

COLLOCATIONS : DEFINITIONS ET PROBLEMATIQUES

Estelle DUBREIL

LINA CNRS UMR 6241 / Université de Nantes

SOMMAIRE :

1. Introduction
2. Les phénomènes collocationnels
 - 2.1. Problèmes de définition
 - 2.1.1. Une nature dynamique
 - 2.1.2. Un continuum libre – figé
 - 2.1.3. Une approche prototypique
 - 2.2. Solutions de description
 - 2.2.1. Critères de figement
 - A. Critère de la transparence
 - B. Critère de l'arbitraire
 - 2.2.2. Critères définitoires
 - A. Critères de Tutin et Grossmann
 - B. Critère de Williams
 - 2.3. Deux approches du concepts de « collocation »
 - 2.3.1. Approche lexicologique - lexicographique
 - A. Définition restreinte
 - B. Typologies
 - 2.3.2. Approche « linguistique de corpus »
 - A. Historique
 - B. Définition large
 - C. Typologies
3. Les problématiques en linguistique appliquée
 - 3.1. La didactique des langues
 - 3.2. La lexicographie
 - 3.3. La terminologie
4. Les problématiques en linguistique théorique
 - 4.1. La modélisation
 - 4.1.1. Pourquoi et comment modéliser les collocation ?
 - 4.1.2. Le modèle des fonctions lexicales – FL
 - A. Fonctions lexicales paradigmatiques vs syntagmatiques
 - B. Fonctions lexicales standard vs non standard
 - C. Fonctions lexicales simples vs complexes vs configurations de FL
 - D. Limites du modèle des fonctions lexicales
 - 4.2. La motivation sémantique
 - 4.2.1. Les procédés sémantiques productifs
 - 4.2.2. Les régularités sémantiques
5. Proposition de motivation argumentative
 - 5.1. Problématique et filiation de la proposition
 - 5.2. Définition retenue
 - 5.2.1. Les collocations textuelles
 - 5.2.2. Les collocations libres vs restreintes
6. Conclusion

Résumé : L'article est consacré aux phénomènes collocationnels. Il en présente et explique les problèmes de définitions et les solutions de descriptions fondées sur des critères et des typologies

variables selon l'approche lexicologique-lexicographique ou linguistique de corpus retenue. Ces variations dans l'acceptabilité du concept sont également le fruit des différences de méthodes de traitement et d'analyse des phénomènes, qui varient elles-mêmes selon les problématiques appliquées ou théoriques posées, les domaines et la finalité des études. Dans cette optique, le chapitre se clôt sur l'étude de la motivation sémantique et sur une proposition de description de l'association collocationnelle en termes de motivation argumentative.

On dit plonger dans le désespoir, mais mettre en rage <*mettre en désespoir, *plonger dans la rage>.

On peut être gravement ou grièvement blessé, mais seulement gravement malade <*grièvement malade> (le fameux exemple de Bally) ; de plus, on a de grands blessés, alors qu'une blessure est grave, pas *grande. On est très fatigué, mais bien reposé <*très reposé (comme dans *Je suis très reposé)>, fabuleusement riche, mais bien connu <*bien riche, *fabuleusement connu> ; et ainsi de suite. En français, on fait un pas, en espagnol, on le donne (= dar un paso) et en anglais, on le prend (= take a step). La pluie est forte en français, mais lourde (= heavy rain) en anglais. On fait un rêve en français, mais on l'a en anglais (= have a dream) ; vous donnez un cours en français, mais vous l'enseignez (= teach a course) en anglais et vous le lisez (= citat' kurs) en russe (même si vous ne le lisez pas !). Toutes ces bizarreries d'usage sont connues depuis longtemps : ce sont des collocations – des expressions phraséologiques d'un certain type, qu'on ne peut pas prévoir et que la linguistique ne savait pas présenter de façon logique et facilement calculable.
(Mel'cuk 1997, p. 23)

1. Introduction

Conformément aux propos de Mel'cuk ci-dessus cités, les collocations sont des expressions appartenant à la catégorie des phénomènes phraséologiques. La phraséologie est un vaste domaine d'investigations linguistiques¹, que Rey et Chantreau, dans leur Dictionnaire des expressions et locutions, définissent comme un :

système de particularités expressives liées aux conditions sociales dans lesquelles la langue est actualisée, c'est-à-dire à des usages (Rey et Chantreau 1989, préface p. IX)

et que le TLFi décrit comme recouvrant :

l'ensemble des tournures typiques d'une langue, soit par leur fréquence, soit par leur caractère idiomatique (TLFi)

Particularités expressives, usages, tournures typiques, caractère idiomatique..., envisagée au sens lexicographique la phraséologie est un phénomène appartenant au patrimoine d'une langue (Mortchev-Bouveret 2004)..., envisagée au sens large (Cowie 1998) la phraséologie regroupe trois catégories de phénomènes : les expressions entièrement lexicalisées (ou figées), les expressions partiellement lexicalisées (ou figées) et les restrictions de sélection (syntaxiques et sémantiques), que Cruse appelle les affinités entre mots (Cruse 1986, p.18). Les phénomènes phraséologiques rassemblent une telle variété de séquences linguistiques que la plupart des courants linguistiques (russe², anglais (Cowie 1998) ou français (Gonzalez Rey 2002)) ont adopté un principe de description fondé sur l'élaboration de typologies.

¹ En témoigne la densité du programme proposé au colloque Phraséologie 2005, la phraséologie dans tous ses états, université catholique de Louvain (Belgique), 13-15 octobre 2005. <http://cecl.fltr.ucl.ac.be/PHRASEO/FRENCH/programmefr.html>

² Les phénomènes collocationnels ont également été traités (définitions et typologies) dans le cadre d'une tradition russe (Telija 1992, 1994, 1998 ; Baranov et Dobrovol'skii 1996).

La typologie des principales séquences phraséologiques proposée ci-dessous est empruntée à Grossmann et Tutin (Grossmann et Tutin 2003, pp. 6-8). Ils distinguent tout d'abord les séquences lexicales phrastiques (ou propositionnelles) des séquences lexicales syntagmatiques.

Les séquences lexicales phrastiques correspondent à des expressions entièrement lexicalisées. Totalement figées, elles n'admettent aucune substitution possible. D'une façon générale, elles comprennent :

- les proverbes – Ex : Pierre qui roule n'amasse pas mousse ;
- les lieux communs – Ex : À la guerre comme à la guerre ;
- les maximes – Ex : Tu ne tueras point ;
- les slogans – Ex : Tous pour un, un pour tous !
- les formules usuelles contextuelles (appelées pragmatèmes par Mel'cuk et holophrases par Galatanu) – Ex : Ça va ?

Sur le plan sémantique, les expressions figées se caractérisent d'abord par leur non-compositionnalité sémantique, c'est-à-dire que le sens du tout ne peut se calculer à partir de la signification des parties. Chaque unité lexicale perd donc sa capacité référentielle. Ensuite, en particulier pour les expressions figées nominales, le référent est souvent unique³.
Ex : nid d'ange

Dans ce cas, le référent de l'expression ne renvoie ni à celui de nid ni à celui d'ange, même si on devine le procédé métaphorique sous-jacent, et l'expression désigne un vêtement pour enfant.

Sur le plan syntaxique :

la contiguïté des éléments, la non-insertion de modifieurs, la non-substituabilité des éléments caractérisent un sous-ensemble de ces expressions, mais, comme les collocations, ces paramètres ne sont pas définitoires de la classe (Tutin et Grossmann 2002, p. 18).

Les séquences lexicales syntagmatiques sont des expressions partiellement lexicalisées (ou semi-figées). On peut y substituer un mot ou un syntagme (nominal, verbal, adverbial, adjectival). On y distingue les unités phraséologiques des groupements usuels (pour reprendre la terminologie introduite par Bally (Bally 1909)) par leur degré de figement.

Les unités phraséologiques désignent des groupes de mots où

les mots qui le(s) composent perdent toute signification et (...) l'ensemble seul en a une ; il faut en outre que cette signification soit nouvelle et n'équivale pas simplement à la somme des significations des éléments (ce qui du reste serait absurde) (Bally 1909, p.74)⁴

³ Tutin et Grossmann (Tutin et Grossmann 2002) relativisent le caractère discriminant de ce critère comparé aux collocations, dans la mesure où certains collocatifs ont une fonction plus « typante » que qualifiante. L'exemple pris est celui de café noir, qui désigne à la fois une propriété du café et une sorte de café. Dans le premier cas, le référent est bien unique, mais pas tout à fait dans le second.

⁴ Les variations terminologiques pour décrire ces phénomènes sont nombreuses, on trouve également expressions figées, locutions, phrasèmes complets (Mel'cuk 1998), unités polylexématiques (Corbin 1997)...

Selon l'opacité de la combinaison, on divise encore les unités phraséologiques en deux sous-catégories :

- les locutions figées opaques : le sens de l'expression n'est pas déductible de la signification des parties.

Ex : cordon-bleu

Un cordon bleu n'est ni un cordon ni *bleu*, le sens n'est pas compositionnel ;

- les locution figées imagées : le sens est imprédictible mais comporte une métaphore ou une métonymie perceptible.

Ex : pomme de terre ; manger les pissenlits par la racine

Le sens de ces expressions est non pas compositionnel mais partiellement dérivé de la signification première des parties.

Les groupements usuels ou collocations correspondent à des restrictions de sélection syntaxico-sémantique et se présentent comme :

des expressions semi-figées binaires, constituées de deux expressions linguistiques⁵
(Grossmann et Tutin 2003, p.8),

la base et le collocatif pour reprendre la terminologie d'Hausmann (Hausmann 1989, 1997).

À l'interface entre discours et langue, par leur instanciation contextuelle et par leur caractère idiomatique, les collocations posent certaines difficultés de définitions. Les solutions de descriptions proposées s'apparentent à une série de critères discutés et discutables selon l'approche choisie. Outre les variations de définitions et la diversité des critères permettant de les décrire, les collocations sont issues des différentes méthodes de traitement et d'analyse des phénomènes, qui varient elles-mêmes selon les problématiques posées, les domaines et la finalité des études. Qu'il s'agisse de problématiques appliquées au domaine didactique, lexicographique, terminologique ou de problématiques théoriques, les objectifs convergent : il s'agit d'une part de chercher à établir une modélisation de l'ensemble des phénomènes collocationnels, c'est-à-dire de mettre au point un modèle simultanément définitoire, descriptif et prédictif, et d'autre part de mettre à jour les mécanismes inhérents à la formation des collocations.

2. Les phénomènes collocationnels

2.1. Problèmes de définition

2.1.1. Une nature dynamique

À l'interface entre langue et discours, le concept de collocation est difficile à définir parce qu'il est dynamique. D'un point de vue quantitatif, les collocations sont importantes, car omniprésentes, tant en langue générale qu'en langue spécialisée, toutes langues

⁵ Dans le cas où une collocation associe plus de deux mots ou expressions, il s'agit généralement d'une superposition de collocations : avoir une faim de loup se décompose en deux collocations, avoir faim et une faim de loup.

confondues. En effet, une demi-page du *Nouvel Observateur* contient par exemple 21 collocations (Mel'cuk 2003, p. 26) et le *Dictionnaire spécialisé au français des affaires* (Binon et al. 2000) recense environ 3 000 vocables autour desquels gravitent pas moins de 11 000 collocations. D'un point de vue qualitatif, les collocations appartiennent bien au système de la langue, en tant que groupements usuels, syntaxiquement et sémantiquement contraints, mais leur développement change avec l'usage. Elles sont soumises à des contraintes sociolinguistiques de registre de langue : familier (péter un boulon, péter la gueule à quelqu'un), soutenu (hanter la conscience, satisfaire un besoin)... Elles témoignent également de l'appartenance d'un texte à un genre particulier (journalistique, être en état d'alerte, scientifique, construire une hypothèse (Pecman 2004a), littéraire, clore un chapitre) et/ou à un domaine de pratique (affaires, contrat de travail (Binon et al. 2000), économie, émettre des actions (Cohen 1986), Internet, visiter un portail (Meynard 2000), pharmacologie, dose d'attaque (Laporte et L'homme 1997)). Enfin, elles subissent également les phénomènes de mode ; toutes les questions sont aujourd'hui cruciales (Grossmann et Tutin 2003, p. 6).

Le langage humain est un processus et non un produit ; aussi, ce qui est habituel évolue au cours du temps selon des facteurs régionaux, techniques ou sociaux. De nouvelles restrictions apparaissent tous les jours ; en ce sens :

la collocation doit donc être conçue comme un phénomène dynamique dont la signification est négociée et non figée (Williams 2001b, p. 3).

2.1.2. Un continuum libre – figé

L'idée qu'il existe un stade intermédiaire de figement entre combinaisons libres et expressions figées n'est pas nouvelle. Elle apparaît déjà chez Bally, sous le terme de séries phraséologiques (Bally 1909, p. 66). De fait, les collocations s'intègrent dans une zone qui oscille entre le syntagme libre et le syntagme figé, ce qui génère un triple problème de définition.

Premièrement, les phénomènes collocationnels correspondent à des syntagmes semi-figés, se situant sur un continuum qui part du syntagme libre et qui va jusqu'au syntagme figé et dont il est difficile, voire impossible, de fixer les bornes. Ils sont situés dans une zone floue, que Thierry Fontenelle (Fontenelle 1997, p. 45) appelle le *fuzzy area*, entre la combinaison libre et l'expression figée, dont les frontières sont loin d'être étanches ; comment savoir où commence et où finit le semi-figé ?

Libre (envie de chocolat) → Semi-figé (ferme intention) → Figé (pomme de terre)

Deuxièmement, sous couvert du terme collocation, les phénomènes rassemblés sont de nature différente, de telle sorte qu'à l'intérieur de cette zone floue les restrictions syntagmatiques se distinguent également par différents degrés de figement. Sur un continuum allant du libre au figé⁶, les collocations sont alors qualifiées de régulière, transparente et opaque⁷.

Régulière (colère destructrice) → Transparente (rouge de colère) → Opaque (colère noire)

⁶ Une analyse diachronique montre une tendance des collocations à évoluer du statut de collocation régulière à celui de collocation opaque, ce qui corrobore leur aspect dynamique.

⁷ Nous recourons à la typologie sémantique proposée par Tutin et Grossmann (Tutin et Grossmann 2002).

Les limites entre les différentes catégories de collocations ne sont pas toujours faciles à établir, notamment à l'intérieur des collocations régulières. Certaines collocations s'associent parfois davantage à des combinaisons libres, elles s'approchent ainsi sensiblement de ce que l'on pourrait appeler à l'instar de Gross (Gross 1995), des classes d'objets :

Les domaines d'arguments de prédicats (verbaux, nominaux ou adjectivaux), quand ils sont libres, peuvent être exprimés en termes de classes d'objets (...) et non d'unités lexicales. Le verbe ressemeler a comme objet un élément de la classe des chaussures (soulier, sandale, botte, etc.). Le verbe lire est caractérisé par deux types de classes : celle des textes (récit, dissertation, roman, rédaction, poème, article, etc.) et celle des supports de textes (livres, journal, cahier, brochure, etc.) (Gross, 1996, pp. 78-79).

2.1.3. Une approche prototypique

Les collocations sont également des phénomènes intuitifs pour tout locuteur natif, pouvant être décrits en termes de prototypes⁸. En effet, à la manière de Martin (Martin 1997b) pour les locutions, il est possible de citer des cas typiques qui fassent l'unanimité et donc tentant de les utiliser comme supports permettant de couvrir les multiples facettes – pas toujours convergentes – du concept de collocation.

Ex : gravement malade ; grièvement blessé ; désirer ardemment ; aimer éperdument
Les célèbres exemples de Bally (Bally 1909) illustrent clairement cette dimension prototypique.

Toutefois, deux raisons poussent Tutin et Grossmann à nous mettre en garde contre cette tentation. D'une part :

une approche par prototype, pour intuitive qu'elle soit, n'apparaît en effet pas suffisamment contraignante pour avoir une capacité prédictive ou explicative (Grossmann et Tutin 2003, p. 9).

D'autre part, le choix d'un prototype ne ferait, selon eux :

que reporter l'étude systématique des critères linguistiques pourtant indispensable pour des traitements formels (Tutin et Grossmann 2002, p. 5).

À défaut de pouvoir définir objectivement et simplement les phénomènes collocationnels, les linguistes cherchent dans un premier temps à décrire ce qui peut être accepté en qualité de collocation et établissent à cet effet des critères.

2.2. Solutions de description

2.2.1. Critères de figement

La comparaison des critères de description proposés par Williams (Williams 2001b) et par Tutin et Grossmann (Tutin et Grossmann 2002) montre que seuls deux paramètres plaçant les phénomènes collocationnels sur une échelle de figement concordent, bien qu'ils soient

⁸ À noter que Tutin et Grossmann (Tutin et Grossmann 2002) évoquent le problème en n'envisageant qu'une catégorie prototypique possible, tandis que Williams (Williams 2001) en projette plusieurs selon le degré de figement de la restriction syntagmatique considérée, à l'intérieur de la classe formée par l'ensemble des phénomènes collocationnels.

discutés. Ces paramètres correspondent à leur caractère transparent et à leur aspect arbitraire.

A. Critère de la transparence

Le critère de la transparence ne s'applique que relativement aux collocations lexicales, et non aux collocations grammaticales comme se souvenir de, eu égard à la dimension sémantique qu'il mobilise. En effet, il réfère à leur caractère interprétable malgré leur imprévisibilité (cf. B. sur le critère de l'arbitraire) :

The term collocation will be used to refer to sequences of lexical items which habitually co-occur, but which are nonetheless fully transparent in the sense that each lexical constituent is also a semantic constituent (Cruse 1986).

De fait, n'importe quel locuteur peut en déduire le sens, alors même que, dans le cas d'un locuteur non natif, les collocations sont impossibles à produire en l'absence de mémorisation préalable.

Ex : buveur invétéré ; heavy drinker

Toutefois, ce critère semble échouer à caractériser l'ensemble de la classe collocationnelle puisque, parfois, le sens ne semble pas réellement transparent.

Ex : colère noire ; peur bleue

Dans ces cas, le sens du collocatif n'est pas immédiatement interprétable, le caractère transparent est donc discutable.

B. Critère de l'arbitraire

Les collocations sont également par nature arbitraires :

Collocations should be defined not just as "recurrent word combinations", but as "arbitrary word combinations" (Benson 1989, p. 3).

Le critère arbitraire qualifie la non-prédictibilité de l'association lexicale :

Si torrentielle peut en effet apparaître en cooccurrence avec pluie, cela paraît nettement moins naturel avec précipitations torrentielles. Cela est encore plus manifeste pour les collocations imagées du type appétit d'ogre ou faim de loup (vs. ?appétit de loup et ?faim d'ogre)⁹ (Tutin et Grossmann 2002, p. 9).

Cet aspect arbitraire est doublement significatif dès lors que l'on aborde le problème de la traduction des collocations. D'une part, d'une langue à l'autre les bases ne sélectionnent pas les mêmes collocatifs :

Ex : circulation intense se traduit en anglais par heavy traffic, non par *circulation lourde

D'autre part, il faut distinguer entre ce qui peut être librement traduit et ce qui exige la connaissance collocationnelle, sachant que ce qui est non arbitraire (un syntagme libre) dans une langue peut ne pas l'être dans une autre (Williams 2001b).

⁹ Conformément aux pratiques linguistiques traditionnelles, les représentations problématiques sont notées « ? » et les représentations asémantiques ou agrammaticales, « * ».

Par ailleurs, le caractère non prédictible de l'association lexicale caractérise principalement l'aspect sémantique qui unit le collocatif à la base, car du point de vue syntaxique les collocations suivent des patrons précis :

Par exemple, les collocations construites autour d'un nom incluront des adjectifs ou des verbes, mais probablement pas des adverbes ou des conjonctions (Tutin et Grossmann 2002, p. 9).

De même que pour le critère de transparence, ce critère nécessite d'être relativisé, au risque d'échouer à caractériser l'ensemble de la classe collocationnelle puisque, parfois, l'association lexicale ne semble pas réellement arbitraire.

Ex : l'âne braie

Du point de vue sémantique, l'aspect arbitraire de l'association est discutable, et la collocation suit un patron syntaxique connu.

2.2.2. Critères définitoires

A. Critères de Tutin et Grossmann

Les critères définitoires sont qualifiés par Tutin et Grossmann (Tutin et Grossmann 2002) de propriétés équivalant à des conditions nécessaires. Ils sont au nombre de trois : le caractère binaire de la collocation, la dissymétrie de ses composants et sa cooccurrence restreinte.

Le caractère binaire : la communauté scientifique s'accorde généralement à décrire une collocation comme étant composée de deux mots ou de deux lexies¹⁰. Mel'cuk utilise également le terme d'« expression linguistique bipartite AB » (Mel'cuk 2003, p. 23).

Dans le cas où la collocation englobe plus de deux mots, elle se laisse généralement décomposer en deux collocations distinctes.

Ex : essayer un échec cuisant se laisse décomposer en essayer un échec et échec cuisant
Toutefois, la fusion des collocations n'est pas toujours réelle.

Ex : prendre peur et peur bleue ne permettent pas de générer *prendre une peur bleue

Dans le cas où la collocation est composée d'un syntagme, Tutin et Grossmann (Tutin et Grossmann 2002, p. 10) préconisent de parler « d'association de constituants » plutôt que de mots ou de lexies.

Ex : fort comme un turc : fort est un mot et comme un turc est un syntagme

La dissymétrie des composants : pour Hausmann et Mel'cuk, les deux composants d'une collocation n'ont pas le même statut :

Dans la collocation, le statut des deux partenaires combinés n'est pas égal (Hausmann 1989, cité dans Tutin et Grossmann 2002, p. 10).

Pour Hausmann, la base est autonome, elle conserve sa signification, tandis que le collocatif est assujéti à la base. De même, pour Mel'cuk, le rapport entre les deux éléments d'une collocation est dissymétrique, puisque la base conserve toujours sa signification habituelle.

Ex : peur bleue

¹⁰ Cette bipartition n'est pas aussi tranchée chez Sinclair, « Collocation is the occurrence of two or more words within a short space of each other » (Sinclair 1991, p. 170), et Cruse (Cruse 1986), par exemple.

Dans cette collocation, le nom peur conserve sa signification habituelle – telle qu'elle est susceptible de figurer dans un dictionnaire de langue française –, tandis que la signification de l'adjectif bleu est quelque peu détournée, au profit d'un sens particulier à ce rapprochement lexical.

La cooccurrence restreinte : le critère concernant la cooccurrence restreinte apparaît comme la continuité du caractère dissymétrique précédemment exposé. En effet, le premier critère témoigne de la faculté de la base à conserver sa signification première, tandis que ce second critère montre la propension du deuxième élément (ou collocatif) à dépendre de la base de façon irrégulière et/ou contrainte, pour reprendre la terminologie de Mel'cuk, auquel nous empruntons les deux exemples ci-dessous (Mel'cuk 2003, p. 24).

Ex : café noir

En français, un café noir est un café – sans produit laitier –. Si café est choisi par le locuteur sans restriction et conserve de fait sa signification habituelle, non seulement le locuteur n'a pas le choix du mot à utiliser pour exprimer ce concept, mais en plus la signification de l'adjectif noir est quelque peu détournée. Sous l'entrée noir, aucun dictionnaire ne mentionne la signification « sans produit laitier », parce que cet adjectif ne prend ce sens que combiné avec café.

Ex : année bissextile

Dans la collocation année bissextile, – année qui a 366 jours –, l'adjectif bissextile a la signification de – qui a 366 jours –, mais il ne se combine qu'avec année (*an bissextil). Il est donc sélectionné à cause de sa signification extrêmement précise et de façon contrainte.

B. Critères de Williams

Les critères discutés par Williams (Williams 2001b) sont au nombre de deux : le caractère habituel de la collocation et sa bonne formation syntaxique.

Le caractère habituel : historiquement, le caractère habituel figure parmi les premiers critères énoncés pour la reconnaissance des collocations :

Collocations of a given word are statements of the habitual or customary places of that word in collocational order (Firth 1957, cité dans Williams 2001b, p. 2).

La difficulté posée par ce critère provient de ce qu'il est effectivement habituel de rencontrer de tels binômes, sans pouvoir instancier la réalité qu'habituel recouvre. Se contenter de ce « qui est (devenu) fréquent, régulier, coutumier » (TLFi) n'est pas satisfaisant, au sens où une telle définition ne permet pas de disposer de paramètres quantitativement discriminants, comparés aux cooccurrences de syntagmes libres. C'est pourquoi est généralement associée au critère habituel une dimension statistique, ce que Sinclair formule ainsi :

SIGNIFICANT COLLOCATION is regular collocation between two items, such that they occur more often than their respective frequencies and length of the text in which they appear would predict (Sinclair et al. 1970, p. 150)¹¹.

Cette dimension statistique prend généralement la forme d'une mesure, les plus courantes étant le z-score, le t-score et l'information mutuelle, que nous détaillons ultérieurement. La limite posée par ce critère tient au fait que chaque mesure statistique est appropriée pour

¹¹ À noter l'évolution du point de vue de Sinclair entre 1970 (two items) et 1991 (two or more words), concernant le caractère binaire des collocations.

une application donnée, ce paramètre définitoire du concept collocation demeure donc fluctuant, puisqu'il est fonction du calcul appliqué.

La bonne formation syntaxique : le critère de grammaticalité tel qu'il est présenté par Williams (Williams 2001b) n'est envisagé que dans la perspective d'une extraction automatique des collocations. De ce fait, ce critère est légitime d'un point de vue lexicographique pour limiter le bruit (Kjellmer 1984) ou, selon les besoins de l'étude, ne sélectionner que les collocations lexicales, comme l'outil Xtract le fait par ajout de filtres linguistiques (Smadja 1993). Toutefois, Williams fait remarquer que :

les collocations textuelles conformes à la définition de cooccurrence dans une fenêtre utilisée par Sinclair n'ont pas besoin de ce critère puisqu'elles recherchent des associations dans un champ sémantique (Williams 2003, p. 5).

Firth précise lui-même que les :

collocations of a given word are statement of the habitual or customary places of that word in collocational order but not in any other contextual order and emphatically not in any grammatical order (Firth 1957, cité dans Williams 2003, p. 5).

Plusieurs recherches sur les thèmes (Berry-Roghe 1973 ; Phillips 1985), sur la polysémie (Clear 1994) ou sur les réseaux collocationnels (Williams 2001b) se dispensent du critère de grammaticalité.

Les critères élaborés et appliqués sont fonction du cadre théorique et du domaine de pratique dans lesquels la recherche se situe. Par répercussion, les conditions nécessaires retenues influent sur les paramètres définitoires entrant dans la définition de collocation mobilisée, ce qui engendre une variation des définitions du concept selon les approches et les problématiques abordées.

2.3. Deux approches du concept de « collocation »

2.3.1. Approche lexicologique – lexicographique

A. Définition restreinte

D'une façon générale, dans la tradition lexicologique (Cruse 1986) et lexicographique (Hausmann 1989 ; Mel'cuk 1998), une conception « restreinte » de la collocation prédomine, entendue comme une association lexicale syntagmatique restreinte entre deux éléments entretenant une relation syntaxique. Cette définition prend sa source dans des études théoriques qui se caractérisent par la volonté d'aboutir à une définition formelle du concept. Reformulant à leur façon la formalisation élaborée par Mel'cuk (Mel'cuk 1998), Tutin et Grossmann proposent la définition suivante : « Une collocation est l'association d'une lexie L (mot simple ou phrasème¹²) et d'un constituant C (généralement une lexie, mais parfois un syntagme, par exemple « à couper au couteau » dans un brouillard à couper au couteau) entretenant une relation telle que :

- C (le collocatif) est sélectionné en production pour exprimer un sens donné en cooccurrence avec L (la base) ;

¹² Expression complètement lexicalisée dans la terminologie mel'cukienne.

- le sens L est habituel. » (Tutin et Grossmann 2002, p. 11).

Cette définition met en avant les paramètres sémantiques et syntaxiques inhérents à l'association collocationnelle, sans résoudre le problème de la diversité des éléments convoqués, auxquels il n'est pas pertinent d'appliquer un traitement homogène. C'est pourquoi, en complément de principes définitoires, les chercheurs s'appuient généralement sur des typologies classant les combinaisons lexicales possibles en fonction des paramètres sémantiques et/ou syntaxiques.

B. Typologies

Typologie selon des critères sémantiques : Tutin et Grossmann (Tutin et Grossmann 2002) proposent une « ébauche » de typologie des collocations fondée sur des critères sémantiques prenant en compte l'idiosyncrasie et l'idiomaticité¹³ du collocatif. Ils distinguent trois sortes de collocations :

- les collocations opaques. C a un sens particulier en cooccurrence avec L différent de sa signification en dehors de cette association. La base conserve sa signification habituelle. Le collocatif semble imprédictible et l'association arbitraire.

Ex : nuit blanche ; peur bleue

Les collocations opaques sont proches des expressions figées (comme cordon-bleu), à la différence près que la base reste interprétable. Les collocatifs sont souvent peu productifs ou donnent lieu à des séries limitées (comme bleu avec frousse ou trouille) ;

- les collocations transparentes. C semble conserver sa signification habituelle mais être difficilement prédictible du point de vue sémantique et/ou syntaxique.

Ex : désir ardent/?désir enflammé ; imprédictibilité sémantique (quasi-synonymie des adverbes sans substitution possible)

Ex : prendre peur ; imprédictibilité syntaxique (absence d'article, *prendre la peur)

Les collocations transparentes constituent le cas prototypique. Le décodage est rendu possible par un sens du collocatif proche de sa signification ou une périphrase comportant une comparaison ou une métaphore facilement interprétable ;

- les collocations régulières. C inclut la signification de L ou a une signification très générique. L'association est motivée et transparente. Soit le collocatif est à base unique, soit il porte sur un paradigme qui semble pouvoir être défini par des traits sémantiques.

Ex : nez aquilin ; base unique (aquilin se définit à l'aide de la base nez)

Ex : grande tristesse ; paradigme sémantique (trait intensif porté par grande)

Les collocations régulières sont proches des expressions libres (comme réelle tristesse).

¹³ L'idiomaticité est liée au caractère non standard de l'association linguistique, tandis que l'idiosyncrasie est liée au caractère unique de cette association (Grossmann et Tutin 2003, p. 9).

La conception étroite de la définition de collocation s'accompagne de restrictions syntaxiques dans la combinatoire des cooccurrents, soit envisagées en combinaison avec une typologie fondée sur des critères sémantiques¹⁴, soit envisagées seules.

Typologie selon des critères syntaxiques : l'association d'une base et d'un collocatif génère un syntagme qui mobilise des constituants spécifiques. L'association de ces constituants n'est pas arbitraire et permet de générer des syntagmes nominaux (nom + adjectif) ou verbaux (verbe + adverbe). C'est pourquoi Hausmann (Hausmann 1989) décrit la classe des collocations par une typologie des paires de constituants que les collocations peuvent générer. Suivant cette démarche, une collocation correspond à la combinaison caractéristique de deux mots dans une des structures suivantes¹⁵ :

- nom + adjectif (épithète) : amour platonique ;
- nom + (prép.) + nom : bourreau des cœurs ;
- nom (sujet) + verbe : la colère s'apaise ;
- verbe + nom (objet) : commettre une agression ;
- verbe + (prép.) + nom : rougir de honte (Hausmann 1999) ;
- verbe + adverbe : exploiter efficacement ;
- adverbe + adjectif : sexuellement transmissible.

Toutefois, Tutin et Grossmann (Tutin et Grossmann 2002, p. 13) font remarquer que la liste des patrons syntaxiques donnée par Hausmann n'est pas exhaustive, comparée, entre autres, aux collocations présentes dans le Dictionnaire explicatif et combinatoire – DEC – (Mel'cuk et al. 1984, 1988, 1992, 1999). Par exemple, les syntagmes construits sur des bases nominales admettent non pas uniquement des adjectifs mais également des prépositions comme de joie. De même, les syntagmes construits sur des bases adjectivales admettent non pas uniquement des adverbes mais aussi des syntagmes prépositionnels comme ivre de colère ou d'autres adjectifs comme ivre mort.

Les limites posées par ce type de typologie proviennent du fait que les patrons syntaxiques des collocations sont susceptibles de variations. Par exemple, la contiguïté de la base et du collocatif n'est pas systématique. On peut dire j'aime le steak bleu, mais également mon steak est vraiment bleu. De même, certaines collocations verbales autorisent des alternances syntaxiques. On peut dire mener une attaque, mais aussi l'attaque a été menée (passivation). Constatant ces limites, Tutin et Grossmann (Tutin et Grossmann 2002) émettent un doute quant à la capacité d'un modèle fondé sur une typologie des constituants syntaxiques à rendre compte de ces associations.

¹⁴ Il apparaît par exemple que les collocations régulières sont les moins contraintes syntaxiquement.

¹⁵ La typologie d'Hausmann repose sur des éléments définitoires identiques à ceux évoqués préalablement (section 2.3.1.1). Les collocations sont des combinaisons de mots binaires aux capacités combinatoires limitées, dont les composants (la base et le collocatif) entretiennent une relation hiérarchisée – la base est sémantiquement autonome et conserve son sens habituel –, tandis que le collocatif est sélectionné en fonction de la base pour exprimer un sens donné (Hausmann 1989, 1997).

2.3.2. Approche « linguistique de corpus »

A. Historique

L'origine contextualiste : en réaction contre les courants linguistiques structuralistes, qui évacuent la question du sens des mots, l'approche pragmatique prônée par Firth place l'étude de la langue en contexte. Influencé par les travaux anthropologiques de Malinowski, qui écrit que :

the real knowledge of a word comes through the practice of appropriately using it within a certain situation (Malinowski 1923, p. 325),

Firth reprend l'idée de contexte :

afin d'insister sur la nécessité d'exploiter ce qui est visible plutôt que d'élaborer des hypothèses abstraites sur la notion de sens (Williams 2003, p. 34).

Dans cette perspective, l'analyse des collocations participe de l'élaboration d'une théorie du sens contextualiste. Aussi, lorsque Firth (Firth 1957) introduit pour la première fois le terme collocation, il entend caractériser certains phénomènes linguistiques de cooccurrence qui engendrent des modifications de sens constatées sur les mots par leur relation de collocation en contexte :

Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words. One of the meanings of night is its collocability with dark, and of dark, of course, collocation with night (Firth 1957, p. 196).

Le contexte est accessible par le texte, donc, concrètement, l'analyse des collocations se fonde sur l'observation des relations sémantiques créées par le contexte, à la surface des textes.

Dans la lignée contextualiste, plusieurs recherches ont été menées en vue de mettre en pratique les principes théoriques de Firth. Halliday s'est attaché à poursuivre la description du système collocationnel en préservant l'importance de l'environnement textuel :

Without our being aware of it, each occurrence of a lexical item carries with it its own textual history, a particular collocational environment that has been built up in the course of the creation of the text and that will provide the context within which the item will be incarnated on its particular occasion. This environment determines the 'instantial meaning', or text meaning, of the item, a meaning which is unique to each specified instance (Halliday and Hasan 1976, p. 289).

Conformément aux postulats firthiens, le mot dérive son sens du contexte et influe simultanément sur ce contexte pour créer l'environnement textuel. La définition du texte est donc primordiale, puisque c'est sur lui que repose toute étude. Pour Halliday, le texte doit représenter une unité sémantique, un tout qui possède une cohésion. La collocation est un élément de cette cohésion, elle a un rôle fonctionnel de cohésion textuelle :

Laugh... joke, blade... sharp, ill... doctor (...) The cohesives effect of such pairs depends not so much on any systematic relationship as on their tendency to share the same lexical environment, to occur in COLLOCATION with one other. In general, any two lexical items having similar patterns of collocation – that is, tending to appear in similar contexts – will

generate a cohesive force if they occur in adjacent sentences (Halliday and Hasan 1976, pp. 285-286).

La fonction de la cohésion textuelle, assurée par les liens lexicaux entre énoncés proches et distants, a été confirmée par Hoey, dont l'origine des travaux repose sur la structuration et les stratégies rhétoriques des textes :

Collocation is the direct product of bonding (and of similar relations between sentences in and across texts). Collocation does not exist in language independent of its grounding in text, and we acquire collocations, as acquire other aspects of language, through encountering texts in the course of our lives (Hoey 1991, p. 219).

Les développements de l'école de Birmingham : à partir des années 1960, le centre de linguistique contextualiste s'est déplacé à Birmingham, sous l'égide de Sinclair (Sinclair 1991). Le lancement du projet OSTI (Sinclair et al. 1970, 2002) marque alors un tournant dans les recherches sur l'émergence du sens en contexte. En effet, ce rapport démontre la nécessité de travailler sur de grands ensembles de textes, les seuls à pouvoir révéler la fonction cohésive des phénomènes collocationnels en contexte. Ce rapport eut pour effet l'élaboration d'un autre projet : le COBUILD, premier grand corpus électronique constitué à des fins lexicographiques, dans lequel l'environnement contextuel des mots devient une réalité descriptible. La nature électronique du corpus ainsi constitué, le développement d'outils informatiques adaptés à l'exploration de ces données textuelles et les analyses qui en résultent ont été à la base d'une nouvelle discipline : la linguistique de corpus. Aussi, le véritable essor de l'école contextualiste est venu avec le développement de cette discipline dans les années 1980, puisque la linguistique de corpus offre des possibilités techniques nouvelles pour analyser de grandes quantités de données à partir desquelles peuvent émerger des régularités parmi lesquelles se trouvent les collocations. Cette approche empirique a permis de remplacer l'introspection fondée sur l'intuition du chercheur par une approche inductive : les hypothèses formulées s'appuient sur des données réelles et le corpus est considéré comme révélateur de la langue (Tognini-Bonelli 2001, p. 3).

B. Définition large

Firth a introduit et décrit le concept de collocation par des exemples, sans toutefois en proposer de réelle définition ni de formalisation. Nous avons vu que les recherches visant à appliquer les principes du sens contextualiste ont abouti à la mise en évidence d'un principe de cohésion textuelle selon lequel les collocations jouent un rôle essentiel, principe à partir duquel a émergé la notion de collocation textuelle. La définition de collocation textuelle mobilise le sens premier du concept de collocation, c'est-à-dire qu'elle désigne des relations de sens portées par des mots qui tendent à apparaître ensemble dans un contexte textuel donné. Aussi, sur le plan sémantique, l'analyse des collocations se dispense de toute étude lexicale préalable. Le syntagme en tant que formulation quasi idiomatique est délaissé au profit du sens contextuel du mot. De même, les relations analysées étant d'ordre pragmatique, dans la perspective contextualiste et néo-contextualiste, le critère de bonne formation syntaxique n'a tout simplement pas lieu d'être :

Les collocations textuelles conformant à la définition de cooccurrence dans une fenêtre utilisée par Sinclair n'ont pas besoin de ce critère puisqu'elles recherchent des associations dans un champ sémantique (Williams 2001b, p. 13).

Cette conception large de collocation :

ne restreint nullement l'association à la pure contiguïté et englobe aussi bien des associations lexicales paradigmatiques (médecin... hôpital) que des associations lexicales syntagmatiques (argument de poids), voire des associations incluant des mots grammaticaux (se souvenir de) (Tutin et Grossmann 2002, p. 8).

Toutefois, Sinclair considère que seules les collocations lexicales méritent l'appellation de collocation :

Collocation in the purest sense ... recognises only the lexical co-occurrence of words (Sinclair 1991, p. 170).

D'ailleurs, ce principe était déjà sous-entendu chez Firth car, parallèlement au terme collocation, il introduit le terme « colligation » pour désigner les collocations grammaticales (Firth 1957, p. 190). Cette distinction se retrouve également dans la typologie de Benson (Benson et al. 1986), qui définit la collocation grammaticale comme :

a phrase consisting of a dominant word (noun, adjective, verb) and a preposition or grammatical structure such as an infinitive or clause (Benson et al. 1986, préface p. IX).

Cette acception uniquement textuelle de collocation, cumulée aux développements de la linguistique de corpus, a eu pour effet d'éloigner progressivement les critères liés à la compétence des locuteurs natifs, liés aux aspects sémantiques (lexicaux) et syntaxiques, au profit d'une instanciation statistique de la collocation textuelle. Dans la lignée de Haskel (Haskel 1971), Jones et Sinclair (Jones and Sinclair 1974), Church et Hanks (Church and Hanks 1990), Aarts (Aarts 1990) et Sinclair (Sinclair 1991), la définition statistique de la collocation textuelle est la suivante :

A collocation is a group of two or more words which are to be found in proximity to each other significantly more often than one would predict, given the frequency of occurrence of each word taken individually. They may or may not be immediate neighbours, and the meaning of the whole may or may not be fully determined by the meaning of the parts (Kilgarriff 1992, pp. 29-30).

Typologie selon des critères contextualistes : la typologie proposée par Partington (Partington 1998) est qualifiée de contextualiste, dans la mesure où les trois types de collocations proposés correspondent à :

une interprétation différente de l'injonction de Firth selon laquelle " you shall know a word by the company it keeps " (Williams 1999b, p. 134).

- la collocation textuelle. Elle se définit comme la cooccurrence de deux mots ou plus dans un espace de texte court ;
- la collocation psychologique (ou associative). Elle doit son existence aux compétences linguistiques permettant de statuer intuitivement sur ce qui est « normal » et/ou « correct » et désigne par voie de conséquence l'association de deux mots ou plus, qui ne semble transparente qu'à un locuteur natif ;
- la collocation statistique. Elle correspond à la probabilité de rencontrer deux mots ou plus juxtaposés dans un espace de texte court.

En l'absence de critère syntaxique ou sémantique propre, la collocation textuelle est également une collocation statistique (Williams 1999b, p. 145).

Typologie selon des critères statistiques : les statistiques ne définissent pas les collocations mais sont un moyen d'évaluer leur seuil de signification, car :

seules des mesures quantitatives peuvent révéler les faits de langage dans une masse de données (Williams 2003, p. 37),

d'où le terme de collocation significative utilisé par Sinclair (Sinclair et al. 1970). Les analyses statistiques ont montré que les collocations sont significatives dans un espace de texte court qui correspond généralement à une fenêtre de + ou - 4 mots à partir du nœud. Les contextualistes parlent de nœud (Jones and Sinclair 1974), et non de base au sens d'élément stable dans une collocation comme Hausmann (Hausmann 1979), parce que le rôle sémantique joué par le nœud n'est pas pris en considération dans le calcul.

Il existe de nombreuses mesures d'association permettant d'identifier les unités lexicales qui cooccurrent le plus souvent dans une collection de textes. Les deux plus courants sont le t-score et l'information mutuelle – IM. L'algorithme t-score mesure le degré d'association entre deux mots, en favorisant l'émergence des combinaisons dont la fréquence est élevée. L'algorithme IM :

compares the probability of observing x and y together (the joint probability) with the probabilities of observing x and y independently (Church and Hanks 1990, p. 23).

À l'inverse, l'IM élimine les fréquences importantes au profit de fréquences plus faibles ; c'est une mesure de confiance. Les deux algorithmes sont donc complémentaires, ils permettent l'émergence d'unités différentes. Le t-score :

by identifying frequent and very reliable collocations, offers the lexicographer a semantic profile of the node word and a set of particular fixed phrases, grammatical frames and typical stereotyped combinations (Clear 1993, p. 282),

tandis que l'IM :

is best consulted for information about pairs which, though not likely to be typical of the usage of the node word, will be strongly associated and tend to form idioms, proverbs and technical phrases (Clear 1993, p. 282).

À l'image des autres critères (binaire, arbitraire, syntaxique...), le critère statistique a également ses limites (Wouden 1997). La mesure t-score pose problème car, en sélectionnant des unités à haute fréquence, elle ramène beaucoup de combinaisons composées de mots grammaticaux. C'est pourquoi la liste des combinaisons recueillies correspond plus exactement à un ensemble de candidats collocations – de collocations potentielles –, qu'il convient de nettoyer pour en extraire les collocations pertinentes au vu des objectifs de la recherche en cours. L'IM présente plusieurs inconvénients. Outre le fait de ne pouvoir différencier les binômes à sémantique compositionnelle des autres et regrouper les mots par classe grammaticale, l'IM a tendance à privilégier les collocations atypiques et bizarres. Pour limiter cet effet indésirable, une solution consiste à éliminer les hapax et les unités de faible fréquence du champ de calcul. Toutefois, cette mesure a les défauts de ses qualités, puisque :

la convergence de différentes expériences montre cependant qu'en donnant un poids important aux événements rares et en soulignant les emplois « spécialisés » le score de cooccurrence fait ressortir les expressions figées, ce qui est précieux dans une perspective lexicographique (Habert et al. 1997, p. 118).

Parallèlement aux mesures statistiques, qui permettent de « calculer les collocations », certaines typologies permettent de les classer relativement à ce critère statistique. Sinclair définit par exemple les collocations descendantes :

When a is node and b is collocate, I shall call his downward collocation – collocation of a with a less frequent word b (Sinclair 1991, p. 115)

et les collocations montantes :

When b is node and a is collocate, I shall call this upward collocation (Sinclair 1991, p. 116)

en fonction de la fréquence des composants (le nœud et le collocatif) de la collocation. Nous reprenons ses exemples pour illustrer les deux concepts. Prenons la collocation anglaise to give an audience, et supposons que les mots to give et audience reviennent respectivement cinq cents fois et trente fois dans un corpus de texte. Pour Sinclair, la relation collocationnelle entre ces deux composants change en fonction du mot à l'étude. Si le lexicographe s'intéresse au verbe to give et trouve que ce mot est fortement associé à audience, cette combinaison sera descendante (downward collocation) puisque le mot audience est moins fréquent que le mot à l'étude. En revanche, si le lexicographe s'intéresse au mot audience plutôt qu'à to give, la collocation est qualifiée de montante (upward collocation), puisque le verbe to give est plus fréquent que le substantif audience.

La mesure statistique des collocations textuelles s'accompagne d'un processus d'extraction automatique des données. Les fonctions statistiques sont souvent incorporées dans des outils informatiques, pour l'extraction automatique comme Xtract (Smadja 1993) ou pour l'extraction terminologique comme ACABIT (Daille 1996a, 1996b ; Daille and Williams 2001).

Typologie selon des critères sémantiques : une définition statistique de la collocation n'exclut pas la possibilité d'une classification indexée sur une typologie fondée sur des critères sémantiques. À la différence de celles fondées sur une définition restreinte de collocation, ces typologies sont plus larges, au sens où elles incluent généralement le concept de collocation ouverte ou libre. Par exemple, Cowie (Cowie 1978 ; Cowie et al. 1983) propose une typologie qui dissocie open collocation et restricted collocation. La première catégorie inclut des associations de mots dont les deux éléments conservent leur sens littéral et dont la combinatoire est assez libre pour sembler quasi illimitée.

Ex : le verbe to run se combine avec machine, business, horse, program...

La seconde catégorie inclut, de façon traditionnelle, des associations de mots dont l'un des constituants conserve son sens littéral tandis que l'autre a un sens figuré. Dans ce cas, le sens du mot A limite les possibilités combinatoires de l'élément B.

Ex : to explore a claim ; to kill a cigarette

De même, Howarth (Howarth 1998) propose une typologie qui répartit les phénomènes collocationnels en cinq catégories, selon leur degré progressif de restriction sémantique, allant des free collocations aux complete restrictions. La première catégorie :

Free collocations, for example, consist of two elements, both of which are used in their literal senses, and which permit the substitution of either element without affecting the sense of the

other component. In the following examples both the verbs publish and put are used in their primary senses, and the substitution of other noun objects (such as publish a book / an article / a prospectus or put a comment / an address / full stop) makes no difference to the sense of the verb. » (Howarth 1998, p. 167).

Tandis que la dernière catégorie – complete restrictions –, rassemble des cas où, par exemple, aucun autre nom ne peut être utilisé avec tel verbe employé dans un sens donné.

Ex : curry favour

Une typologie aussi large couvre naturellement l'ensemble des phénomènes collocationnels possibles, la difficulté à laquelle Howarth n'échappe pas étant de trancher sur le statut sémantique des éléments ainsi rassemblés :

A decision can then be taken concerning where the cut-off point is between restricted and free collocations. According to the present analysis collocability at Level 1 is felt to be insufficiently constrained to be restricted, so Levels 2-5 constitute the category of restricted collocations. » (Howarth 1998, p. 168).

Les variations de définitions et la diversité des critères permettant de décrire les collocations sont issues des différentes méthodes de traitement et d'analyse des phénomènes, qui varient elles-mêmes selon les problématiques posées, les domaines et la finalité des études. Qu'il s'agisse de problématiques appliquées ou théoriques, l'objectif ultime est le même ; il s'agit de chercher à en établir une modélisation au sens propre du terme, c'est-à-dire un modèle simultanément définitoire, descriptif et prédictif du système complexe formé par l'ensemble des phénomènes collocationnels.

3. Les problématiques en linguistique appliquée

3.1. La didactique des langues

Nous regroupons sous l'appellatif « didactique des langues » à la fois l'apprentissage et l'enseignement des langues maternelles et secondes.

Concernant l'apprentissage des collocations en langue maternelle, nous ne disposons à l'heure actuelle d'aucune étude systématique. Si, du point de vue compréhension-interprétation, ces phénomènes ne semblent pas poser de problème majeur selon leur caractère intuitif, en revanche, du point de vue production, les erreurs ou problèmes dans leur maniement sont courants. Par exemple, chez les jeunes enfants (4-5 ans), la tendance est à la normalisation sémantique cumulée à une certaine proximité phonétique des unités lexicales substituées.

Ex : vaisseau spécial pour vaisseau spatial

Ex : pot d'échauffement pour pot d'échappement

Ex : faire un coin d'œil pour faire un clin d'œil

Ex : à la fin du pestacle, on s'allume pour à la fin du spectacle, on salue (collocation textuelle paradigmatique, comme pour médecin... hôpital)

Pour les publics plus âgés, Grossmann et Tutin pensent que :

la modélisation visée devrait donc à la fois préciser les types collocatifs à l'origine de difficultés en réception et en production et fournir des aides " en ligne " lors de la production d'écrits (Grossmann et Tutin 2003, p. 10).

Concernant l'apprentissage des collocations en langue seconde, les études (Granger 1998 ; Howarth 1998) montrent qu'elles sont difficiles à acquérir et surtout difficiles à produire, même pour des apprenants de niveau avancé ayant été sensibilisés à ces phénomènes :

S. Granger plaide pour un enseignement contrastif, à partir d'une modélisation en langue maternelle permettant de prendre conscience du phénomène (Grossmann et Tutin 2003, p. 10).

Ce qui ne résout pas le problème de savoir comment les intégrer dans la pratique didactique:

Il s'agit là d'un domaine encore largement à défricher. La conception de ressources lexicales spécifiques pour les apprenants apparaît également indispensable et on remarquera que de nombreux projets lexicographiques dans le domaine des collocations ont été avant tout conçus à destination des apprenants en langue seconde (Grossmann et Tutin 2003, p. 10).

Par exemple, dans le domaine de l'enseignement et de l'apprentissage de langues intelligemment assisté par ordinateur (ELIAO), LogoTax développé par Ludewig, est une application qui essaie de franchir le fossé habituel entre dictionnaires classiques de mots et expressions usuelles dans un texte. Ce système s'adresse aux étudiants en allemand ayant acquis déjà un niveau avancé et les aide à composer leur propre dictionnaire collocatif en leur permettant de sélectionner et mémoriser des collocations qui les intéressent pour leur propre production de langage. En fouillant dans un corpus LogoTax essaie de trouver et de classer des références textuelles illustrant l'usage d'une collocation sélectionnée par l'étudiant et d'en générer ensuite une entrée (Ludewig 2001).

3.2. La lexicographie

Dans une perspective lexicographique, les lexicographes sont confrontés au problème du codage des collocations dans les dictionnaires. Ce codage est nécessaire compte tenu des difficultés que les collocations posent en didactique et en traductologie, par exemple, mais il est rendu difficile, premièrement à cause de la nature idiomatique et/ou idiosyncrasique (aspects syntaxiques et sémantiques) des phénomènes collocationnels et deuxièmement par le peu de modèles théoriques sur lesquels s'appuyer, capables d'en proposer une modélisation.

À la suite d'Hausmann (Hausmann 1989) et de Mel'cuk (Mel'cuk et al. 1984, 1988, 1992, 1999), la plupart des dictionnaires de collocations comme le BBI Dictionary of English Word Combinations (Benson et al. 1986) ou encore le Dictionnaires des cooccurrences (Beauchesne 2001) ont adopté un système de codage à partir de la base et non du collocatif. Par exemple, la collocation peur bleue est codée à partir de la base peur. Toutefois, l'accès à l'information par la base n'est pas une évidence. En espagnol, l'adjectif solo dans café solo, traduit en français par café noir, pourrait être décrit à l'aide d'une correspondance entre le sens sans lait et l'adjectif solo dans l'article lexicographique de café, sans qu'aucune information ne soit perdue (Alonso Ramos 2003, p. 45). Mais, si Alonso Ramos souligne que les collocations ne peuvent être décrites uniquement dans l'article lexicographique de la base, elle précise également que l'article du collocatif ne saurait suffire, dans la mesure où le lexicographe n'a pas les moyens d'y décrire l'interaction des sens des deux constituants.

Les collocations ont donné lieu à beaucoup d'ouvrages ; à titre d'exemples, nous citons les plus connus : le BBI Dictionary of English Word Combinaison (Benson et al. 1986), la base de données de Fontenelle (Fontenelle 1997, 2003) et le Dictionnaire explicatif et combinatoire (Mel'cuk et al. 1984, 1988, 1992, 1999).

Le BBI Dictionary of English Word Combination est un dictionnaire de collocations destiné aux apprenants et aux usagers de l'anglais qui regroupe 18 000 entrées et 90 000 collocations. Préalablement à la consultation du dictionnaire, le Practical Guide to the BBI informe l'utilisateur que deux types de collocations sont recensés. D'une part, les collocations grammaticales, répertoriées sous leur mot fort, sont divisées en huit classes tenant compte de phénomènes comme la transitivité, la passivation... D'autre part, les collocations lexicales, divisées en sept classes, formées par des enchaînements du type verbe + nom, adjectif + nom, nom + verbe... Les auteurs tiennent compte des phénomènes de polysémie. Ils donnent également des conseils pour trouver les réponses de façon optimale. Par exemple, concernant une collocation lexicale, si elle est composée d'un nom, chercher en premier lieu ce nom, si elle est composée de deux noms, chercher en premier lieu le deuxième nom, en l'absence de nom, chercher l'adjectif...

Dans le domaine de la lexicographie computationnelle et plus précisément de la lexicographie informatisée, la base de données de Fontenelle (Fontenelle 1997, 2003) transforme le dictionnaire Collins-Robert en base de données enrichie sémantiquement, afin de rendre accessible l'ensemble des connaissances métalinguistiques et collocationnelles. Dans cet ouvrage, l'auteur rend explicite l'organisation structurelle d'un domaine sémantique en décrivant la façon dont les éléments sont liés et en se servant de la théorie Sens-Texte (Mel'cuk 1997). En utilisant l'appareil descriptif des 56 fonctions lexicales standard, l'auteur a codé environ 70 000 paires d'items linguistiques, en montrant parfois les imprécisions de la théorie Sens-Texte et la nécessité de créer de nouvelles fonctions lexicales.

Le Dictionnaire explicatif et combinatoire (du Français contemporain) – DEC¹⁶ – est également un dictionnaire de collocations, mais élaboré dans une perspective lexicographique théorique, sous-tendu par le modèle des fonctions lexicales, lui-même inclus dans la théorie Sens-Texte. Selon cette théorie, un DEC doit fournir, de façon systématique et rigoureuse, les informations permettant à un locuteur de construire toutes les expressions linguistiques correctes de n'importe quelle pensée, et ce dans n'importe quel contexte : c'est un dictionnaire de production. En conséquence, les lexies françaises sont décrites sous toutes leurs facettes : sémantique, syntaxique, lexico-combinatoire, morphologique, etc., d'où une décomposition de chaque article en pas moins de dix zones.

Pour chaque lexie à décrire, il s'agit de donner la définition, d'indiquer les connotations, de spécifier la cooccurrence syntaxique (c'est-à-dire les données du régime syntaxique avec tous les détails des conditions d'utilisation contextuelles), de recenser la cooccurrence lexicale restreinte au moyen des fonctions lexicales, sans oublier les indications sur la flexion, les marques d'usage ainsi que les spécifications orthographiques, prosodiques, pragmatico-culturelles et même encyclopédiques.

Quatre volumes du DEC ont été publiés, ce qui correspond à la description formelle de 510 vocables du français. Chaque volume est divisé en deux parties, l'une théorique, l'autre descriptive. La partie théorique comporte des textes présentant les principes, règles et critères qui régissent les descriptions lexicales ou justifiant la démarche adoptée, ainsi que des analyses des phénomènes linguistiques pertinents. La partie descriptive constitue le dictionnaire à proprement parler : les descriptions lexicographiques formelles y sont présentées. Toutefois, le métalangage du DEC et sa formalisation extrême le rendent peu susceptible d'un emploi grand public. C'est pourquoi des versions simplifiées comme le

¹⁶ Cette présentation du DEC reprend la présentation générale proposée sur le site de l'université de Montréal : <http://www.olst.umontreal.ca/decfr.html>

DAFLES (Verlinde et al. 2003) et le dictionnaire de combinatoire – DiCo¹⁷ (Polguère 2000b, 2003, 2005) sont en cours d'élaboration. Le DiCo est une base de données lexicale du français, dont la finalité première est de décrire chaque lexie selon deux axes : les dérivations sémantiques qui la lient à d'autres lexies de la langue et les collocations qu'elle contrôle. Cette description s'accompagne d'une modélisation des structures syntaxiques régies par la lexie et d'une modélisation de son sens, sous forme d'étiquetage sémantique. Au-delà de la construction d'une base de données décrivant le lexique du français, le projet DiCo vise l'élaboration de deux types de produits lexicographiques. D'abord, la production automatique à partir du DiCo (par compilation de celui-ci) de lexiques de systèmes de TAL (Lareau 2002 ; Steinlin et al. 2004), ensuite la création de versions « vulgarisées » du DEC, à l'usage du grand public, objectif poursuivi par le projet intitulé lexique actif du français – LAF (Polguère 2000a).

3.3. La terminologie

Traditionnellement, partant des notions propres à un domaine (finances, bourse, informatique...), la tâche du terminologue est de dresser :

des listes d'unités lexicales rassemblées sous un terme de nature nominale et de les répartir soit par catégories grammaticales, soit par sens (L'homme 2003, p. 90),

pour en dresser un répertoire. De ce fait, les travaux en terminologie accordent plus de place à la construction de taxinomies qu'à l'aspect syntagmatique des mots ; d'où un intérêt minime pour les phénomènes phraséologiques, parmi lesquels figurent les collocations. Toutefois, depuis quelques années, plusieurs recherches portent sur les phénomènes combinatoires (syntaxiques et sémantiques) dans les textes spécialisés (Mortchev-Bouveret 1998). Deux grandes tendances se dégagent ; l'une qui vise à recenser et à classer les cooccurents dans un domaine particulier (Cohen 1986) ou à rendre compte du fonctionnement des termes en incluant dans les entrées terminographiques des phénomènes relevant de la combinatoire lexicale (Binon et al. 2000 ; Meynard 2000 ; Caigon 2001) ; l'autre qui cherche à caractériser ces groupes lexicaux dénommés combinaisons lexicales spécialisées – CLS – (Heid 1994 ; Laporte et L'homme 1997) et/ou à proposer des repères méthodologiques pour les situer par rapport aux collocations (terme réservé aux lexicographes) (L'homme 2000). Comme pour les lexicographes, dans ce dernier cas se pose à nouveau le problème du codage ou de l'encodage des CLS sous la base ou le cooccurrent (L'homme 2003). Cette problématique est d'autant plus présente qu'avec le développement de banques de données terminologiques informatisées (Termium ; Grand dictionnaire terminologique – GDT – ; Eurodicautom) les rubriques consacrées aux associations lexicales peuvent désormais rendre compte des multiples possibilités combinatoires de certains cooccurents, travail facilité du point de vue théorique par la modélisation possible des collocations.

4. Les problématiques en linguistique théorique

4.1. La modélisation

4.1.1. Pourquoi et comment modéliser les collocations ?

¹⁷ Le DiCo est consultable en ligne à l'adresse suivante : <http://olst.ling.umontreal.ca/dicouebe/>

La modélisation des collocations, même si elle se révèle délicate, est une étape obligée (Grossmann et Tutin 2003, p. 8).

De fait, elle est nécessaire d'un point de vue applicatif pour l'apprentissage des langues secondes, la lexicographie, la terminologie ou encore le TAL, nous venons de le voir. Elle est également indispensable d'un point de vue théorique pour mieux appréhender les phénomènes collocationnels sur le plan linguistique. Le caractère idiomatique et/ou idiosyncrasique des collocations posé par le problème de leur non-compositionnalité sémantique nécessite de mettre en évidence simultanément les caractéristiques et les mécanismes qui régulent les associations syntaxiques et sémantiques. D'une façon générale, une modélisation :

peut se définir comme une description formelle, explicite et cohérente des propriétés syntaxiques et sémantiques des collocations (Grossmann et Tutin 2003, p. 10).

En particulier, une modélisation devrait être adéquate à la fois au niveau descriptif et au niveau explicatif. Sur le plan descriptif¹⁸, il s'agit de disposer :

- d'une description sémantique fine de la collocation (définition) et de ses propriétés syntaxiques.

Ex : pour peur bleue, la définition doit être accompagnée de données syntaxiques mentionnant le caractère inséparable du nom et de l'adjectif ;

- d'une description de la relation sémantique liant le collocatif à la base, au moins dans les cas où la relation est partiellement compositionnelle.

Ex : pour peur bleue, la relation est intensive (mais de quel type ? Sur quel composant de la lexie repose-t-elle ?) ;

- d'une description de la relation syntaxique liant le collocatif à la base.

Ex : pour peur bleue, bleue est un modifieur dépendant de peur.

Sur le plan explicatif, les collocations étant difficilement prédictibles mais la plupart du temps facilement analysables, il s'agit de pouvoir rendre compte des procédés sémantiques d'association entre les deux composants.

4.1.2. Le modèle des fonctions lexicales – FL

Mel'cuk signale que, dans le modèle Sens-Texte :

la définition proposée pour la collocation couvre TOUS les cas d'expressions qui sont linguistiquement particulières et nécessitent l'inclusion dans le dictionnaire sous l'entrée de leur base, et SEULEMENT de tels cas : elle est donc nécessaire et suffisante » (Mel'cuk 2003, p. 26).

Il reprend la terminologie d'Hausmann concernant le constituant libre d'une collocation qu'il appelle également la base et l'autre constituant, le collocatif (sélectionné en fonction du premier). Dans ce modèle, les collocations sont des expressions semi-figées appelées semi-phrasèmes (café noir, commettre un crime) qui s'opposent à des locutions complètement ou partiellement figées appelées phrasèmes complets (pas question, prendre une veste) ou

¹⁸ Nous reprenons les trois critères proposés par Grossmann et Tutin (Grossmann et Tutin 2003, p. 11).

quasi-phrasèmes (point-virgule, bande dessinée) et aux pragmatèmes (bonne nuit, à consommer avant...).

La combinatoire lexicale restreinte constitue un élément fort problématique pour toute description dictionnaire et, par ricochet, pour la linguistique théorique. Le caractère capricieux et imprévisible de la cooccurrence lexicale est notoire (Mel'cuk 1997, p. 23).

Or l'approche Sens-Texte propose un modèle fonctionnel de ce phénomène, les fonctions lexicales – FL :

Puisque le trait définitoire central de collocation est la dépendance fonctionnelle du collocatif vis-à-vis de la base, il est évident que les fonctions lexicales (= FL) représentent LE moyen pour la description des collocations (Mel'cuk 2003, p. 27).

Une FL se définit comme :

une dépendance, ou correspondance, f qui associe à une unité lexicale L , appelée l'ARGUMENT de f , un ensemble d'unités lexicales $f(L)$ – la VALEUR de f . Chaque FL correspond à un sens très général (à la limite qui peut être zéro) et à un rôle syntaxique profond (Mel'cuk 1988, p. 6)¹⁹.

Les fonctions lexicales correspondent à des fonctions mathématiques dont la formule est du type : $f(x) = y$, où x est l'argument ou le mot clé et y , la valeur. Il existe trois binômes de FL, décrits en fonction de trois axes.

A. Fonctions lexicales paradigmatiques vs syntagmatiques

Les FL paradigmatiques vs syntagmatiques se distinguent selon le sens associé à f et le comportement syntaxique des éléments de $f(L)$. Les FL décrivent les propriétés combinatoires des unités lexicales sur les axes paradigmatiques et syntagmatiques (les cooccurrences lexicales restreintes, c'est-à-dire les collocations).

Une FL f paradigmatique s'utilise dans le texte – au lieu de – L .

Ex : fumeur, qui est S1(FUMER), s'utilise dans le texte au lieu du verbe fumer, permettant la paraphrase il fume = il est un fumeur

Les FL paradigmatiques modélisent donc des dérivations sémantiques de L , au sens propre (nom d'agent, nom de propriété, adjectif relatif...), comme au sens figuré (synonymes, antonymes...), et à ce titre ne sont donc pas pertinentes pour la description des collocations.

Une FL f syntagmatique s'utilise dans le texte – à côté de – L .

Ex : Magn(FUMER) = comme un sapeur s'utilise dans le texte à côté du verbe fumer, permettant l'énoncé il fume comme un sapeur

Les FL syntagmatiques modélisent donc la cooccurrence lexicale restreinte de L (intensifieur, verbes supports...) et, à ce titre, entrent directement dans la description des collocations.

B. Fonctions lexicales standard vs non standard

Les FL standard et non standard mobilisent un critère d'universalité de f .

Une FL standard est (en principe) universelle, au sens où elle apparaît dans toutes les langues. À ce titre, elle s'applique à beaucoup de mots clés, possède énormément

¹⁹ Pour une définition formalisée de fonction lexicale, consulter (Mel'cuk 2003, p. 28 ; Kahane et Polguère 2001).

d'éléments comme valeur de FL et participe au paraphrasage en tant qu'objet des règles universelles de paraphrase. Les FL standard ont des noms latins universels (Magn, du latin magnus, qui signifie grand, FL appelée intensificateur) et sont désignées par des symboles spéciaux. Il en existe une soixantaine (57 exactement).

Ex : S1(L) : nom d'agent (fumeur de fumer) ; Magn(L) : intensificateur = fumeur invétéré

Ex : Oper1(L) : verbe support (porter plainte)

En principe, une FL non standard n'apparaît pas (mais le peut !) dans toutes les langues. À ce titre, elle s'applique à peu de mots clés (voire à un seul), possède également peu d'éléments comme valeur de FL (voire un seul) et ne participe au paraphrasage – en ce sens que les règles universelles de paraphrase n'utilisent aucune FL non standard. Les FL non standard sont décrites par des minidéfinitions dans la langue où elles s'appliquent.

Ex : sans produit laitier(CAFE) = (café) noir

Ex : première chez un enfant qu'il perd vers l'âge de 7 ans(DENT) = (dent) de lait

Ex : qui tend à changer(TEMPS) = (temps) incertain

C. Fonctions lexicales simples vs complexes vs configurations de FL

Les FL simples vs complexes mobilisent un critère relatif à la structure syntaxique de f.

Une FL simple est constituée d'un seul nom de FL, tandis qu'une FL complexe représente une série de FL liées syntaxiquement qui ont le même mot clé et sont exprimées comme un tout ; une configuration de FL est une série de FL qui ne sont pas syntaxiquement liées, mais qui ont le même mot clé et sont exprimées aussi comme un tout (Mel'cuk 2003, p. 29).

Ex : FL simples standard : Magn(malade) = très, gravement

Ex : FL complexes : CausFunc1(CONFUSION) (=causer que Func1) : semer la confusion)

Ex : Configurations de FL : (Magn + IncepReal1)(LARMES) (=commencer à réaliser intensément) : fondre en larmes)

D. Limites du modèle des fonctions lexicales

Pour les FL syntagmatiques en particulier, le modèle des FL vise une description exhaustive, systématique et transparente des collocations contrôlées par chaque lexie vedette. En ce sens, le modèle des FL satisfait aux exigences imposées par le principe d'une modélisation : d'une part, les FL permettent de décrire formellement la relation syntaxique et sémantique qui lie le collocatif à la base, d'autre part, la description sémantique de la collocation est déduite de la description sémantique proposée par la FL. Toutefois, sans remettre l'apport fondamental pour la description des collocations que représentent les FL, Mel'cuk lui-même (Mel'cuk 2003) ne cache pas que le modèle comporte encore à l'heure actuelle quelques imperfections. Premièrement, si l'extrême majorité des collocations peut être décrite par les FL, il demeure un petit groupe de collocations qui sont couvertes par le régime du mot clé (de la base), et non par des FL de la lexie vedette.

Ex : assurance vie, où la vie est ce qui est assuré, donc un actant sémantique et syntaxique de ASSURANCE.

Deuxièmement, toutes les collocations décrites par des FL n'utilisent pas toujours des FL standard. Un ensemble important de collocations est couvert par les FL non standard, qui ne peuvent être théoriquement prévues et doivent donc être recueillies empiriquement, par une recherche minutieuse.

Au niveau de la description syntaxique, il arrive par exemple que la description proposée pour certains verbes supports ne soit pas suffisante. Comme leur nom l'indique, les verbes

supports servent de support à des prédicats nominaux (Gross 1981, 1991, 1993, 1994). Leur signification est générale ; ils n'ont pas de fonction prédicative. Ils apportent aux substantifs prédicatifs qu'ils accompagnent des informations relatives au temps, à la personne, au nombre, etc. Des verbes comme être, avoir, faire ou donner ont généralement un emploi de verbe support. Le verbe support permet d'actualiser le prédicat.

Ex : être heureux ; avoir peur ; faire un voyage ; donner une gifle

Par voie de conséquence, Gross parle également de noms supports (Gross 1986).

Ex : douanier → agent des douanes

Dans ces cas, le modèle des FL rencontre des difficultés à indiquer certaines informations relatives à la passivation, à la modification du nom ou encore aux changements de déterminants. Pourtant :

le traitement de ces phénomènes serait grandement amélioré par un traitement plus systématique des propriétés syntaxiques à partir du collocatif (Grossmann et Tutin 2003, p. 13 ; Tutin 1997).

Enfin, la littérature fait état de plusieurs remarques allant du métalangage, parfois complexe et opaque (Wanner 1996), aux règles de combinaisons et de configurations fantaisistes en passant par l'amélioration du marquage des registres et des niveaux de langue et de familiarité des lexies.

Disposant d'un modèle de formalisation possible des collocations, une autre problématique a trait à la motivation des deux éléments. En effet, au niveau sémantique, un intérêt doit être porté à l'étude des mécanismes permettant de générer les collocations.

4.2. La motivation sémantique

La question de la motivation sémantique des collocations est souvent évacuée dans les modèles lexicaux, tout en étant pourtant centrale en linguistique. Dans de nombreuses collocations, l'arbitraire n'est pas total, et le caractère analysable de l'association entre la base et le collocatif peut être catégorisé en deux grands types de motivations : les procédés sémantiques productifs et les régularités sémantiques.

4.2.1. Les procédés sémantiques productifs

Les procédés sémantiques productifs correspondent aux cas de la métaphore, de la comparaison et de la métonymie et concernent de ce fait les collocations imagées.

Ex : rouge de colère ; faim de loup ; océan de bonheur

Ces motivations :

ont leur source dans le principe cognitif permettant d'appréhender un phénomène sous l'angle d'un autre, dans l'ancrage anthropologique de la représentation, ou encore dans des stéréotypes sociaux (Grossmann et Tutin 2003, p. 14).

Par exemple, nombreuses sont les collocations construites par métaphores empruntées au domaine du bâtiment pour caractériser l'argumentation (Lakoff et Johnson 1985).

Ex : construire une argumentation ; les fondements de son argumentation

Ces procédés stylistiques sont en effet largement utilisés dans les associations collocationnelles. De fait, l'expression de la collection et de la quantification, l'appréciation méliorative ou dépréciative, l'intensité (Tutin et Grossmann 2002)... sont générées grâce à

des moules productifs qui reposent sur certains types de métaphores ou de comparaisons. Les collocations créées de cette manière sont donc partiellement régulières, et il est possible de préciser quelques-unes de leurs règles de formation.

En français, la collection et la quantification sont exprimées par un syntagme binominal en de dont le collocatif, le premier nom est collectif et revêt une valeur quantitative importante, d'où la valeur métaphorique. Dans ce cas, la valeur métaphorique tient au fait que le nom collectif perd son autonomie référentielle.

Ex : une forêt de mâts ; une armée de marmitons ; un essaim d'hélicoptères

Le choix du nom collectif n'est pas arbitraire et suit certains critères d'appropriation référentielle (*essaim de mâts).

Selon le même principe, il existe aussi des noms qui n'ont pas de valeur collective mais qui ont également une valeur quantitative acquise métaphoriquement.

Ex : une avalanche de questions ; un déluge de paroles

Pour pouvoir prédire ces associations, il faut préciser les conditions d'application référentielle liées au sémantisme des expressions. Un quantitatif d'origine métaphorique :

semble pouvoir être considéré comme un outil permettant d'exprimer la quantité, chaque fois qu'il est question d'« objets » (abstraites ou concrètes) qui nous « tombent dessus » en grande quantité, soudainement et par surprise (Tutin et Grossmann 2002, p. 16).

Les collocatifs quantitatifs sont donc sélectionnés en fonction de la cohérence de l'élément métaphorique avec le signifié de la base.

De même que pour la quantification, l'appréciation méliorative ou dépréciative se construit avec un syntagme binominal en de. Toutefois, à la différence du premier cas, cette fois, le collocatif est en deuxième position. Il suit la préposition pour marquer l'appréciation positive ou négative, comme dans une santé de fer, un morceau de roi, un caractère de cochon.

Les appréciatifs, fortement liés à des stéréotypes culturels ou à des informations encyclopédiques (une voix de stentor, une voix de sirène), ont un comportement différent, a priori moins facilement prédictible (Tutin et Grossmann 2002, p. 17).

Lorsque le collocatif introduit est un nom propre, le mécanisme stylistique mobilisé correspond généralement à une comparaison, qu'il s'agisse de structure binominale ou construite avec des adjectifs relationnels.

Ex : la force d'Hercule ; une force herculéenne

On trouve également des structures en comme pour évoquer l'appréciation.

Ex : fort comme un turc

Quelle que soit la structure syntaxique retenue, on se trouve soit dans le cas de collocations transparentes, lorsque la signification de la comparaison est largement partagée et repose sur des stéréotypes culturels que la langue a lexicalisés dans de nombreuses expressions (fer, acier = force), soit dans le cas de collocations opaques, lorsque la motivation est de type historique ou mythologique, et la collocation est forcément plus idiosyncrasique. Dans ce dernier cas, c'est plutôt l'adjectif qui sélectionne sémantiquement le nom.

Ex : de bénédictin sélectionne de manière standard travail et autorise minutie

Dans ce cas, il s'agit de collocations régulières, la signification de minutie étant incluse dans celle de bénédictin.

Concernant l'intensité, le système de production des collocatifs intensifs est très proche des appréciatifs. On retrouve, entre autres, des adjectifs relationnels, des structures binominales en de et des comparaisons en comme.

Autant pour les adjectifs relationnels la prédictibilité est possible et apparaît relativement transparente (Ex : un froid sibérien ; un froid polaire), autant pour les structures binominales et les comparaisons la situation est plus complexe. Tantôt la motivation est accessible et donc prédictible. Par exemple dans les structures binominales où le collocatif correspond à un nom d'animal incarnant un défaut (Ex : têtu comme une bourrique) ou une qualité (Ex : malin comme un singe) ; dans ce cas, la motivation est sous-tendue par un symbolisme culturel. De même, les cas de comparaisons en comme, où le collocatif est un intensifieur.

Ex : blanc comme neige

À l'inverse, il est des collocations fondées sur des structures binominales imagées qui sont imprédictibles. Par exemple, le collocatif peut être un nom d'animal, sans que celui-ci n'incarne de qualité « culturellement fondée » (Ex : une mémoire d'éléphant ; un froid de canard). De même, il est des cas de comparaisons quasiment figées qui sont idiomatiques (Ex : un temps à ne pas mettre un chien dehors) et donc imprédictibles.

Le problème est que la cohérence de tels réseaux est difficile à établir, ce qui signifie que ces collocations sont peu prédictibles. Premièrement, en général, plusieurs réseaux sémantiques sont simultanément mobilisés. Lakoff et Johnson montrent que l'argumentation est décrite par des métaphores liées à l'opposition fond vs surface, mais aussi par des métaphores de forme comme dans les grandes lignes de l'argumentation. Deuxièmement, la motivation est souvent obscurcie par des phénomènes liés à l'usage. Certaines collocations métaphoriques ou métonymiques perdent la transparence de leur motivation par l'usure et leur appartenance à un style noble « dépassé », créant un décalage de désuétude.

Ex : vie orageuse ; remords dévorant

De même, les collocations à valeur intensive construites par comparaisons ont tendance à être familières, sinon vulgaires.

Ex : mauvais comme la gale ; con comme la lune

Dans ces cas, le collocatif est sémantiquement motivé mais ne peut être intégré dans un réseau, car il est idiosyncrasique.

Face à ces limites :

un programme cohérent en la matière viserait à mettre en évidence, pour les collocations imagées, les réseaux collocationnels les plus productifs, à examiner la diversité de leurs réalisations syntaxiques et à spécifier les restrictions d'utilisation (Grossmann et Tutin 2003, p. 15).

Ainsi, beaucoup de collocations sont construites par processus figuraux comme la métaphore, la métonymie ou encore la comparaison. Il apparaît que, pour les collocatifs à valeur quantitative, la motivation sémantique est forte, tandis que, pour les collocatifs à valeur appréciative ou intensive, bien qu'ils recourent à des structures régulières, le caractère prédictif du collocatif n'est pas toujours assuré (Tutin et Grossmann 2002, p. 17).

4.2.2. Les régularités sémantiques

Bien qu'on ait souvent insisté sur le caractère irrégulier des collocations, il apparaît possible, dans de nombreux cas, de préciser les règles d'association entre la base et le collocatif, en fonction de leur sémantisme respectif (Grossmann et Tutin 2003, p. 15).

En particulier pour les collocations les moins imagées, les régularités sémantiques correspondent à des traits sémantiques récurrents (Grossmann et Tutin 2005).

Ex : affreux sens intensif avec des noms de sentiments à polarité négative comme désespoir
Dans leur étude sur les intensifieurs des « noms de joie », Grossmann et Tutin (Grossmann and Tutin à paraître) dégagent trois principes explicatifs des associations entre adjectif intensif – AdjIntens – et nom de joie – Ndj – : le renforcement, la convergence de polarité et la cohérence avec le type sémantique du Ndj. Ces trois types de motivations résident dans l'idée d'une « congruence sémantique », pour reprendre le terme des auteurs, ce qui correspond à une logique des signifiés lexicaux qui tendanciellement favorise les affinités. La congruence sémantique requiert que les dimensions sémantiques repérées pour le nom et pour l'adjectif ne soient pas contradictoires. Dans la mesure où l'idée de bonheur semble, de façon privilégiée, être liée à la durée, les collocatifs seront plutôt du type paisible, calme, s'opposant à l'idée même d'intensité, telle qu'elle peut s'exprimer à travers des adjectifs comme vif, aigu ; cela n'empêche évidemment pas des combinaisons du type vif sentiment de bonheur dans lesquelles vif caractérise sentiment et non bonheur. Par ailleurs, même pour les collocations les moins imagées, on retrouve parfois les réseaux métaphoriques. Par exemple la métaphore du fond vs surface, structure de nombreuses expressions.

Ex : la profondeur des sentiments ; être profondément ému

Le renforcement concerne par exemple les cas où le nom comporte déjà le trait intensif et où l'utilisation de l'adjectif intensif conduit à un renforcement, par l'utilisation d'un intensif adjectival de degré élevé comme un superlatif.

Ex : tristesse est associé à grande ; désespoir est associé à immense ou encore à total et non à grand, parce que le désespoir est grand par définition, donc l'intensif doit être plus fort qu'avec tristesse, par exemple.

Les auteurs illustrent la convergence de polarité avec le trait agréable. Si la base comporte le trait agréable, le collocatif revêt en principe le trait appréciatif (au sens mélioratif) ou un trait neutre. À l'inverse, si la base comporte le trait de polarité désagréable, le collocatif revêt le trait dépréciatif ou neutre. Par exemple, affreux est un adjectif de polarité dépréciative. Une étude sur corpus a montré qu'il n'exprime l'intensité qu'en cooccurrence avec un substantif orienté négativement.

Ex : affreux désespoir ; *affreux bonheur les auteurs précisent que cette combinaison est possible, stylistiquement marquée, mais que dans ce cas l'adjectif perd son statut d'intensif.

Là encore, une étude sur corpus a également montré que dans certains cas, en particulier lorsque l'adjectif est antéposé, le trait de polarité positive s'efface.

Ex : parfait bonheur ; parfait désespoir dans ces deux cas, les auteurs précisent que le sens de parfait qui est mobilisé correspond à la signification total et non à la signification excellent.

Enfin, le principe de cohérence avec le type sémantique du Ndj est un principe plus flou. Il concerne les cas où une base conserve une ambivalence d'ordre sémantique telle que la possibilité de dénoter une émotion, un état ou une qualité, par exemple. Dans ce cas, ce principe d'association montre que :

certaines contraintes de sélection liées aux AdjIntens peuvent favoriser ou bloquer l'association avec ces noms polysémiques (Grossmann and Tutin à paraître).

Ex : absolu ou total s'associent préférentiellement à un Ndj traduisant un état comme bonheur ou désespoir mais semblent incompatibles avec un Ndj comme chagrin.

La question de la motivation sémantique des collocations est donc traitée selon deux aspects, les procédés sémantiques productifs et les régularités sémantiques. Les trois mécanismes décrits dans les régularités sémantiques, notamment le renforcement et la convergence de polarité, sont également pertinents dans la perspective d'une description argumentative de la motivation des entités mobilisées dans certains phénomènes collocationnels.

5. Proposition de motivation argumentative

5.1. Problématique et filiation de la proposition

Adhérant au point de vue de Tutin et Grossmann concernant la nécessité d'étudier les mécanismes permettant de générer les collocations. L'idée de décrire argumentativement la motivation des entités mobilisées dans certains phénomènes collocationnels provient du pouvoir explicatif et prédictible²⁰ « potentiel » de la sémantique des possibles argumentatifs, modèle lexical de description de la signification (Galatanu 2004, 2006). La signification lexicale y est conçue comme une entité pluridimensionnelle : sa description mobilise simultanément des informations sémantiques, internes à la langue, et des informations discursives, externes donc contextuelles. Le caractère mixte du modèle repose sur une interaction nécessaire entre le sémantique et le discursif. Aussi, le recours à la sémantique des possibles argumentatifs inscrit logiquement cette recherche dans la filiation contextualiste. En effet, de notre point de vue, l'approche contextualiste des phénomènes collocationnels revêt également un caractère mixte, puisque l'analyse des collocations se fonde sur l'observation empirique des relations sémantiques – dimension sémantique –, créées par le contexte – dimension discursive –. Rappelons que lorsque Firth introduit pour la première fois le terme collocation, il entend caractériser certains phénomènes linguistiques de cooccurrence qui engendrent des modifications de sens constatées sur les mots par leur relation de collocation en contexte. L'approche contextualiste étudie les collocations dites textuelles, eu égard au phénomène de cohésion textuelle, inhérent à toute unité sémantique formée par le texte, dont ces collocations rendent compte et auquel elles participent. Pour les mêmes raisons, l'hypothèse d'une motivation argumentative des entités collocationnelles est propre à désigner les collocations textuelles comme objet d'étude.

5.2. Définition retenue

5.2.1. Les collocations textuelles

Dans la perspective contextualiste et néo-contextualiste, la définition de collocation textuelle mobilise le sens premier du concept de collocation, c'est-à-dire qu'elle désigne des relations de sens portées par des mots qui tendent à apparaître ensemble dans un contexte textuel donné. Firth et Sinclair considèrent que seules les collocations lexicales méritent l'appellation de collocation. Pour notre part, les colligations étant constituées de mots grammaticaux dénués de contenu notionnel, elles sont impropres à une description argumentative fondée sur la signification des entités qui la composent. Aussi, nous laissons les collocations grammaticales de côté, au profit des collocations lexicales, construites avec des mots pleins, susceptibles de contenir des enchaînements argumentatifs séquentiels. Par ailleurs, les

²⁰ Prédictible est employé au sens de calculable et rejoint par ce fait, l'une des perspectives de recherches mentionnées par Mel'cuk (Mel'cuk 1997, p. 23), le facilement en moins (cf. Présentation de cet article) !

relations analysées étant d'ordre argumentatif, a fortiori pragmatique, nous considérons également que le critère de bonne formation syntaxique n'a pas lieu d'être.

Les développements successifs de la linguistique de corpus ont eu pour effet d'éloigner progressivement les critères liés à la compétence des locuteurs natifs, aux aspects sémantiques et syntaxiques, tout en privilégiant l'instanciation statistique de la collocation textuelle. Dans la lignée de Kilgarriff, Sinclair, Hoey (etc.), nous attribuons au critère statistique un caractère discriminant, sans toutefois recourir aux mesures statistiques t-score et IM, généralement réservées aux procédures d'extraction automatique, ce qui n'est pas le cas dans ce travail. Au contraire, le critère statistique a fait lui-même l'objet d'une révision dans une perspective qualitative (Dubreil 2005a, 2005b). Conformément aux analyses statistiques menées par Jones et Sinclair, les collocations jugées significatives sont retenues dans un espace de texte court qui correspond à une fenêtre de plus ou moins 4 mots autour du nœud. Le terme « nœud » n'est employé que pour désigner l'unité recherchée par concordance. Contrairement à la terminologie contextualiste généralement utilisée, nous conservons les termes de base et de collocatif introduits par Hausmann, précisément parce que le rôle sémantique joué par la base intervient dans les processus sémantico-discursifs analysés.

À cette acception statistique de la collocation textuelle, nous ajoutons une distinction fondée sur une typologie sémantique.

5.2.2. Les collocations libres vs restreintes

Les typologies sémantiques dédiées à la description – classification des collocations textuelles sont généralement larges, au sens où les distinctions qu'elles impliquent couvrent des phénomènes collocationnels très différents. Nous reprenons celle de Cowie, qui dissocie les collocations libres des collocations restreintes. La première catégorie inclut des associations de mots dont les deux éléments conservent leur sens littéral et dont la combinatoire est assez libre pour sembler quasi illimitée. La seconde catégorie inclut, de façon traditionnelle, des associations de mots dont l'un des constituants conserve son sens littéral tandis que l'autre a un sens figuré ; dans ce cas, le sens du mot A limite les possibilités combinatoires de l'élément B.

Les problématiques de recherches en linguistique théorique montrent que dans de nombreuses collocations, l'arbitraire n'est pas total. Les collocations libres sont même définies par le caractère prédictible de l'association entre la base et le collocatifs :

Cooccurrence that are free and predictable in the sense that they can be formed without any help by learner of German who knows these lexical items and who has mastered the elements of German grammar (Hausmann 1979, cité dans Benson 1989, pp. 2-3).

Toutefois, cette prédictibilité n'a pas de réalité autre que « psychologique » ou « associative », au sens défini par Partington (Partington 1998). Concernant les collocations restreintes, englobant les collocations opaques, transparentes et régulières définies par Grossmann et Tutin, les auteurs ont montré le caractère analysable de l'association entre la base et le collocatif par motivation sémantique. En particulier pour les collocations régulières, les processus sémantiques en œuvre correspondent à des traits sémantiques récurrents permettant également une forme de prédictibilité.

La sémantique des possibles argumentatifs permettant de formuler une hypothèse sur la nature sémantico-discursive de la relation qui unit la base au collocatif dans les collocations textuelles (Dubreil 2007a, 2007b), nous entrevoyons la possibilité de mettre à jour un même principe descriptif – explicatif, voire prédictible pour les collocations libres et les collocations restreintes. Ce faisant, les collocations libres ne seraient pas « réellement » libres, au sens où leurs associations seraient contraintes par un principe argumentatif, principe également sous-jacent à leur prédictibilité, autre que psychologique ; tandis que la prédictibilité des collocations restreintes reposerait sur un nouveau mécanisme de génération, autre que syntaxique et/ou sémantique.

6. Conclusion

À l'interface entre la langue et le discours, les collocations posent la question du statut de l'irrégularité dans le lexique. Leur caractère idiomatique et idiosyncrasique est à l'origine de la difficulté à les définir, difficulté accrue par la nature dynamique et le caractère fluctuant des phénomènes qu'englobe le concept de collocation. Aussi, les solutions de descriptions adoptées prennent la forme d'une série de critères discutés et discutables selon l'approche choisie, soit lexicologique – lexicographique, soit linguistique de corpus. Ces deux approches retiennent une acception différente du concept de collocation. La première est caractérisée de « restreinte », fondée sur des paramètres syntaxico-sémantiques, orientés vers une définition formelle du concept. La seconde est qualifiée de « large », reposant essentiellement sur des critères statistiques hérités d'une conception contextualiste du sens, à laquelle les collocations participent par la fonction cohésive qu'elles assurent au sein des textes. Chacune de ces approches privilégie le recours à ses propres typologies, compléments nécessaires à la délimitation et à la catégorisation des phénomènes observés. Outre les différentes approches, les variations de définitions et la diversité des critères permettant de les décrire, les collocations sont issues des différentes méthodes de traitement et d'analyse des phénomènes, qui varient elles-mêmes selon les problématiques posées, les domaines et la finalité des études. Qu'il s'agisse de problématiques appliquées au domaine didactique, lexicographique, terminologique, ou de problématiques théoriques, les objectifs convergent : il s'agit d'une part de chercher à établir une modélisation de l'ensemble des phénomènes collocationnels, c'est-à-dire de mettre au point un modèle simultanément définitoire, descriptif et prédictif, et d'autre part, de mettre à jour les mécanismes inhérents à la formation des collocations. De fait, la question de la motivation notamment sémantique des collocations est souvent évacuée dans les modèles lexicaux, tout en étant pourtant centrale en linguistique.

BIBLIOGRAPHIE

[Aarts 1990] J. Aarts, *Corpus Linguistics: an appraisal*, in *Computers in Literary and Linguistics Research*, J. Hamesse and A. Zampolli (eds.), Paris, Champion Slatkine, pp.13-28.

[Alonso Ramos 2003] M. Alonso Ramos, *La nature des collocatifs : leur statut en tant qu'unités lexicales*, dans *Les collocations : analyse et traitement*, F. Grossmann et A. Tutin (éds.), *Travaux et recherches en linguistique appliquée*, Amsterdam, de Werelt, pp. 45-60.

[Bahns 1993] J. Bahns, *Lexical collocations: a contrastive view*, *ELT Journal*, Vol. 47(1), pp. 56-63.

[Bally 1909] C. Bally, *Traité de stylistique française*, Genève, Paris, Georg et Klincksieck.

- [Bally 1932] C. Bally, *Linguistique générale et linguistique française*, Berne, Francke.
- [Baranov and Dobrovol'skii 1996] A.N. Baranov and D.O. Dobrovol'skii, *Cognitive modeling of actual meaning in the field of phraseology*, *Journal of Pragmatics*, Vol. 25(3), pp. 409-429.
- [Beauchesne 2001] J. Beauchesne, *Dictionnaires des cooccurrences*, Montréal – Toronto, Guérin.
- [Benson 1989] M. Benson, *The Structure of the Collocational Dictionary*, *International Journal of Lexicography*, 2-1, pp. 1-14.
- [Benson et al. 1986] M. Benson, E. Benson and R. Ilson, *The BBI Dictionary of English Word Combinations*, Amsterdam, John Benjamins (2nd edition, 1997).
- [Berry-Roghe 1973] G.L.M. Berry-Roghe, *The computation of Collocations and their Relevance in Lexical Studies*, in A.J. Aitken, R. Bailey and N. Hamilton-Smith (eds.), *The Computer and Literary Studies*, Edinburgh, Edinburgh University Press .
- [Benson 1985] M. Benson, *Collocations and Idioms*, in R. Ilson (ed.), *Dictionaries, Lexicography and Language Learning*, Oxford, Pergamon, The British Council, pp. 61-68.
- [Binon et al. 2000] J. Binon, S. Verlinde, J. Van Dyck et A. Bertels, *Dictionnaire d'apprentissage du français des affaires*, Paris, Didier.
- [Church and Hanks 1990] K.W. Church and P. Hanks, *Word Association Norms, Mutual Information, and Lexicography*, *Computational Linguistics*, 16(1), pp. 22-29.
- [Clear 1993] J. Clear, *From Firth principles: Computational tools for the study of collocation*, in M. Baker, G. Francis, and E. Tognini-Bonelli (eds.), *Text and technology: In honour of John Sinclair*, Amsterdam, John Benjamins.
- [Clear 1994] J. Clear, *I can't see the sense in a large corpus*, *COMPLEX '94 - Proceedings of the 3rd International Conference on Computational Lexicography*, Hungary, Budapest, pp. 33-48.
- [Cohen 1986] B. Cohen, *Lexique de cooccurrents – Bourse et conjuncture économique*, Montréal, Linguatex.
- [Corbin 1997] D. Corbin, *Locutions, composés, unités polylexématiques : lexicalisation et mode de construction*, dans M. Martins-Baltar (éd.), *La locution, entre langue et usages*, Fontenay-St-Cloud, ENS Editions, pp. 55-102.
- [Cowie 1978] A. Cowie, *The place of illustrative material and collocations in the design of a learner's dictionary*, in P. Strevens (ed.), *In Honour of A.S. Hornby*, Oxford, Oxford University Press.
- [Cowie 1998] A. Cowie, *Phraseology: Theory, Analysis and Applications*, A.P. Cowie (ed.) Oxford, Clarendon Press.
- [Cowie et al. 1983] A. Cowie, R. Mackin and I. R. McCaig, *Oxford Dictionary of Current Idiomatic English*, Vol. 2 - English Idioms, Oxford, Oxford University Press.
- [Cruse 1986] D. A. Cruse, *Lexical Semantics*, Cambridge, Cambridge University Press.

- [Daille 1996a] B. Daille, ACABIT : une maquette d'aide à la construction automatique de banques terminologiques monolingues ou bilingues, dans A. Clas, P. Thoiron et H. Béjoint (éds), *Lexicomatique et Dictionnairiques*, FMA, Beyrouth, pp. 123-136.
- [Daille 1996b] B. Daille, Study and Implementation of Combined Techniques for Automatic Extraction of Terminology, in P. Resnik and J. Klavans (eds), *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*, MIT Press, pp. 49-66.
- [Daille and Williams 2001] B. Daille and G. Williams (eds.), *COLLOCATION: Computational Extraction, Analysis and Exploitation*, Workshop in ACL'01, France, Toulouse.
- [Dubreil et Daille 2005a] E. Dubreil et B. Daille, Analyse sémantico-discursive des collocations lexicales en corpus spécialisé : la base « connaissance-s », « Mots, Termes et Contextes », Actes des 7^{èmes} journées scientifiques LTT, 6-10 septembre 2005, Bruxelles, (à paraître).
- [Dubreil 2005b] E. Dubreil, Principe de composition sémantico-discursif des collocations lexicales : de l'arbitraire au motivé ?, Actes des 4^{èmes} Journées de la Linguistique de Corpus (JLC), 15-17 septembre 2005, Lorient, Université de Bretagne Sud.
- [Dubreil 2007a] E. Dubreil, Analyse des mécanismes discursifs d'axiologisation du sens par le contexte, *CJCPrax 2007* (Colloque Jeunes Chercheurs Praxiling) « Analyse du discours et contextes : quelles relations ? », Montpellier, 10-11 mai 2007.
- [Dubreil 2007b] E. Dubreil, Proposition de typologie argumentative des collocations textuelles, Actes des 5^{èmes} Journées de la Linguistique de Corpus (JLC), septembre 2007, Lorient, Université de Bretagne Sud, (à paraître).
- [Firth 1957] J.R. Firth, *Modes and Meaning*, Papers in linguistics 1934-1951, Oxford, Oxford University Press.
- [Fontenelle 1997] T. Fontenelle, Turning a Bilingual Dictionary into a Lexical-Semantic Database, *Lexicographica*, Series Maior 79, Tübingen, Niemeyer.
- [Fontenelle 2003] T. Fontenelle, Collocations et traitement automatique du langage naturel, dans *Les collocations : analyse et traitement*, F. Grossmann et A. Tutin (éds.), Travaux et recherches en linguistique appliquée, Amsterdam, de Werelt, pp. 75-87.
- [Galatanu 2004] O. Galatanu, La sémantique argumentative intégrée et ses enjeux pour l'analyse linguistique du discours, dans *Les actes du Congrès International d'Etudes Françaises : La Rioja, Croisée des Chemins ; Vol. 2*, M.J. Salinero Cascante et Ignacio Inarrea Las Veras (éds.), pp. 213-225.
- [Galatanu 2006] O. Galatanu, La dimension axiologique de la dénomination, dans M. Riegel, C. Schnedecker, P. Swiggers et I. Tamba (éds.), *Aux carrefours du sens. Hommages offerts à Georges Kleiber*, Louvain, Peeters, (à paraître).
- [Gonzalez Rey 2002] I. Gonzalez Rey, *La phraséologie du français*, Toulouse, Presses Universitaires du Mirail.
- [Granger 1998] S. Granger, Prefabricated patterns in advanced EFL writing: collocations and formulae. In AP Cowie (ed.) *Phraseology: Theory, Analysis, and Applications*. Oxford, Clarendon Press, pp. 145-160.

[Gross 1995] G. Gross, Une sémantique nouvelle pour la traduction automatique : les classes d'objets, La Tribune des Industries de la Langue et de l'Information électronique, N° 17-19, Paris, pp. 16-19.

[Gross 1996] G. Gross, Les expressions figées en français. Noms composés et autres locutions, Paris, Ophrys.

[Grossmann et Tutin 2003] F. Grossmann et A. Tutin, Quelques pistes pour le traitement des collocations, Les collocations : analyse et traitement, F. Grossmann et A. Tutin (éds.), Travaux et recherches en linguistique appliquée, Amsterdam, de Werelt, pp. 5-21.

[Grossmann et Tutin 2005] F. Grossmann et A. Tutin, Joie profonde, affreuse tristesse, parfait bonheur : sur la prédicativité des adjectifs intensifiant certains noms d'émotion. Cahiers de lexicologie, N° 86, 2005-1, pp. 1-18.

[Grossmann and Tutin à paraître] F. Grossmann and A. Tutin, Motivation of Lexical Associations in Collocations : the Case of Intensifiers denoting 'Joy', dans Festschrift In Honour of Igor Mel'cuk, L. Wanner (ed.), Amsterdam, John Benjamins, (à paraître).

[Habert et al. 1997] B. Habert, A. Nazarenko et A. Salem, Les linguistiques de corpus, Paris, Armand Colin.

[Halliday and Hasan 1976] M.A.K Halliday and R. Hasan, Cohesion in English, London, Longman.

[Haskel 1971] P.I. Haskel, Collocations as a measure of stylistic variety, in R.A. Wisbey (ed.), The computer in literary and linguistic research, Papers from a Cambridge symposium, Cambridge, Cambridge University Press, pp. 159-168.

[Hausmann 1979] F.-J. Hausmann, Un dictionnaire des collocations est-il possible ?, Travaux de littérature et de linguistique de l'Université de Strasbourg, XVII:1, pp. 187-195.

[Hausmann 1989] F.-J. Hausmann, Le dictionnaire de collocations, dans F.-J. Hausmann et al. (éds.), Wörterbücher : ein internationales Handbuch zur Lexicographie. Dictionnaires, Berlin, De Gruyter, pp. 1010-1019.

[Hausmann 1997] F.-J. Hausmann, Tout est idiomatique dans les langues, dans M. Martins-Baltar (éd.). Langages, La Locution entre langues et usages, Fontenay Saint-Cloud, ENS éditions, pp. 277-290.

[Hausmann 1999] F.-J. Hausmann, Le dictionnaire de collocations – Critères de son organisation, dans N. Greiner et al. (éds.), Texte und Kontexte in Sprachen und Kulturen, Festschrift für Jörn Albrecht, Trier : WVT Wissenschaftlicher Verlag, pp. 121-139.

[Heid 1992] U. Heid, Décrire les collocations. Deux approches lexicographiques et leur application dans un outil informatisé, Terminologie et traduction, Vol. 2(3), pp. 523-544.

[Heid 1994] U. Heid, On Ways Words Work Together - Topics in Lexical Combinatorics, Euralex'94 Proceedings, Amsterdam, pp. 226-257.

[Hoey 1991] M. Hoey, Patterns of Lexis in Text, Oxford, Oxford University Press.

[Howarth 1998] P. Howarth, The Phraseology of Learners' Academic Writing, in A.P. Cowie (ed.) Phraseology: Theory, Analysis, and Applications. Oxford, Clarendon Press, pp. 161-186.

- [Jones and Sinclair 1974] S. Jones and J. Sinclair, English lexical collocations, Cahiers de Lexicologie, N° 24, pp. 15-61.
- [Kahane and Polguère 2001a] S. Kahane and A. Polguère, Formal foundation of lexical functions, in workshop proceedings: COLLOCATION: Computational Extraction, Analysis and Exploitation, 39th Annual Meeting and 10th Conference of the European Chapter of the Association for Computational Linguistics, Toulouse, July 7th 2001, pp. 8-15.
- [Kahane et Polguère 2001b] S. Kahane et A. Polguère, Un langage formel d'encodage des fonctions lexicales et son application à la modélisation des collocations, dans B. Daille et G. Williams (coord.), Journée d'étude de l'ATALA « Collocations ». <http://atala.biomath.jussieu.fr/je/010113/Kahane.resume.pdf>
- [Kennedy 1990] G. Kennedy, Collocations: Where grammar and vocabulary teaching meet in S. Anivan (ed), Language Teaching Methodology for the Nineties, SAEMEO Regional Language Centre, Singapore, pp. 215-229.
- [Kilgarriff 1992] A. Kilgarriff, Polysemy, dissertation, University of Sussex.
- [Kilgarriff 2001] A. Kilgarriff, Comparing Corpora, International Journal of Corpus Linguistics, Vol. 6(1), pp. 97-133.
- [Kjellmer 1984] G. Kjellmer, Some Thoughts on Collocational Distinctiveness, in J. Aarts, W. Meijs (eds.), Corpus Linguistics: recent advances in the use of computer corpora in English Language research, Amsterdam, Rodopi, pp. 163-171.
- [Lakoff et Johnson 1985] G. Lakoff et M. Johnson, Les métaphores dans la vie quotidienne, Paris, Minuit.
- [Laporte et L'homme 1997] I. Laporte et M.-C. L'homme, Recensement et consignation des combinaisons lexicales en langue de spécialité : Exemple d'application dans le domaine de la pharmacologie cardiovasculaire, Terminologies nouvelles Vol. 16, pp. 95-101.
- [Lareau 2002] F. Lareau, La synthèse de textes comme outil de développement et de vérification de modèles linguistiques formels, Mémoire de maîtrise, Département de linguistique et traduction, Université de Montréal.
- [L'homme 2000] M.-C L'homme, Understanding Specialized Lexical Combinations, Terminology, Vol. 6(1), pp. 86-110.
- [L'homme 2003] M.-C L'homme, Les combinaisons lexicales spécialisées (CLS): Description lexicographique et intégration aux banques de terminologie, dans Les collocations : analyse et traitement, F. Grossmann et A. Tutin (éds.), Travaux et recherches en linguistique appliquée, Amsterdam, de Werelt, pp. 89-103.
- [Ludewig 2001] P. Ludewig, LogoTax - un outil exploratoire pour l'étude de collocations en corpus, dans traitement automatique des langues – TAL, Vol. 42(2), Traitement automatique des langues et linguistique de corpus, Paris, Hermès, pp. 623-642.
- [Malinowski 1923] B. Malinowski, The problem of Meaning in Primitive Languages, in C.J. Ogden and I.A. Richards (eds.), The meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism, New York, Harcourt, Brace and World, pp. 296-336.

- [Martin 1997a] J. Martin, Du bon usage des corpus dans la recherche sur le discours spécifique, *Revue ASp* Vol. 15-18, pp. 75-83.
- [Martin 1997b] R. Martin, Sur les facteurs du figement lexical, dans M. Martins-Baltar (éd.), *La locution, entre langue et usages*, Fontenay Saint-Cloud, ENS éditions, pp. 291-307.
- [Mel'cuk 1988] I. Mel'cuk, Paraphrase et lexique dans la théorie linguistique *Sens-Texte* : Vingt ans après. *Cahiers de lexicologie*, N° 52(1), pp. 5-50, et N° 53(2), pp. 5-53.
- [Mel'cuk 1997] I. Mel'cuk, *Vers une linguistique Sens-Texte, Leçon inaugurale*. Paris, Collège de France, 78 pages. <http://www.ling.umontreal.ca/olst/FrEng/melcukColldeFr.pdf>
- [Mel'cuk 1998] I. Mel'cuk, Collocations and Lexical Functions, in A.P. Cowie (ed.), *Phraseology. Theory, Analysis, and Applications*, Oxford, Clarendon Press, pp. 23-53.
- [Mel'cuk 2003] I. Mel'cuk, Collocations : définition, rôle et utilité, dans *Les collocations : analyse et traitement*, F. Grossmann et A. Tutin (éds.), *Travaux et recherches en linguistique appliquée*, Amsterdam, de Werelt, pp. 23-31.
- [Mel'cuk et al. 1984, 1988, 1992, 1999] I. Mel'cuk, *Dictionnaire explicatif et combinatoire du français contemporain, Recherches lexico-sémantiques*, Vols. I, II, III, IV, Montréal, Presses de l'Université de Montréal.
- [Meyer and Mackintosh 1996] I. Meyer, K. Mackintosh, The corpus from a terminographer's viewpoint, *International Journal of Corpus Linguistics*, Vol. 2, pp. 257-285.
- [Meynard 2000] I. Meynard, *Internet, Répertoire bilingue de combinaisons lexicales spécialisées français-anglais*, Montréal, Linguatech.
- [Mortchev-Bouveret 1998] M. Mortchev-Bouveret, Approche de la dénomination en langue spécialisée, *Meta* Vol. XLIII(3), pp. 393-410.
- [Mortchev-Bouveret 2004] M. Mortchev-Bouveret, Etude comparative de la phraséologie dans six dictionnaires de langue française, in A. Häcki Buhofer and H. Burger (eds.), *Phraseology in Motion, Proceedings zu Europhras Basel 2004*, Baltmannsweiler, Schneider Verlag, pp. 289-298.
- [Partington 1998] A. Partington, *Patterns and Meanings*, Amsterdam, John Benjamins.
- [Pearson 1998] J. Pearson, *Terms in Context*; Amsterdam, John Benjamins.
- [Pecman 2004a] M. Pecman, Exploitation de la phraséologie scientifique pour les besoins de l'apprentissage des langues, *Journée d'étude de l'ATALA « Traitement Automatique des Langues et Apprentissage des Langues »*, édité par Grenoble, LIDILEM, Université de Grenoble Stendhal, pp. 145-154.
- [Pecman 2004b] M. Pecman, L'enjeu de la classification en phraséologie, *EUOPHRAS 2004*, 26-29 août 2004, Université de Bâle, Suisse, Baltmannsweiler, Schneider Hohengehren Verlag, pp. 127-146.
- [Pecman 2005] M. Pecman, Les apports possibles de la phraséologie à la didactique des langues étrangères, *Apprentissage des Langues et Systèmes d'Information et de Communication (ALSIC)*, N° 8(1), pp. 109-122.
- [Phillips 1985] M. Phillips, *Aspects of Text Structure: An investigation of the lexical Organisation of Text*, Amsterdam, North Holland.

- [Polguère 2000a] A. Polguère, Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French, Proceedings of EURALEX'2000, Stuttgart, pp. 517-527.
- [Polguère 2000b] A. Polguère, Une base de données lexicale du français et ses applications possibles en français, Revue de Linguistique et de Didactique des Langues, N° 21, pp. 75-97.
- [Polguère 2003] A. Polguère, Étiquetage sémantique des lexies dans la base de données DiCo, Revue TAL, Vol. 44(2), pp. 39-68.
- [Polguère 2005] A. Polguère, Typologie des entités lexicales d'une base de données explicative et combinatoire, Journée d'étude de l'ATALA « Interface lexique-grammaire et lexiques syntaxiques et sémantiques », École nationale supérieure des télécommunications (ENST), Paris.
- [Rey et Chantreau 1989] A. Rey et S. Chantreau, Dictionnaire des expressions et locutions, Le Robert (éds.), Collection Les usuels.
- [Sinclair 1987] J. Sinclair, Collocation: A progress report. In R. Steele and T. Threadgold, (eds.), Language Topics: Essays in Honour of Michael Halliday, Vol. II, Amsterdam, John Benjamins, pp. 319-331.
- [Sinclair 1991] J. Sinclair, Corpus, concordance, collocation, Oxford, Oxford University Press.
- [Sinclair 1996] J. Sinclair, Preliminary recommendations on Corpus Typology, Rapport technique, CEE, EAGLES (Expert Advisory Group on Language Engineering Standards).
- [Sinclair et al. 1970, 2002] J. Sinclair, S. Jones and R. Daley, English Lexical Studies, Report to Office for Scientific and Technical Information (OSTI) on Project C/LP/08, University of Birmingham.
- [Smadja 1993] F. Smadja, Retrieving Collocations from Text: Xtract, Computational Linguistics, Vol. 19(1), pp. 143-177.
- [Smadja and McKeown 1991] F. Smadja and K.R. McKeown, Using collocations for language generation, Vol. 7(4), pp. 229-239.
- [Steinlin et al. 2004] J. Steinlin, S. Kahane, A. Polguère et A. El Ghali, De l'article lexicographique à la modélisation objet du dictionnaire et des liens lexicaux, Actes de EURALEX'2004, Lorient, pp. 177-186.
- [Stubbs 1995] M. Stubbs, Collocations and semantic profiles, Functions of Language, Vol. 2(1), pp. 23-55.
- [Stubbs 1996] M. Stubbs, Text and Corpus Analysis, Oxford, Blackwell.
- [Swales 1990] J. Swales, Genre analysis, Cambridge, Cambridge University Press.
- [Telija 1992] V.N. Telija, Lexicographic description of words and collocations: feature-functional model, Euralex '90, Barcelona, pp. 315-320.
- [Telija 1994] V.N. Telija, Lexical collocation: denominative and cognitive aspects, Euralex '94, Amsterdam, pp. 368-377.

[Telija et al. 1998] V.N. Telija, N.G. Bragina, E.O. Oparina and I.I. Sandomirskaja, Phraseology as a language of culture: its role in the representation of a cultural mentality, in *Phraseology: Theory, Analysis, and Applications*, A.P. Cowie (ed.), Oxford, pp. 55-75.

[Tognini-Bonelli 2001] E. Tognini-Bonelli, *Corpus Linguistics at Work*, Amsterdam, John Benjamins.

[Tutin 1996] A. Tutin, The formalization of collocations for natural language processing, *Proceedings of COMPLEX'96*, Hungary, Budapest, pp. 243-256.

[Tutin 1997] A. Tutin, Coder les collocations dans un lexique formel de TALN, *Revue française de linguistique appliquée*, Vol. II, pp. 43-58.

[Tutin 2004] A. Tutin, Pour une modélisation dynamique des collocations dans les textes, *Actes de EURALEX'2004*, France, Lorient, pp. 207-219.

[Tutin 2005] A. Tutin, Le dictionnaire de collocations est-il indispensable ? *Revue Française de Linguistique Appliquée. Dictionnaires : nouvelles approches, nouveaux modèles*. Vol. X(2), pp. 31-48.

[Tutin et Grossmann 2002] A. Tutin et F. Grossmann, Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif, *Revue Française de Linguistique Appliquée, Lexique : recherches actuelles*, Vol. VII, pp. 7-25.

[Verlinde et al. 2003] S. Verlinde, T. Selva et J. Binon, Les collocations dans les dictionnaires d'apprentissage : repérage, présentation et accès, dans *Les collocations : analyse et traitement*, F. Grossmann et A. Tutin (éds.), *Travaux et recherches en linguistique appliquée*, Amsterdam, de Werelt, pp. 105-115.

[Wanner 1996] L. Wanner, *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam, John Benjamins.

[Williams 1999a] G. Williams, Looking in before looking out: Internal selection criteria in a corpus of plant biology, *Papers in computational Lexicography, Proceedings of Complex'99*, Hungary, Budapest, pp. 195-204.

[Williams 1999b] G. Williams, Les réseaux collocationnels dans la construction et l'exploitation d'un corpus dans le cadre d'une communauté de discours scientifique, thèse de doctorat soutenue le 15/12/99, à l'Université de Nantes.

[Williams 2001a] G. Williams, Les réseaux collocationnels dans la construction et l'exploitation d'un corpus dans le cadre d'une communauté de discours scientifique, Lille, Presses Universitaires de Septentrion.

[Williams 2001b] G. Williams, Sur les caractéristiques de la collocation, *TALN*, Tours, pp. 9-16.

[Williams 2002] G. Williams, In search of representativity in specialized corpora: categorisation through collocation, *International journal of Corpus Linguistics*, Vol. 7(1), pp. 43-64.

[Williams 2003] G. Williams, Les collocations et l'école contextualiste britannique, dans *Les collocations : analyse et traitement*, F. Grossmann et A. Tutin (éds.), *Travaux et recherches en linguistique appliquée*, Amsterdam, de Werelt, pp. 33-44.

[Williams 2005] G. Williams (éd.), *La linguistique de corpus*, Rennes, Presses Universitaires Rennes.

[Wouden 1997] T. van der Wouden, *Negative contexts: Collocation, polarity and multiple negation*, London, Routledge Studies in Germanic Linguistics.