, qu'il s'agisse de les décrire à partir de critères internes à la langue (sémantiques, syntaxiques, lexicaux, etc.) ou de critères externes à celle-ci (cognitifs, sociaux, géographiques, diachroniques, stylistiques, etc.). Comment en effet expliquer que le locuteur natif préfère utiliser le verbe *tirer* – plutôt que le verbe *sortir, extraire* ou *tendre* –, pour exprimer verbalement l'action de « faire sortir » *la langue* de sa bouche ? Et comment expliquer cette différence de sens, que le locuteur natif perçoit de manière intuitive, entre l'enfant qui *tire la langue* pour faire une grimace à son voisin et l'enfant qui *tire la langue* pour que le médecin puisse l'ausculter ? Si aucune réponse simple ne semble à ce jour pouvoir être apportée à ces questions, on s'accorde néanmoins actuellement sur le fait que ces unités pour le moins complexes se laissent difficilement appréhender tant par le locuteur natif que par le locuteur non natif.

Partant de ce constat, nous tenterons de déterminer, dans cet ouvrage, en quoi les unités phraséologiques construites avec des verbes fréquents peuvent être comprises comme des phénomènes linguistiques « complexes », en français écrit langue première (L1) et langue étrangère ou seconde (L2). L'ouvrage est ainsi structuré en quatre parties qui visent chacune à répondre à l'une des questions suivantes : Qu'est-ce qu'une unité phraséologique (chapitre 1) ? Comment mesurer le degré de figement des unités phraséologiques (chapitre 2) ? Quelle est la place occupée par ces unités en acquisition d'une L2 (chapitre 3) ? Comment identifier et décrire les « erreurs » phraséologiques produites par des apprenants avancés du français L2 (chapitre 4) ?

Dans le premier chapitre, nous abordons le phénomène phraséologique sous l'angle de la description en français L1, afin d'en dégager les caractéristiques définitoires et de fournir à l'analyse des données de

corpus un ancrage théorique solide, héritier de la littérature tant francophone qu'anglo-saxonne. Si nous devions les définir ici en quelques mots, nous dirions que les *unités phraséologiques*, appelées aussi *séquences (semi-)figées*, sont des unités linguistiques polylexicales (i.e. constituées de plusieurs unités lexicales) qui se sont conventionalisées au fil du temps et dont les constituants co-apparaissent de manière contrainte dans l'usage.

Dans le second chapitre, nous proposons une méthode d'analyse permettant de mesurer le degré de figement des unités phraséologiques, à travers une étude de cas menée sur 200 séquences verbo-nominales construites avec le verbe *prendre* dans des corpus journalistiques et littéraires en français contemporain L1 (400 000 mots) (par ex., par ordre croissant de degré de figement : *prendre un fruit* > *prendre une décision* > *prendre contact* > *prendre le taureau par les cornes*).

Quelques notions issues du domaine des recherches en acquisition d'une L2 seront abordées dans le troisième chapitre. Une attention particulière y sera accordée au rôle que les unités phraséologiques jouent dans le processus acquisitionnel, et en particulier au stade avancé d'apprentissage d'une L2. Ces notions constituent un préalable théorique aux analyses sur corpus d'apprenants du français L2 du chapitre suivant.

L'objectif du quatrième chapitre est à la fois descriptif et méthodologique : il s'agit (i) de cerner les traits phraséologiques caractéristiques de l'interlangue avancée du français académique L2 en identifiant les « erreurs » spécifiques aux apprenants dans leur production d'unités phraséologiques (par ex. : *\*donner un coup d'œil* pour *jeter un coup d'œil*, *prendre un \*moment de réflexion* pour *prendre un temps de réflexion*) et (ii) de décrire étape par étape l'élaboration de l'outil d'analyse qui nous a permis d'identifier et de décrire ces « erreurs ». Cette analyse interlangagière consiste en l'analyse collocationnelle de 906 séquences verbo-nominales (dont 703 unités phraséologiques) construites avec les verbes à haute fréquence *prendre* et *donner*. Le corpus se compose de textes en français académique produits par des apprenants avancés anglophones (185 000 mots), néerlandophones (90 000 mots) et par des étudiants francophones natifs (185 000 mots). Nous évoquons en outre dans le dernier point de ce chapitre quelques pistes didactiques et pédagogiques.

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 ex. : *prendre des gants*) (par opposition au semi-figement

très répandu dans l'usage, par ex. *prendre la parole*), 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 (syntaxique, 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 s'inspirant, entre autres, de la méthode d'analyse paramétrique et statistique de données de corpus (décrite dans Degand et Bestgen, 2004), de la sémantique contextuelle (Rastier *et al.*, 1994), de la linguistique de corpus (McEnery et Wilson, 1996 ; Habert *et al.*, 1997 ; Kennedy, 1998), de l'analyse contrastive sur corpus informatisés d'apprenants (Granger, 2004) et de l'analyse d'erreurs (James, 1998).

Cette étude des unités phraséologiques s'inscrit dans une conception large du figement et porte une attention toute particulière aux séquences *semi-figées* (par ex. : *donner envie, prendre connaissance de qqch.*), 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.

Il est important de souligner que si la méthode adoptée tend bien à restreindre les biais méthodologiques liés aux analyses strictement quantitatives ou qualitatives, elle ne permet cependant pas d'éviter certains inconvénients liés à l'une ou l'autre approche. Outre le fait que nous avons constamment essayé de mettre le doigt, au fil des pages, sur ces inconvénients méthodologiques, nous mettrons en avant dans la conclusion de l'ouvrage les limites majeures de notre recherche, en même temps que nous y mettrons en lumière quelques perspectives susceptibles de prolonger ce travail.



## CHAPITRE 1

### Définir les unités phraséologiques

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) ou élargie (approche statistique) : 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. chapitre 4). Notre propos dans le présent chapitre est de déterminer quels sont les critères définitoires des unités phraséologiques et d'en examiner les principales catégories existantes. Nous nous attarderons en outre sur la notion de collocation, 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.

#### 1. Phraséologie au sens strict (approche fonctionnelle)

La phraséologie au sens strict a pour objet d'étude les *unités phraséologiques* (UP), i.e. toute unité polylexicale constituée de deux ou plusieurs mots graphiques catégoriellement liés, contigus ou non, qui co-apparaissent de manière préférentielle dans l'usage. Les tenants de l'approche fonctionnelle catégorisent prioritairement les UP sur la base du critère sémantique de non-compositionnalité (*vs* compositionnalité, i.e. *transparency*) et du critère lexical de restriction paradigmatique (*vs* ouverture des paradigmes, i.e. *commutability*).

Transparency refers to whether the elements of the combination and the combination itself have a literal or a non-literal meaning, and commutability refers to whether and to what degree the substitution of the elements of the combination is restricted. (Nesselhauf, 2005 : 14)

Cette conception correspond à une approche dite « phraséologique » (Nesselhauf, 2005) ou « fonctionnelle » (François et Manguin, 2006) du phénomène de figement. Elle est la perspective traditionnellement adoptée par les phraséologues de tradition européenne occidentale et d'Europe de l'Est, tels que Mel'čuk (1993, 1995, 1998), G. Gross (1996), Gréciano (1983, 1997a, 1997b), Fiala *et al.* (1997), Hausmann (1997), Martins-Baltar (1995, 1997), González Rey (2002),

Mejri (1997a, 2003a, 2005), Lamiroy (2003, 2008), Lamiroy et Klein (2005) ou Svensson (2004), pour ne citer qu'eux. González Rey (2002) et Gréciano (1997b) distinguent par ailleurs deux types de phraséologie : une phraséologie générale, relevant du langage courant, et une phraséologie terminologique ou spécialisée, relevant des langues de spécialité (par ex. : *porter plainte* (juridique) ; *établir un devis* (commerciale) ; *arrêt cardiaque* (médicale) ; *mordre le guidon* (cyclisme) ; *faire l'avion* (cyclisme) ; *mieux vaut règle que rente* (commerciale) ; *de mauvais contrat, longue dispute* (juridique)). Les UP en langue de spécialité sont appelées des « phraséotermes » par les deux auteures. Dans le monde anglo-saxon, nous retenons comme auteurs principaux adoptant une perspective phraséologique fonctionnelle Cowie (1998), Moon (1998), Nesselhauf (2004, 2005) et Wray (2002, 2008). S'inscrivent aussi généralement dans cette conception restreinte les psycholinguistes et les acquisitionnistes qui effectuent des recherches expérimentales sur les modes d'accès aux acceptions littérale et/ou figurée des unités phraséologiques à valeur idiomatique (Cacciari et Glucksberg, 1995 ; Everaert *et al.*, 1995 ; Gibbs, 1995 ; Langlotz, 2001 ; Schmitt, 2004).

L'approche fonctionnelle de la phraséologie envisage le phénomène du figement lexical par opposition à la liberté combinatoire, en situant ces deux pôles sur l'axe horizontal des relations syntagmatiques entre unités linguistiques. Une *séquence* de plusieurs mots graphiques est ainsi dite *libre* (SL) quand elle ne fait l'objet d'aucune restriction idiosyncrasique : elle se définit sémantiquement par son caractère transparent, syntaxiquement par sa flexibilité distributionnelle, lexicalement par une ouverture des paradigmes, pragmatiquement par sa liberté d'emploi en situation de communication. Les seules contraintes linguistiques qui la régissent sont les règles de syntaxe et les règles de restriction de sélection : « In a free phrase, the signified, the signifier, and the syntactics are constructed exclusively according to the general rules of the language ; a free phrase is thus 100 % compositional and replaceable by any other sufficiently synonymous phrase. » (Mel'čuk *et al.*, 1995 : 176). Nous présentons ci-dessous un cas de SL (ex. 1) extrait de notre corpus d'apprenants anglophones, avec les modifications que cette séquence accepte en cotexte (ex. 1a, 1b, 1c, 1d) :

Exemple de SL (FL2-AN, <TAN062>, Occ. 100)<sup>1</sup> :

(1) Dès qu'elle avait mis son maillot, Katrina **prit** une serviette<sup>2</sup> et se précipita en bas

---

<sup>1</sup> Pour chaque exemple cité nous indiquons entre parenthèses la langue source, le code du texte source et le numéro de l'occurrence. Les exemples sont en outre retranscrits dans leur version originelle, c'est-à-dire avec les erreurs (orthographiques, syntaxiques, lexicales, coquilles, etc.) produites par les apprenants.

(1a) Dès qu'elle/ il avait[/ eut] mis son maillot, Katrina(/ Marie/ la nageuse/ Pierre) **prit**(/ **s'empara de/ saisit/ attrapa**) une serviette(/ essuie de bain/ un drap de bain) et se précipita en bas [Substituabilité des constituants de la SVN par un constituant quasi-synonyme, avec Sujet [humain] et N Objet [prenable] : liberté sur l'axe paradigmatique et liberté au niveau des règles de restriction de sélection]

(1b) Dès qu'elle/ il avait[/ eut] mis son maillot, (la petite) Katrina (qui était pressée de plonger dans la piscine) **prit** (précipitamment) une (confortable) serviette (qu'elle venait d'acheter) et se précipita en bas [Insertion acceptable de modificateurs au sein de la SVN : liberté sur l'axe syntagmatique]

(1c) (Les serviettes se trouvaient dans l'armoire). Dès qu'elle avait[/ eut] mis son maillot, Katrina en prit une et se précipita en bas [Pronominalisation acceptable : liberté au niveau des règles syntaxiques]

(1d) Dès qu'elle avait [eut] mis son maillot, une serviette fut prise par Katrina et cette dernière se précipita en bas [Passivation acceptable : liberté au niveau des règles syntaxiques]

Par contraste, les UP sont marquées par un haut degré de collocabilité, le principe de *collocabilité* se définissant comme la « tendance syntactico-sémantique des mots à entrer en combinaison avec d'autres termes, en nombre limité parmi une grande quantité de combinaisons possibles » (González Rey, 2002 : 65-66). Envisageant le figement de manière scalaire sur un axe bipolaire, une des problématiques centrales de la phraséologie est de pouvoir déterminer où se situe le seuil minimal de figement :

La difficulté de décider du caractère figé d'une expression provient de son caractère graduel, en particulier dans la zone des figements minimaux où se situent les cas moins prototypiques : le véritable problème ne correspond pas à la limite supérieure mais à la limite inférieure du figement, c'est-à-dire non pas aux expressions fortement figées dont personne ne remet en cause le caractère idiomatique (ex. *porter le chapeau, vendre la mèche*), mais en revanche aux expressions peu figées ou semi-figées (ex. *je vous avoue que, attirer les foudres de qqn, se dessiner à l'horizon*, etc.). (Lamiroy et Klein, 2005 : 136)

Hausmann affirme ainsi que « l'idiomaticité des locutions, des collocations, des constructions, cela signifie – au lieu d'une combinabilité libre et illimitée des mots – toutes sortes de figements, de semi-figements, de restrictions, de préférences contextuelles » (1997 : 284-285).

---

<sup>2</sup> Dans les exemples, les caractères gras et les soulignements sont de nous. Le caractère gras est utilisé pour mettre en évidence le verbe à haute fréquence et le soulignement pour mettre en relief les syntagmes nominaux compléments du verbe. Le soulignement simple concerne les compléments sans préposition et le double soulignement les compléments prépositionnels.

Autrement dit, « l’idiomaticité n’est pas à la périphérie des faits de langue, elle est au cœur. Elle est l’essence même de la langue » (1997 : 284). Mejri (2003b) pose quant à lui des frontières notionnelles entre le domaine de la combinatoire libre et celui du « figement absolu » en distinguant trois stades de contraintes linguistiques, du moins contraignant au plus contraignant : (i) la syntaxe libre, considérée comme étant « le degré minimal du figement »<sup>3</sup>, avec les contraintes sur la sélection de traits sémantiques et la syntaxe des positions (i.e. la valence) ; (ii) l’attraction lexicale, « où certains items ne fonctionnent qu’avec d’autres soit parce qu’ils leur sont spécifiques (exemples : *intimer* ne fonctionne qu’avec *ordre*, ce qui donne *intimer l’ordre de* ; *perpétrer* avec *crime*, ce qui donne *perpétrer un crime*), soit parce que leur emploi avec certains items, bien qu’il ne soit pas propre, ne se conçoit qu’idiosyncrasiquement pour une langue (comme c’est le cas de *caresser*, *nourrir l’espoir* ou *négocier un virage*) » ; (iii) le figement à proprement parler, « où il y a une vraie saturation lexicale des positions, totale ou partielle [...] (saturation partielle : *tirer les vers du nez à quelqu’un* ; saturation totale : *chercher midi à quatorze heures*) » (Mejri, 2003b : 28). À l’instar de Mejri (2003b, 2005), Lamiroy et Klein (2005) soulignent le caractère contraignant de la syntaxe libre, lié aux règles de sélection de traits sémiques et à la valence verbale, et excluent du champ phraséologique les expressions sur lesquelles pèse ce seul type de contraintes. Les tenants de l’approche fonctionnelle fixent donc traditionnellement le seuil minimal du figement à une non-compositionnalité au moins partielle avec une restriction paradigmatique forte sur au moins un des deux constituants.

En outre, comme le souligne Mejri (2005 : 183-185), les propositions terminologiques foisonnent pour recouvrir le phénomène du figement. Nous avons pour notre part préféré désigner notre champ d’étude et notre objet d’étude par les termes respectifs de *phraséologie* et d’*unités phraséologiques*. Nous opérons par là explicitement une distinction par rapport au terme de *figement*, que nous réservons au processus langagier en tant que tel, et au terme de *séquences figées*, qui se réfère à une catégorie particulière d’unités phraséologiques hautement figées. Quelles sont les raisons majeures qui ont motivé ce choix terminologique ? Tout

---

<sup>3</sup> Bien que reconnaissant les contraintes de la syntaxe libre comme étant « le degré minimal de figement », Mejri conclut paradoxalement que le figement ne relève que du stade le plus contraignant. Il opère donc de manière implicite une distinction entre d’une part le *figement* comme processus langagier (i.e. les *contraintes*) et, d’autre part, les *figements* qui en sont le produit (i.e. les *séquences figées*).

<sup>4</sup> La saturation est totale si l’on ne tient pas compte du sujet dont les paradigmes sont ouverts. En effet, seules les règles de restriction de sélection lexicale N [+ humain] pèsent sur ce dernier.

d'abord, la phraséologie englobe à la fois les unités caractérisées par un degré minimal de figement sémantique (approche fonctionnelle) et les unités polylexicales caractérisées au minimum par une attraction lexicale (approche statistique) (cf. Mejri, 2003b, *supra*). Par ce choix, nous n'excluons donc *a priori* aucun type d'unités pouvant faire l'objet de recherches en phraséologie. Bien plus, nous considérons comme potentiellement phraséologique toute unité polylexicale dont les constituants entretiennent une relation contrainte sur l'axe syntagmatique. Ensuite, cette appellation a l'avantage d'être reconnue (et reconnaissable) internationalement par les différents spécialistes en phraséologie, quels que soient leur langue de recherche et leur domaine d'application. En bref, ce choix tente de répondre modestement au manque de consensus terminologique entre linguistes, bien que nous reconnaissons que ces désaccords sont souvent la conséquence involontaire de méthodes d'investigation et d'appartenance à l'un ou l'autre courant linguistique (linguistique de corpus, psycholinguistique, linguistique théorique, linguistique cognitive, etc.). Pour ne citer que les appellations les plus courantes, les UP, qui sont intégrées dans le champ de la *phraséologie* (González Rey, 2002) ou du *figement lexical* (Mejri, 1997a), sont désignées en français par les termes suivants : *expressions figées* (G. Gross, 1996 ; Klein et Rossari, 2003 ; Lamiroy, 2003, 2008 ; Klein, 2007 ; Svensson, 2004 ; González Rey, 2002, 2007), *séquences figées* (Mejri, 1997a), *phrases figées* (M. Gross, 1988), *expressions idiomatiques* (Gréciano, 1983), *locutions* (G. Gross, 1996 ; Fiala *et al.*, 1997 ; Martin, 1997 ; Martins-Baltar, 1997), *séquences préfabriquées* (Forsberg, 2005, 2006, 2008), *formules* (Krieg-Planque, 2009) ou encore *phrasèmes* (Mel'čuk, 1993 ; Mel'čuk *et al.*, 1995). En anglais, nous rencontrons les termes de *fixed expressions* (Moon, 1998), *idiomatic expressions* (Cacciari et Glucksberg, 1995), *prefabs* (De Cock, 2003, 2004), *phrasemes* (Mel'čuk, 1995, 1998), *formulaic sequences* (Wray, 2002, 2008 ; Schmitt, 2004 ; Corrigan *et al.*, 2009 ; Forsberg, 2009) ou *idioms* (Everaert *et al.*, 1995 ; Moon, 1998 ; Cowie, 1998). Notons que notre acception d'*unité phraséologique* correspond au terme *phrasème* tel que défini par Mel'čuk (1993 : 83) : « Un **phrasème** de la langue L est une expression multilexématique de L qui ne peut pas être produite, à partir d'une situation donnée ou d'un sens donné, selon un dictionnaire de mots de L et à partir des règles générales standard de L ». Ces deux appellations ont en commun la même racine lexicale, correspondant au terme anglais *phrase* (en français, *syntagme*). C'est cette acception qui a donné son nom à la phraséologie comme discipline linguistique s'occupant des contraintes pesant sur les relations entre unités linguistiques sur l'axe syntagmatique.

Nous proposons une première définition générale des UP :

Une *unité phraséologique* est une séquence polylexicale constituée de deux ou plusieurs mots graphiques catégoriellement liés, contigus ou non. Les UP se caractérisent linguistiquement par : (i) un certain degré de fixité syntaxique (blocage des propriétés transformationnelles et ordre des constituants inaltérable) ; et/ou (ii) un certain degré de figement sémantique (non-compositionnalité au moins partielle) ; et/ou (iii) un certain degré de figement lexical (restriction paradigmatique) ; et/ou (iv) une contrainte sur l'emploi en situation de communication. Ces contraintes se traduisent généralement par un certain degré de collocabilité mesurable statistiquement en termes de fréquence de cooccurrence des constituants.

Selon cette définition pluridimensionnelle, i.e. *polyfactorielle* (cf. Lamiroy et Klein, 2005 : 135), et scalaire du figement, nous opérons une distinction sur une échelle de figement entre les UP semi-figées et les UP hautement figées. Nous illustrons ci-dessous cette gradation en montrant quels types de modifications une occurrence semi-figée (ex. 2) et une occurrence figée (ex. 3) acceptent respectivement en termes de *fixité*, de *non-compositionnalité* et de *restriction lexicale*.

Exemple d'UP semi-figée (FL2-AN, <TAN141>, Occ. 227) :

(2) Les onze pays qui participent à ce projet **ont** pourtant **pris** un grand risque en acceptant la transition à l'euro

(2a) C'est un grand risque que les onze pays qui participent à ce projet **ont** pourtant **pris** en acceptant la transition à l'euro/ Un grand risque a pourtant **été pris** par les onze pays qui participent à ce projet en acceptant la transition à l'euro [Flexibilité syntaxique : topicalisation du N postverbal en *c'est... que* et transformation passive acceptables]

(2b) Les onze pays qui participent à ce projet **ont** pourtant **\*saisi/ \*attrapé** un grand risque en acceptant la transition à l'euro [Non-compositionnalité partielle : premier stade du processus d'abstraction sémantique de l'élément verbal, exprimé à travers la sélection d'une acception non prototypique de *prendre* (≠ « préhension ») ; maintien du sens prototypique de *risque*]

(2c) Les onze pays qui participent à ce projet **ont** pourtant **couru/ \*saisi/ \*tiré/ \*attrapé** un grand risque en acceptant la transition à l'euro [Restrictions paradigmatiques partielles : substituabilité par un V quasi-synonyme spécifique]

Exemple d'UP figée (FL2-AN, <TAN145>, Occ. 530) :

(3) Les Présidents et Premiers Ministres **donnent** un grand coup de collier afin de maintenir le cap dans une situation qui est sur le fil du rasoir

(3a) C'est un grand coup de collier que les Présidents et Premiers Ministres **donnent** afin de maintenir le cap dans une situation qui est sur le fil du ra-

soir [Flexibilité syntaxique : topicalisation du N postverbal en *c'est... que* acceptable]

(3b) Les Présidents et Premiers Ministres **\*confèrent/ \*offrent un grand** \*N afin de maintenir le cap dans une situation qui est sur le fil du rasoir [Non-compositionnalité totale : a-référenciation maximale de l'UP, exprimée à travers le caractère démotivé des constituants, l'inalysabilité et la dimension métaphorique de l'UP]

(3c) Les Présidents et Premiers Ministres **donnent/ \*confèrent/ \*offrent un/\*des/\*le grand/\*immense/\*énorme coup de collier/\*N** afin de maintenir le cap dans une situation qui est sur le fil du rasoir [Restrictions paradigmatiques totales : non-substituabilité des constituants de l'UP (verbe, déterminant, modifieur et nom) par un lexème quasi-synonyme]

### 1.1. Critères de figement

Dans son étude sur les critères de figement, Svensson (2004 : 29) n'hésite pas à souligner que « le *critère de figement* n'est pas un terme univoque. Tantôt il semble décrire ce qui est typique pour certains types d'expressions figées, tantôt il est utilisé pour vraiment trancher entre les syntagmes figés et les syntagmes non figés. » Nous avons pour notre part tenté de discerner, parmi les critères de figement repris dans les littératures antérieures, ceux qui caractérisaient le mieux les UP par opposition aux séquences libres (SL). À l'instar de G. Gross, nous pensons que la *polylexicalité* est le critère définitoire minimal des séquences figées au sens strict : « La première condition nécessaire pour qu'on puisse parler de figement est que l'on soit en présence d'une séquence de plusieurs mots et que ces mots aient, par ailleurs, une existence autonome » (G. Gross, 1996 : 9-10). Par ailleurs, les critères constitutifs des UP par rapport aux SL ont été dégagés par Lamiroy (2003), ainsi que par Svensson (2004) dans son étude sur les critères de figement en français contemporain. Les critères retenus comme pertinents par ces auteurs sont les suivants :

Lamiroy (2003 : 7)	Svensson (2004 : 42)	Définition
	Mémorisation	Rôle de la mémorisation pour les expressions figées
	Contexte unique	Rôle des mots utilisés uniquement dans les expressions figées
Opacité sémantique	Non-compositionnalité	Contribution au sens de l'expression par chaque mot qui y figure
	Syntaxe marquée	Importance des constructions syntaxiques rares
Paradigme limité	Blocage lexical	Impossibilité d'effectuer des commutations
Restrictions morphosyntaxiques	Blocage grammatical	Impossibilité de faire des changements syntaxiques

**Tableau 1. Critères de figement**

Les données sur lesquelles portera notre analyse de corpus seront étudiées en regard des critères linguistiques internes de *blocage grammatical/ restrictions morphosyntaxiques*, de *non-compositionnalité/ opacité sémantique* et de *blocage lexical/ paradigme limité*. Nous les désignerons quant à nous par les termes de *fixité* (figement syntaxique), de *non-compositionnalité* (figement sémantique) et de *restriction paradigmatique* (figement lexical). Ces trois types de contraintes sont à envisager chacun en termes de degrés de figement et s'inscrivent donc dans un continuum allant des séquences les plus libres aux plus figées. En effet, comme l'a souligné G. Gross, « le figement peut être partiel si la contrainte qui pèse sur une séquence donnée n'est pas absolue, s'il existe des degrés de liberté » (1996 : 154).

Sans imposer *a priori* de hiérarchie entre les dimensions langagières syntaxique, sémantique et lexicale, nous montrerons dans notre étude sur le degré de figement en français L1 (cf. chapitre 2) que l'indication du degré de figement se reflète tant dans les possibilités transformationnelles d'une séquence que dans son caractère plus ou moins compositionnel ou ses possibilités de sélection paradigmatique.

Nous mettrons tout d'abord en évidence deux critères récurrents dans les diverses études menées sur le figement : la mémorisation cognitive, i.e. l'intégration mémorielle des UP comme des unités autonomes, et l'institutionnalisation, i.e. la reconnaissance partagée du caractère conventionnel des UP par la communauté linguistique concernée. Ensuite, nous aborderons les quatre critères linguistiques que sont la polylexicalité, la fixité, la non-compositionnalité et la restriction paradigmatique.

### 1) *Mémorisation et institutionnalisation*

[Le processus de figement lexical consiste en] la transformation d'une séquence syntaxique, un syntagme plus ou moins long et complexe, apparais-

sant en discours, en une séquence dont le signifiant et le signifié sont stabilisés : comme tel, ces syntagmes, apparemment semblables à des énoncés libres, s'inscrivent dans la mémoire des locuteurs, qui peuvent les réutiliser sans avoir à les « produire » en discours lors de chaque énonciation ; de ce fait, ils fonctionnent plus ou moins comme les unités du lexique, ils y sont « codés ». (Mortureux, 2003 : 12-13)

Considérons tout d'abord le critère de *mémorisation*. Il est généralement admis que l'interprétation sémantique des UP repose sur des mécanismes d'ordre cognitif. La dimension cognitive du figement a été explorée par de nombreux psycholinguistes qui axent leurs recherches expérimentales sur le phénomène d'activation du sens figuré dans les séquences verbales figées idiomatiques. Les UP sont ainsi conçues comme des unités enregistrées en mémoire discursive *as a whole*, i.e. comme une unité autonome sémantiquement : « elles ne sont pas à définir comme un phénomène à fondement essentiellement linguistique mais comme un phénomène à fondement psycholinguistique, et plus exactement mémoriel » (Grunig, 1997 : 225).

Reprenant la définition qu'en donne Moon (1998 : 7), l'*institutionnalisation* est quant à elle « the process by which a string or formulation becomes recognized and accepted as a lexical item of the language (Bauer, 1983 : 48 and *passim*) : it is a necessary but not sufficient condition for a string to be classified as an FEI [Fixed Expression or Idiom] ». Les unités phraséologiques, qui se conventionalisent au fil du temps, peuvent trouver leur origine dans un changement linguistique impliquant un processus de lexicalisation (par ex. : *porte-manteau*, *fou rire*) (Mortureux, 2003), de grammaticalisation (par ex. : *aller* + Infinitif (mouvement > intention > futur), *ne... pas*, *à moins que*) (Marchello-Nizia, 2006 ; Torres Cacoullas, 2006 ; Bybee et Torres Cacoullas, 2009 ; Torres Cacoullas & Walker, 2009) ou de pragmatization (par ex. : *prenons l'exemple de...*, *t'sais*, *bon appétit !*) (Erman et Kotsinas, 1993 ; Dostie, 2004 ; Bolly, 2010, sous presse, à paraître). Elles peuvent en outre être introduites en langue via un mécanisme d'innovation langagière (par ex. : *zone piétonne*, *faire son come-back*).

Le figement est un processus qui obéit au temps et qui traduit d'une certaine manière les mécanismes profonds du fonctionnement des langues naturelles. C'est la langue qui, par le biais du discours, se renouvelle, s'adapte et évolue. L'étude de ce phénomène, de nature fondamentalement sémantique, pourrait contribuer à construire une sémantique capable de rendre compte des mécanismes cognitifs sous-jacents à l'intelligence humaine, qui demeure indissociablement liée aux facultés langagières. (Mejri, 1997b : 256)

Ces deux critères de *mémorisation* et d'*institutionnalisation*, que nous distinguons, sont en fait intimement liés : ils participent l'un et l'autre au processus dynamique de création langagière.

## 2) Polylexicalité

Une *unité polylexicale* se définit traditionnellement comme une catégorie grammaticale composée de plusieurs mots (lexèmes) séparés par un blanc, c'est-à-dire soudés de manière *discontinue* (cf. tableau 2) : « [l]es limites de cette polylexicalité sont généralement fixées, au niveau inférieur, à deux mots (*voir rouge*), et au niveau supérieur, à la phrase composée (*aller où le roi ne va qu'à pied*) et même au texte (prières, chansons, refrains, poèmes, etc.) » (González Rey, 2002 : 53). Mais c'est sans compter sur la tendance des mots à l'agglutination ou à la coalescence, qui voit plusieurs constituants initialement autonomes se fondre en un seul mot par une soudure graphique *continue*.

D'autres problèmes sont posés également tels que l'acceptation des mots composés séparés par des traits d'union (*porte-plume*) ou des apostrophes (*aujourd'hui*), ou bien des dérivés (*malheureux*) : s'agit-il d'unités phraséologiques au même titre que les constructions polylexicales séparées par des blancs ? (González Rey, 2002 : 53)

La problématique liée au figement qui est soulevée par le critère de polylexicalité tient à la délimitation floue existant entre les deux domaines de la morphologie et de la phraséologie<sup>5</sup>. Ce qui nous ramène à la question centrale et plus que jamais actuelle, e.a. en lexicométrie et en traitement automatique du langage, de la définition linguistique du *mot*. La définition de Meillet, selon laquelle « un mot est défini par l'association d'un sens donné à un ensemble donné de sons susceptible d'un emploi grammatical donné » (1921 : 30), s'avère ainsi insuffisante quand il s'agit d'étudier l'écrit, puisque l'unité linguistique y reste le mot graphique délimité par deux blancs.

Les noms complètement soudés ou liés par un trait d'union ne posent aucun problème. C'est l'autonomie orthographique des constituants de l'unité composée qui rend la tâche du linguiste très difficile. Or cette autonomie n'a rien avoir ni avec la fréquence des mots ni avec « l'âge » de l'unité. Des mots comme *pomme de terre*, *jeune fille* ou *trait d'union* illustrent parfaitement le caractère arbitraire de la soudure des unités polylexicales. (Mejri, 1997a : 132)

Si les *unités polylexicales* se distinguent des constructions syntaxiques libres, dont les constituants conservent leur autonomie syntaxique et sémantique respective, elles se distinguent aussi des formes

---

<sup>5</sup> Dans son acception large (cf. tableau 2), la morphologie inclut la *morphosyntaxe*, puisque cette dernière étudie les combinaisons morphématiques internes et les combinaisons syntaxiques externes du mot (graphique) qui permettent de « former des mots, des syntagmes et des phrases » (Dubois *et al.*, 2001 : 312).