



HAL
open science

Collocations régulières et irrégulières: esquisse de typologie du phénomène collocatif.

Agnès Tutin, Francis Grossmann

► **To cite this version:**

Agnès Tutin, Francis Grossmann. Collocations régulières et irrégulières: esquisse de typologie du phénomène collocatif.. *Revue Française de Linguistique Appliquée*, 2002, *Lexique: Recherches actuelles*, 7 (1), pp.7-25. hal-01336278

HAL Id: hal-01336278

<https://hal.science/hal-01336278>

Submitted on 6 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

COLLOCATIONS RÉGULIÈRES ET IRRÉGULIÈRES : ESQUISSE DE TYPOLOGIE DU PHÉNOMÈNE COLLOCATIF

Agnès Tutin, Francis Grossmann

Pub. linguistiques | « *Revue française de linguistique appliquée* »

2002/1 Vol. VII | pages 7 à 25

ISSN 1386-1204

Article disponible en ligne à l'adresse :

<https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2002-1-page-7.htm>

Pour citer cet article :

Agnès Tutin, Francis Grossmann « Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif », *Revue française de linguistique appliquée* 2002/1 (Vol. VII), p. 7-25.

Distribution électronique Cairn.info pour Pub. linguistiques.

© Pub. linguistiques. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif

Agnès Tutin et Francis Grossmann
Université de Grenoble

Résumé : Les collocations constituent une thématique essentielle dans les phénomènes de combinatoire lexicale, mais restent souvent définies de façon floue. Dans cet article, notre premier objectif est de proposer une définition plus claire du phénomène, en nous appuyant sur des paramètres rendant compte du degré de figement. Nous présentons ensuite quelques caractéristiques syntaxiques, puis quelques mécanismes sémantiques productifs dans la genèse des collocations imagées. Nous distinguons enfin les collocations de notions voisines.

Abstract: Collocations are an essential phenomenon in lexical combinatorics but often are fuzzily defined. In this paper, we aim at defining clearly this concept and propose parameters accounting for the degree of idiomaticity. We present then some syntactic characteristics and some semantic processes yielding colourful collocations. We finally distinguish collocations from close concepts.

1. Introduction

Dans la langue, certains mots présentent des affinités et tendent à apparaître ensemble : *ferme intention*, *argument de poids*, *brouillard à couper au couteau*, *rendre visite* ou *prêter attention*. Ces expressions sont souvent appelées collocations.

De telles associations s'opposent intuitivement à des expressions complètement figées comme *pomme de terre* ou *nid d'ange*, pour lesquelles le sens du tout est difficilement prédictible à partir du sens des parties : l'interprétation sémantique ne se fait pas de façon compositionnelle. Ces expressions s'opposent également à des associations complètement libres du type *argument intéressant*, *envie de chocolat*, dont le sens est non seulement analysable mais aussi prédictible. On caractérise souvent les collocations par le fait que le sens est transparent en réception (il se « devine »), alors que, pour un locuteur non natif, il est difficile de produire les lexèmes adéquats.

A la frontière entre le préconstruit et le libre, les collocations constituent une problématique réelle en linguistique et en linguistique appliquée, distincte de celles des expressions figées et largement étudiée depuis quelques années comme en témoignent les ateliers et les conférences organisés autour de cette notion¹ même si les contours du

¹ Par exemple Journée d'étude de l'ATALA : « La collocation » à Paris en janvier 2001, Journées d'étude à Grenoble « Modélisation des collocations » en septembre 2001, Atelier « Collocation:

phénomène semblent parfois mal définis et si le terme reçoit des interprétations divergentes selon les auteurs et les écoles.

Dans cet article, nous nous fixons comme première tâche, sans prétendre à l'exhaustivité, de faire le point sur les différentes approches de la notion de collocation et nous en proposons une définition inspirée du modèle mel'cukien. Nous essayons ensuite de caractériser ces associations sur le plan syntaxique, puis nous examinons les mécanismes sémantiques dont elles sont souvent issues. Nous distinguons ensuite les collocations de notions proches comme les expressions figées et les restrictions de sélection sémantique. L'étude part du postulat selon lequel, s'il existe indéniablement un continuum dans le degré de figement sémantique des collocations, il apparaît néanmoins souhaitable de proposer une typologie plus fine à l'intérieur de cette classe permettant des traitements linguistiques adaptés à chaque type.

2. La collocation : un phénomène aux contours flous

2.1. Deux approches de la notion de collocation

Les collocations au sens premier sont *des mots qui tendent à apparaître ensemble*. Cette notion large du concept 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*)².

Cette conception large introduite par Firth dans les années 50 est reprise chez des fonctionnalistes anglais contemporains (Halliday, Sinclair)³. Pour Halliday & Hasan (1976), les collocations ont comme rôle fonctionnel de contribuer à la cohésion du texte.

laugh ... joke, blade ... sharp, ill... doctor [...] The cohesive 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 another. 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 & Hasan, 1976, 285-286)

[rire ... blague, lame ... aiguisé, malade ... médecin [...] L'effet cohésif de telles paires ne dépend pas tant d'une relation systématique que de leur tendance à partager le même environnement, à apparaître en COLLOCATION l'un avec l'autre. En général, n'importe quelle paire d'éléments lexicaux ayant les mêmes patrons de collocation – c'est-à-dire ayant tendance à apparaître dans les mêmes contextes – généreront une force cohésive s'ils apparaissent dans des phrases adjacentes.]

John Sinclair (1991, 70) n'adopte pas une autre définition de la notion de collocation, essentiellement définie par des critères de type statistique. Cette approche large est encore

Computational Extraction, Analysis and Exploitation » de la conférence ACL 2001, à Toulouse, Atelier « Lexicographic applications of computational approaches to collocations: Restricted collocations in dictionaries » de la conférence EURALEX, août 2002, Colloque « Computational Approaches to Collocations », Vienne, juillet 2002.

² Par exemple, Benson, Benson et Ilson (1990) incluent dans les collocations grammaticales les patrons de sous-catégorisation.

³ Cf. Williams (à paraître).

partagée par certains chercheurs en TAL (par exemple, Smadja & McKeown, 1991) ou par des chercheurs se situant dans la lignée des contextualistes anglais (Williams, 1999). Néanmoins, depuis quelques années, dans la communauté des lexicologues et des lexicographes, il semble que la conception étroite – la collocation est une cooccurrence lexicale privilégiée de deux éléments linguistiques entretenant une relation syntaxique – soit largement privilégiée. Cette définition prend sa source dans les études théoriques en lexicologie (Cruse, 1986) et en lexicographie théorique (Mel'cuk, 1984 ; Haussman, 1989). C'est celle que nous adopterons dans la suite de notre présentation.

2.2. Quels critères pour définir les collocations ?

L'idée qu'il existe un stade intermédiaire de figement entre expressions mémorisées et combinaisons libres n'est pas nouvelle. Elle apparaît déjà chez Bally dans son *Traité de stylistique française* (1909) sous le terme de *séries phraséologiques*.

Entre les extrêmes (groupements passagers et unités indécomposables) se placent des groupes intermédiaires appelés *séries phraséologiques* (p.ex. les séries d'intensité ou les périphrases verbales) (p. 66).

Il y a série ou groupement usuel lorsque les éléments du groupe conservent leur autonomie, tout en laissant voir une affinité évidente qui les rapproche, de sorte que l'ensemble présente des contours arrêtés et donne l'impression du « déjà vu » (p. 70).

Les exemples cités par Bally, *gravement malade* et *grièvement blessé, désirer ardemment* et *aimer éperdument*, où les adverbes n'apparaissent pas facilement interchangeables dans les séries, sont souvent encore présentés comme des prototypes de collocations. Bally remarque déjà que les séries phraséologiques sont particulièrement productives dans les séries d'intensité (cf. la fonction **Magn** dans la lexicologie explicative et combinatoire de Mel'cuk et ses collègues (1995)) et dans le cas des périphrases verbales (cf. la fonction **Oper** dans le même modèle). Les collocations sont évidemment productives dans les clichés, les qualificatifs galvaudés de la « mauvaise » littérature, raillés par Flaubert dans le *Dictionnaires des idées reçues*⁴. Les métaphores et les comparaisons y abondent souvent (*faim de loup, s'ennuyer à mourir, dormir comme une souche, déluge de paroles*) comme on le verra en 4. à côté d'expressions moins marquées stylistiquement (*avoir faim, prendre peur, essuyer un échec*).

La notion de collocation, définie provisoirement comme cooccurrence lexicale privilégiée de deux éléments linguistiques entretenant une relation syntaxique, reçoit diverses appellations dans la littérature linguistique. Mel'cuk et ses collègues (1998 ; 1995) préfèrent le terme de *semi-phrasème* qui, à certains égards, paraît plus explicite que le terme de *collocation* qui englobe aussi l'acception décrite plus haut : *semi* suggère en effet que le figement n'est pas complet, et le terme de *phrasème*, bien que peu usité, est monosémique. Le terme de *collocation* étant plus répandu, nous le préférons malgré ses inconvénients.

⁴ Par exemple, aux articles ACCIDENT et CHALEUR.

ACCIDENT : Toujours déplorable ou fâcheux (comme si on devait jamais trouver un malheur une chose réjouissante...).

CHALEUR : Toujours insupportable. Ne pas boire quand il fait chaud.

Nous avons abordé les collocations à travers quelques exemples prototypiques. Mais comment définir formellement la notion ?

Si l'on se base sur les textes fondateurs définissant la notion de collocation (Cruse, 1986 ; Mel'cuk, 1998 ; Hausmann, 1989), plusieurs critères peuvent être mis au jour.

- **L'aspect arbitraire (la non prédictibilité)** de l'association lexicale : la collocation n'est pas prévisible (Hausmann : « [la] transparence n'empêche nullement la collocation d'être imprédictible »). 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*). Néanmoins, si l'arbitraire caractérise souvent les réalisations lexicales, on relève que les collocations suivent des patrons syntaxiques précis. Par exemple les collocations construites autour d'un nom incluront des adjectifs ou des verbes, mais probablement pas des adverbes ou des conjonctions. Par ailleurs, sur le plan sémantique, comme on le verra en 4., l'association est souvent favorisée par certaines relations (comme l'intensité).
- **La transparence et le non-figement sémantiques de la collocation.** Cruse et Hausmann soulignent que les collocations peuvent être interprétées par le locuteur, même si leur forme n'est pas prévisible⁵. Leur sens se déduit. Les exemples de Hausmann (*célibataire endurci*, *feuilleter un livre*) seront facilement compris par un locuteur non natif du français qui serait toutefois tout à fait incapable de les produire s'il ne les avait mémorisés comme tels. Chez Mel'cuk, les faits linguistiques décrits sont plus complexes. Dans certains cas de figure, le sens n'est pas réellement transparent. C'est le cas d'expressions comme *peur bleue* ou *colère noire*, dont l'adjectif n'est pas immédiatement interprétable. La « transparence » ne caractérise donc peut-être pas l'ensemble de la classe.
- **Le caractère binaire de la collocation.** Pour Hausmann comme pour Mel'cuk, la collocation est composée essentiellement de deux mots ou de deux lexies. Cette bipartition n'est pas aussi marquée chez Cruse. Même si ce caractère duel est vérifié dans la plupart des collocations, il nous paraît plus pertinent de parler d'association de constituants (plutôt que de lexies) pour englober des exemples comme les suivants :

fort comme un turc
un bruit à crever les tympans

pour lesquels l'élément qualifiant se réalise par un syntagme : il y a ici association d'un mot (par exemple, *fort*, *bruit*) avec un syntagme (*comme un turc*, *à crever les tympans*). Il paraît en revanche justifié de considérer les collocations comme constituées de deux parties dont le statut est d'ailleurs inégal. Il arrive cependant que des expressions englobent davantage que deux termes : *essuyer un échec cuisant*, *avoir une peur bleue*. Dans la plupart des cas, il est possible de considérer qu'on a simplement affaire à deux

⁵ « La collocation se distingue d'autre part des locutions [...] par son non-figement et sa transparence. » (Hausmann 1989). « 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).

collocations distinctes qui ont été fusionnées : *essuyer un échec et échec cuisant, avoir peur et peur bleue*, même si la fusion est parfois bloquée⁶.

- La **dissymétrie des composants de la collocation (un élément conserve son sens habituel)**. Mel'cuk et Hausmann considèrent tous deux que les composants de la collocation n'ont pas le même statut (Hausmann : « Dans la collocation, le statut des deux partenaires combinés n'est pas égal »). Pour Hausmann, un élément est autonome (la **base**), il a son sens habituel. L'autre (le **collocatif**) dépend du premier. Pour Mel'cuk, le rapport est aussi nettement dissymétrique, la base conservant son sens habituel.
- La **notion de cooccurrence restreinte, de sélection lexicale**. Mel'cuk et Hausmann s'intéressent surtout au phénomène collocatif dans le cadre de la production, l'encodage. Hausmann suggère dans sa formulation que la base impose la sélection du collocatif. Cette perspective apparaît encore plus claire chez Mel'cuk pour qui la collocation doit être envisagée dans le cadre de la production. Pour lui, il y a collocation, lorsque, voulant « produire » une suite de deux expressions, le choix d'une des expressions n'est pas libre, mais imposé par l'autre. Pour lexicaliser le sens d'« intense » en cooccurrence avec *peur*, le locuteur choisira *bleue*. La cooccurrence n'est pas libre, mais restreinte⁷.

Les cinq paramètres mis en évidence permettent de caractériser des collocations prototypiques comme *célibataire endurci* ou *pluie torrentielle*, mais n'engloberont pas des expressions comme *peur bleue, il pleut à verse, l'âne braie* ... parce que le critère d'imprédictibilité pourra paraître contestable avec *l'âne braie*, alors qu'inversement la transparence ne caractérise pas clairement *peur bleue* ou *il pleut à verse*. Une solution possible serait d'adopter une définition à l'aide d'un prototype, un peu à la façon de Robert Martin (1997) pour les locutions⁸. Mais cette solution ne nous semble pas souhaitable pour deux raisons. D'une part, la notion nous paraît trop floue pour qu'un prototype puisse être dégagé. Le prototype ajouterait encore du flou au flou. D'autre part, le choix du prototype ne fait que reporter l'étude systématique des critères linguistiques pourtant indispensable pour des traitements formels en lexicographie et en TAL.

Nous préférons définir les collocations à partir de propriétés : le caractère binaire, la dissymétrie et la notion de sélection lexicale seront des conditions nécessaires. La transparence et l'arbitraire seront des paramètres plaçant la collocation sur une échelle de figement.

⁶ La fusion des collocations n'est cependant pas toujours possible : *prendre peur + peur bleue* => **prendre une peur bleue*.

⁷ Cette perspective de production, bien que compréhensible intuitivement, ne va pas sans poser de problèmes. Le premier est celui de la procédure de « production » qui, interne au sujet parlant, n'est pas observable. Il n'y a pas de trace linguistique tangible de ce processus. Le second est le traitement de la difficile limite entre cooccurrence restreinte et cooccurrence libre. Si *peur bleue* est plus idiomatique (et donc plus « restreint »), *grande peur* est aussi possible, mais dans des contextes probablement différents.

⁸ Cette solution avait d'ailleurs été suggérée par Pierre Corbin pour la définition des collocations (communication personnelle).

Reformulant à notre façon la définition proposée par Mel'cuk (1998)⁹, nous définirons les collocations de la façon suivante :

Une **collocation** est l'association d'une **lexie (mot simple ou phrasème)**¹⁰ **L** 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 syntaxique telle que :

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

Sur le plan sémantique, plusieurs cas de figure se présentent :

1. Le sens du collocatif en cooccurrence avec L est différent du sens qu'il a en dehors de cette association (*peur bleue, colère noire*). Dans ce cas (**collocations opaques**), l'association paraît arbitraire et non transparente sur le plan sémantique.

2. Le collocatif peut avoir un sens interprétable (la collocation est « transparente ») mais apparaître difficilement prédictible (*faim de loup, brouillard à couper au couteau*,

⁹ The signified of a collocation includes 'intact' the signified of the one of its two constituent lexemes – say, of A [...]; A is freely chosen by the speaker strictly because of its signified. But the other component of its signified, i.e. 'C' is 'problematic': it is expressed by B, which is chosen CONTINGENT on A (this makes the signifier of a collocation to be not unrestrictedly constructed). [...]

A Collocation **AB** of L is a semantic phraseme of L such that its signified 'X' is constructed out of the signified of the one of its two constituent lexemes - say, of A - and a signified 'C' ['X' = 'AHB'] such that the lexeme **B** expresses 'C' contingent on **A**.

The formulation « **B** expresses 'C' contingent on **A** » covers four major cases, which correspond to the following four major types of collocations :

1. **either** 'C' ('B', i.e. **B** does not have (in the dictionary) the corresponded signified;
 - and** [a. 'C' is empty, that is, the lexeme **B** is, so to speak, a semi-auxiliary selected by **A** to support it in a particular syntactic configuration;
 - or** b. 'C' is not empty but the lexeme **B** expresses 'C' only in combination with **A** (or with a few other similar lexemes)];
2. **or** 'C' = 'B', i.e., **B** has (in the dictionary) the corresponding signified;
 - and** [a. 'B' cannot be expressed with **A** by any otherwise possible synonym of **B**;
 - or** b. 'B' includes (an important part of) the signified 'A', that is, it is utterly specific, and this **B** is 'bound' by **A**].

Examples (lexeme **A** is in small caps)

Case **1a** : collocations with support (= 'light') verbs, such as [*to*] *do a FAVOR*, [*to*] *give a LOOK*, [*to*] *take a STEP*, [*to*] *launch an APPEAL*, [*to*] *lay SIEGE* [*to* N].

Case **1b** : collocations such as *black COFFEE*, *French WINDOW*, Fr. *BIÈRE bien frappée* < *battue >
...

Case **2a** : collocations with intensifiers, such as *strong* < *powerful > *COFFEE*, *heavy* < *weighty > *SMOKER*, *deeply* < *profoundly > *MOVED*, [*to*] *ILLUSTRATE vividly*.

Case **2b** : collocations such as *The HORSE neighs*, *aquiline NOSE*, *rancid BUTTER* or *artesian WELL*.

¹⁰ C'est-à-dire une expression complètement figée dans la terminologie mel'cukienne.

grièvement blessé, avoir faim). Le collocatif n'a pas de statut lexical (c'est un syntagme comme *de loup* ou *à couper au couteau*) ou bien a un sens décodable en cooccurrence avec la base (*grièvement, avoir*) On parlera de **collocations transparentes**.

3. Le collocatif inclut le sens de la base ou a un sens très générique (*nez aquilin, l'âne brait, grande tristesse*). L'association est motivée et transparente. Le collocatif est à base unique (on définirait le collocatif à l'aide de la base, comme dans le cas de *aquilin*) ou bien au contraire porte sur un paradigme qui semble pouvoir être défini à l'aide de traits sémantiques. On parlera de **collocations régulières**.

Les collocations opaques sont proches des expressions perçues comme figées, mais le sens de la base reste interprétable. Les collocatifs sont généralement peu productifs (ils sont souvent uniques, en tout cas, avec ce sens-là) ou donnent lieu à des séries limitées. Par exemple, *bleu* dans le sens d'«intense» apparaît également en cooccurrence avec *trouille* ou *frousse*, bien que de façon moins productive¹¹.

Les collocations transparentes constituent le cas prototypique. La collocation se comprend, par exemple pour un locuteur non natif, mais ne peut pas être facilement produite. Le décodage est rendu possible par un sens du collocatif proche du sens habituel ou une périphrase comportant une comparaison ou une métaphore facilement interprétable.

Les collocations régulières sont évidemment proches des expressions libres. Dans le premier cas, le collocatif est à base unique, mais le collocatif contient le sens de la base dans sa définition et n'a pas d'autre sens. Par exemple, pour reprendre un exemple examiné dans Alonso Ramos & Mantha 1997, l'adjectif *bissextile* ne porte guère que sur le nom *année*, même si l'on rencontre des variantes¹². Par définition, *bissextile* impose ce type de restriction lexicale. Le cas de *aquilin* est un peu différent. Littéralement, il signifie 'de l'aigle', mais s'est spécialisé comme qualificatif pour le mot *nez*. Dans le second cas, à l'inverse, le collocatif porte sur un large paradigme de mots, qui semblent pouvoir être définis à l'aide de caractéristiques sémantiques. C'est par exemple, le cas de *grand* qui semble être l'intensif standard pour les noms d'émotion ou *ressentir* qui est le verbe opérateur par défaut de la même catégorie. Des contre-exemples sont cependant souvent fréquents (cf. discussion plus loin en 5.2).

La typologie esquissée sera sûrement mise à mal par quelques cas-limites mais cette catégorisation nous paraît toutefois préférable à la solution du continuum qui permet difficilement d'adopter des solutions concrètes pour un traitement formel.

3. Caractéristiques syntaxiques des collocations

En tant que cooccurrences lexicales privilégiées, les collocations – dans le sens étroit que nous nous avons donné à ce terme – présentent des particularités syntaxiques, qui varient

¹¹ Dans le corpus *Frantext* de 1950 à 2000, soit plus de 37 M de mots, on relève le nombre d'occurrences suivant : *peur bleue* (26), *trouille bleue* (2), *frousse bleue*(2), *pétoche bleue* (0), *effroi bleu* (0), *terreur bleue* (1).

¹² Par exemple, nous avons trouvé sur la Toile : *l'an 2000 sera bissextile...Pour être bissextile, une année ...*

selon le statut des associations. Par exemple, les collocations « régulières » apparaissent, du point de vue syntaxique, comme les moins contraintes.

3.1. Constituants mis en jeu dans les collocations

Les collocations mettent en jeu des constituants spécifiques puisque les éléments constitutifs forment généralement un syntagme. Les associations de parties du discours ne sont pas arbitraires, par exemple un adverbe et un substantif, mais permettent de former des syntagmes nominaux (adjectif + substantif) ou des syntagmes verbaux (verbe + substantif). Hausmann (1989, 1010) décrit la classe des collocations par un inventaire des paires de parties du discours que les collocations peuvent constituer :

On appellera collocation la combinaison caractéristique de deux mots dans une des structures suivantes :

- a) substantif + adjectif (épithète)
- b) substantif + verbe
- c) verbe + substantif (objet)
- d) verbe + adverbe
- e) adjectif + adverbe
- f) substantif + (prép.) + substantif

La liste d'associations fournie par Hausmann apparaît cependant partielle si l'on consulte quelques exemples du *Dictionnaire Explicatif et Combinatoire* (Mel'cuk et al., 1984, 1988, 1992, 2000), qui recense de nombreuses collocations. Autour des bases nominales, on peut ainsi avoir, non seulement des épithètes adjectivales, mais aussi d'autres combinaisons comme les suivantes (la liste n'est pas exhaustive):

<i>Patron syntaxique</i>	<i>Exemples</i>
prép + substantif	de joie
adj + prép + substantif (le substantif est complément de l'adjectif)	ivre de colère
V + nom (le nom est un complément « second »)	Avoir quelqu'un en visite

Tableau 1. *Quelques combinaisons de catégories syntaxiques formant des collocations*

Il faudrait donc élargir la liste de combinaisons de Hausmann en incluant celles-ci..

Par ailleurs, la notion de partie de discours proposée ici, en particulier pour le collocatif, paraît peut-être trop limitative. Par exemple, les collocatifs qui portent sur les adjectifs ne sont pas seulement des adverbes au sens strict, mais des constituants ayant un fonctionnement adverbial (le collocatif à fonctionnement adverbial apparaît en italique dans le Tableau 2 ci-dessous).

<i>Nature du constituant</i>	<i>Exemple</i>
Syntagme prépositionnel	Saoul <i>comme</i> une barrique
Adjectif	Ivre <i>mort</i>
Infinitive	Bête à manger <i>du foin</i>

Adverbe	Grièvement blessé
---------	-------------------

Tableau 2. Quelques exemples de collocatifs adverbiaux associés aux adjectifs

Les collocations mettent donc en jeu des constituants variés pouvant former des syntagmes.

Un repérage systématique des associations syntaxiques dans le cadre des collocations n’a, à notre connaissance, jamais été effectué et serait utile, tant pour les applications de TAL (en particulier pour le repérage automatique des collocations dans les textes) qu’en lexicographie.

3.2. Statut syntagmatique de la collocation

Si les collocations mettent en jeu prototypiquement des constituants contigus formant un syntagme, cela n’est pas toujours le cas. D’une part, la contiguïté n’est pas systématique (Heid, 1994). Certaines associations **substantif + épithète** comme *steak bleu* peuvent apparaître séparées (*mon steak est trop bleu*) : les éléments de la collocation *peuvent* donc former un syntagme, mais ne forment pas toujours un syntagme. Par ailleurs, dans les collocations verbales, certaines alternances syntaxiques sont possibles (passivation, relativation, etc.), bien que la souplesse soit souvent moins grande qu’avec les combinaisons libres.

Par ailleurs, il arrive que la combinaison du collocatif et de la base, même lorsqu’ils sont contigus, ne forme pas un constituant au sens traditionnel. C’est par exemple le cas de l’association **substantif (sujet) + verbe** dans *le problème réside (dans cet aspect)*, puisque dans les modèles à base de constituants, le sujet et le verbe ne forment pas une unité. De plus, comme cela a été souvent souligné (Heid, 1994 ; Iordansjaka, Kim & Polguère, 1996), la base de la collocation n’est pas systématiquement la tête du syntagme, comme on peut le constater dans les exemples suivants :

Exemple	Base de la collocation	Tête du syntagme
Célibataire endurci	Célibataire	Célibataire
De dépit	Dépit	De
Rendre visite	Visite	Rendre
Vert de peur	Peur	Vert

Tableau 3. Bases, collocatifs et têtes des syntagmes

Il ressort de ces remarques que les modèles à base de constituants ne sont peut-être pas les plus qualifiés pour rendre compte de ces associations. Les modèles syntaxiques de dépendance comme celui de la théorie Sens-Texte (Mel’cuk, 1988) qui relie des mots plutôt que des constituants sont probablement mieux adaptés à la description du phénomène.

3.3. Restrictions syntaxiques des collocations

Les collocations étant partiellement figées, elles présentent des restrictions syntaxiques bien connues par rapport aux combinaisons libres.

Tout d’abord, elles se construisent parfois de façon atypique (souvent archaïque) par rapport aux règles standard de la langue, comme en témoigne par exemple l’absence

fréquente de déterminant dans les constructions à verbe support : *avoir faim, avoir soif, rendre visite, avoir rendez-vous*, qui n'est cependant pas du tout systématique (*essuyer un échec, perpétrer un délit*).

Par ailleurs, elles présentent souvent moins de flexibilité dans leur comportement syntaxique au niveau de la distribution syntaxique. Par exemple, les collocatifs épithètes ayant un fonctionnement prédicatif ne peuvent pas tous apparaître comme attributs, alors que dans les combinaisons libres, ce contexte est pratiquement toujours possible avec les adjectifs ayant cette fonction sémantique (Gross, 1988). On pourra ainsi opposer *peur bleue* à *peur immense*.

Jean a eu une peur bleue.
Jean a ressenti une peur immense.
*La peur de Jean a été bleue.
La peur de Jean a été immense.

Mais là aussi, ce comportement n'est pas la règle. La collocation *steak bleu* présente ainsi moins de figement syntaxique que *peur bleue*.

Léo aime le steak quand il est bleu.

Enfin, les collocations verbales peuvent connaître un certain nombre d'alternances syntaxiques. On relève ici aussi des particularités par rapport aux combinaisons libres. Par exemple, les constructions à verbe support, largement décrites dans le cadre du lexique grammaire (Giry-Schneider, 1987 ; Gross, 1981 ; Vivès, 1993) présentent des propriétés bien connues. On peut ainsi comparer les expressions *mener une attaque* (où *mener* est verbe support) et *raconter une attaque* (où *raconter* est un verbe plein) :

- Les contraintes sur les déterminants¹³ :

Luc raconte (une+cette+ton) attaque contre la citadelle.
Luc mène (une+*cette+*ton) attaque contre la citadelle.

- La double analyse du groupe NPred Prép N :

C'est une attaque contre la citadelle que Luc raconte.
* C'est une attaque que Luc raconte contre la citadelle.
C'est une attaque contre la citadelle que Luc mène.
C'est une attaque que Luc mène contre la citadelle.

- L'effacement du verbe support dans certains contextes.

L'attaque contre la citadelle menée par Luc = L'attaque de Luc contre la citadelle.
L'attaque contre la citadelle racontée par Luc ≠ L'attaque de Luc contre la citadelle.

Ainsi, bien que les collocations présentent souvent des spécificités sur le plan syntaxique par rapport aux combinaisons libres et des degrés de figement divers, le figement syntaxique ne peut pas être considéré en tant que tel comme une de leurs propriétés définitives.

4. Régularités et irrégularités sémantiques dans les collocations imagées

¹³ Les exemples sont tirés de Vivès, 1993.

Dans de nombreuses collocations, l'arbitraire n'est pas total. D'une part, l'association lexicale est souvent favorisée par certaines relations sémantiques entre collocatif et base, comme les séries d'intensité de Bally¹⁴. D'autre part, certains procédés stylistiques comme la comparaison, la métaphore et la métonymie semblent largement utilisés dans ces associations qui, si elles comportent des valeurs arbitraires, ne se forment pas toujours de façon complètement aléatoire. Ainsi la quantification, l'appréciation méliorative ou dépréciative, l'intensité (etc.) sont produites grâce à des moules productifs qui reposent sur certains types de métaphore ou de comparaison¹⁵. Les collocations produites de cette manière sont donc partiellement régulières, et il est possible de préciser quelques-unes de leurs règles de formation. Bien que les collocations imagées puissent comporter tous les types de constituants mentionnés plus haut, nous nous limiterons ici, par commodité, aux collocations associant un nom à SP ou à un adjectif.

4.1. L'expression de la collection et de la quantification dans les structures binominales en *de*

En français, l'expression de la collection s'effectue souvent grâce à un syntagme binominal en *de*, dont le premier nom est un nom collectif¹⁶ qui peut avoir une valeur métaphorique (Benninger, 2001). Il y a par ailleurs fréquemment un rapport entre collection et quantification, dans la mesure où la valeur métaphorique fournie par le nom collectif (qui perd alors son autonomie référentielle) permet de signaler une quantité importante :

un forêt de mâts
une armée de marmitons
un essaim d'hélicoptères

Ces noms collectifs, dans cette structure particulière (Dét-N1-*de*-(Dét)-N2) et avec cette valeur quantitative¹⁷, sont toujours, comme le signale N. Flaux (1999), citée par Benninger (2001, 28), interprétés métaphoriquement. Le choix du nom à valeur quantitative n'est pas purement arbitraire, et respecte des critères d'appropriation référentielle¹⁸, parfois très lâches quand le nom est quasi grammaticalisé comme collectif (cf. *une foule de questions*). Les collocations sont des catachrèses, qui se distinguent, par leur degré de lexicalisation, de métaphores plus créatives (cf. *une forêt de mains frémissantes*¹⁹). On peut ajouter à ces noms collectifs jouant quasi naturellement le rôle de quantificateurs, des substantifs qui n'ont pas valeur de collectif, mais qui revêtent parfois également une valeur quantitative,

¹⁴ Certaines de ces relations particulièrement productives sont codées par les fonctions lexicales syntagmatiques standard du *Dictionnaire Explicatif et Combinatoire* (Mel'cuk et al., 1995).

¹⁵ Nous n'entrerons pas ici dans les distinguos parfois subtils qui permettent d'opposer comparaison et métaphore : la structure prépositionnelle binominale dont il sera question ci-dessous, peut, suivant les cas, engendrer plutôt des métaphores (dans les structures quantitatives) ou des comparaisons (avec les appréciatifs et les intensifs).

¹⁶ Une définition rigoureuse de cette notion n'est pas des plus simples. On se référera à la définition fournie par Benninger (2001, 22) : « des noms dont le singulier dénote une pluralité d'entités isolables, de même type et perçues comme un tout ».

¹⁷ Evidemment, un *essaim de guêpes* ne justifie pas une interprétation métaphorique, *essaim* n'ayant pas de valeur quantitative et fournissant le nom collectif approprié pour *guêpes*.

¹⁸ Ainsi on parlera d'un *essaim d'hélicoptères* mais non d'un *essaim de mâts*.

¹⁹ Exemple de J. Lanzmann, cité par Benninger (2001, 28).

acquise métaphoriquement, et perdent tout comme les noms précédents, leur autonomie référentielle :

- une avalanche de questions
- un déluge de paroles
- une brochette d'importantes personnalités

Si *avalanche* et *déluge* sont des quantificateurs du type *beaucoup*, *brochette* joue ici un rôle équivalent à *quelques*, en signalant une quantité restreinte de personnes qui paraissent ensemble, comme des morceaux de viande sur une brochette. Vouloir prédire la possibilité d'association de tels collocatifs à valeur quantitative suppose, on le voit, que soient précisées les conditions d'application référentielle, liées au sémantisme des expressions. *Brochette* s'emploie plaisamment et limite son application aux personnes. Des collocations très diverses sont autorisées avec *avalanche* (*une avalanche de questions, de problèmes, de cadeaux, de produits, de bonnes idées, de demandes, de propositions, de nouveautés, de surprises, de protestations*, etc.), même s'il existe également quelques restrictions, comme avec les noms non comptables avec ? *une avalanche d'eau, de brouillard ? une avalanche de bonheur, de désespoir*. En dehors de ces restrictions de sélection²⁰, un quantitatif d'origine métaphorique, comme *avalanche*, semble pouvoir être considéré comme un outil permettant d'exprimer la quantité, chaque fois qu'il est question d'« objets » (abstraits ou concrets) qui nous « tombent dessus » en grande quantité, soudainement et par surprise. La motivation sémantique de telles expressions reste, en définitive, bien vivace.

4.2. Appréciatifs mélioratifs et péjoratifs

On retrouve dans cette catégorie, à côté d'adjectifs relationnels²¹ (*une douceur angélique*), un syntagme binominal en *de*, qui se différencie du précédent en ce que le collocatif se trouve cette fois en deuxième position et suit la préposition, pour marquer l'appréciation positive ou négative :

- des nerfs d'acier
- un moral d'acier
- une santé de fer
- un morceau de roi
- une voix de sirène
- un caractère de cochon
- un travail de cochon

Alors que les quantitatifs sont sélectionnés en fonction de la cohérence de l'élément métaphorique avec le signifié de la base²², 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. Le recours à des

²⁰ Qu'il faudrait préciser.

²¹ On sait que ces adjectifs indiquent une relation par définition non gradable : **une douceur très angélique*. Mais ce critère semble moins clair ici que dans les exemples prototypiques comme *discours présidentiel*. Les relationnels à valeur appréciative peuvent plus aisément être intensifiés par des adverbes (cf. *une force presque, tout à fait herculéenne*) et semblent donc avoir des caractéristiques spécifiques.

²² On peut ainsi prédire que tous les noms concrets comportant les dimensions sémantiques de verticalité et de hauteur (*cheminée, tour, mât, antenne, pylône, jambe, ...*) peuvent être quantifiés par le collectif *forêt*.

noms propres est fréquent dans cette catégorie, que ce soit dans les structures binominales ou avec des adjectifs relationnels (*une force herculéenne, une querelle byzantine*). On est plus proche de la comparaison que de la métaphore, la préposition *de* introduisant l'élément comparant (*une voix de stentor = une voix comme celle de Stentor*). La description introduite par le comparant excède parfois la simple valeur appréciative : *un travail de bénédictin* ne fournit pas seulement une appréciation positive sur le travail réalisé, mais ajoute des informations du type : « qui requiert du soin et de la patience ». Outre les structures binominales en *de*, et les adjectifs relationnels, on trouve pour exprimer l'appréciation des comparaisons en *comme* (cf. *raisonner comme un pied*). Nous revenons plus bas sur les structures comparatives en *comme*, utilisées plus massivement encore pour l'expression de l'intensité. Retenons que, quelle que soit la structure syntaxique utilisée, on se trouve tantôt 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 (cf. *acier, fer = 'force'*), tantôt dans le cas de collocations opaques, lorsque la motivation est de type historique ou mythologique, et la collocation forcément plus idiosyncrasique. Contrairement à certaines idées reçues, dans ce type de collocations, c'est plutôt l'adjectif ou le SP qui, sémantiquement, sélectionne le nom : *de bénédictin* sélectionne de manière standard le nom *travail*, mais autorise aussi *minutie* (*il oeuvrait avec une minutie de bénédictin*) : on retrouve alors le cas n°3 de notre typologie, celui des collocations régulières, le sens de *minutie* étant inclus dans celui de *de bénédictin*.

4.3. Intensifs

Le système de production des collocatifs intensifs est très proche de celui des appréciatifs (on sait d'ailleurs que nombre d'appréciatifs ont aussi une valeur intensive). On retrouve par exemple le même type d'adjectifs relationnels (*un froid sibérien, un froid polaire*) et de structures binominales en *de*. Dans les structures binominales ou les systèmes comparatifs comportant un collocatif à nom d'animal, la situation semble plus variée que pour les appréciatifs. On trouve des collocations dans lesquelles l'animal est censé incarner typiquement tel défaut ou telle qualité (*têtu comme une bourrique, malin comme un singe, une mémoire d'éléphant*, etc.), mais on rencontre également des exemples dans lesquels l'animal ne représente pas la propriété : *un froid de canard* n'est pas un froid propre au canard comme la ruse l'est au singe. De même, dire que l'on est *malade comme un chien* ne signifie pas que les chiens ont pour propriété d'être malades. Cette diversité s'explique par le fait que l'on a souvent affaire à des comparaisons incomplètes ou tronquées, dans lesquelles la motivation, qui ne repose pas forcément sur un symbolisme culturel, est généralement peu accessible (cf. *froid de canard*). A l'inverse, on trouve aussi des comparaisons développées, presque figées : *un temps à ne pas mettre un chien dehors* (collocation transparente mais peu prédictible). Les comparaisons introduites par des structures en *comme* sont pour la plupart très idiomatiques et peu prédictibles, sauf peut-être en ce qui concerne les noms typiques de propriétés (cf. pour la couleur : *blanc comme de la neige, noir comme l'ébène, comme la suie*).

En résumé, si nombre de collocations relèvent de processus figuraux, on trouve certaines régularités dans le codage des relations sémantiques exprimées. On a vu que c'était assez nettement le cas pour les collocatifs à valeur quantitative, qui conservent une forte motivation sémantique. Les appréciatifs et les intensifs recourent à des structures régulières, même si le caractère prédictif du collocatif n'est pas toujours assuré. Enfin,

certaines collocations imagées sont parfaitement régulières, une fois que l'on a précisé les conditions sémantiques qui justifient l'emploi du collocatif : ainsi *formalisme desséchant*, qui comporte un collocatif métaphorique lexicalisé, peut s'appliquer à un ensemble de noms référant à des constructions intellectuelles qui, justement parce qu'elles sont intellectuelles, peuvent s'éloigner de la réalité empirique, ou des faits sensibles (*une théorie, une doctrine, une opinion, une conception desséchante*). On retrouve dans ce cas des collocations que l'on peut considérer comme transparentes, la décryptage de la métaphore ne posant en principe pas de problème.

5. Collocations et notions voisines

5.1. Collocation et expressions figées

Il est usuel, dans une échelle du figement, de considérer les collocations comme un cas intermédiaire entre des combinaisons libres du type *argument judicieux* et des expressions opaques, mémorisées comme telles du type *nid d'ange, prendre le taureau par les cornes*. Ces dernières, qui bénéficient de multiples appellations sans qu'aucune ne soit consensuelle (« expressions figées », « locutions », « unités phraséologiques » (Bally, 1909), « phrasèmes complets » (Mel'cuk et al., 1995), « unités polylexématiques » (Corbin, 1997)) semblent, sur le plan sémantique, se distinguer nettement des collocations.

La principale caractéristique sémantique est la non-compositionalité du sens des expressions figées, le sens du tout ne pouvant pas être calculé à partir du sens des parties, chaque élément perdant aussi sa fonction référentielle²³. Dans *nid d'ange*, le sens de l'expression ne peut pas être calculé à partir des sens de *nid* et d'*ange*. Le référent de l'expression ne renvoie ni à celui d'*ange*, ni à celui de *nid*. Bien entendu, de nombreux procédés linguistiques (métaphores, métonymies, etc.) peuvent souvent expliquer la genèse de ces expressions, comme celle des collocations d'ailleurs (cf. 4.). En synchronie, ces unités sont cependant rétives aux processus traditionnels d'interprétation, tant en analyse qu'en production, par exemple pour le traitement automatique ou dans la maîtrise d'une langue seconde. La lexicologie explicative et combinatoire (Mel'cuk et al., 1995) accorde d'ailleurs de façon méritée²⁴ aux « phrasèmes complets » le statut d'unité linguistique au même titre que les mots simples alors que les collocations seront codées sous l'article de la base.

Un second critère sémantique, d'ordre référentiel, est souvent mis en avant, en particulier pour les expressions figées de type nominal : le fait que l'expression renverrait à un référent unique. Ce paramètre paraît en fait délicat à utiliser, surtout si l'on vise à distinguer les différents types de collocations des expressions figées. Certains collocatifs ont une fonction plus typante que qualifiante. Par exemple, dans l'expression *café noir*, l'élément *noir* indique à la fois une propriété du café et une sorte de *café*, alors que *endurci*

²³ Nous n'entrons pas dans le débat de savoir si le figement est cause ou conséquence de la lexicalisation (cf. Corbin, 1997).

²⁴ Si l'on suit le plaidoyer de Hausmann (1997) : « ... entre les mots et les locutions il y a dans nos dictionnaires un abîme. Les mots bénéficient d'articles, les locutions n'en bénéficient pas. Notre lexicographie est ainsi faite que les mots ont dans la société lexicale un statut de grand, de noble, alors que les locutions ne sont que des roturières, souvent laissées pour compte, en tout cas des parents pauvres. On pourrait parler de la grande misère des locutions françaises. ... »

associé à *célibataire* qualifie ce dernier. Si l'on construit des ontologies, les collocations du type *café noir* peuvent être considérées comme des hyponymes du mot base, ici *café*. Un café noir est une sorte de café alors qu'un célibataire endurci n'est pas une sorte de célibataire. De ce fait, les collocations à collocatifs typants tendent à être perçues comme des unités référentielles. Cependant, sur le plan linguistique, ces collocations peuvent être analysées de façon semi-compositionnelle et il nous semble essentiel dans l'étude de la langue de dissocier le plan sémantique du plan référentiel.

D'autres critères syntaxiques comme la contiguïté des éléments, la non insertion de modificateurs, la non substituabilité des éléments caractérisent un sous-ensemble de ces expressions, mais comme pour les collocations, ces paramètres ne sont pas définitoires de la classe²⁵. Comme le souligne Gaston Gross (1997), sur le plan syntaxique, il n'y a pas d'opposition binaire entre expressions libres et expressions figées, mais plutôt des degrés de figement. Sur le plan sémantique, néanmoins, la classification des expressions selon le type de figement comme dans le système mel'cukien (« phrasèmes complets » sans compositionnalité, « semi-phrasèmes » pour les collocations où un des éléments conserve son sens habituel, « quasi-phrasèmes » pour des exemples comme *donner le sein* où le sens du tout exprime davantage que le sens des parties) nous paraît préférable à la solution du continuum souvent prônée²⁶. Le codage lexicographique impose une nécessaire formalisation de ces phénomènes, qui paraît souhaitable pour formuler des hypothèses linguistiques.

5.2. Collocations et restrictions de sélection sémantique

À côté des collocations opaques (*peur bleue*, *beau comme un camion*) et des collocations transparentes mais non prédictibles (*dormir comme un nouveau-né*, *une envie folle*, *célibataire endurci*), apparaissent certaines associations lexicales apparemment régies par des régularités. Les cooccurrences semblent sémantiquement explicables et paraissent même parfois prédictibles. Dans ce cas, on peut légitimement se demander si les associations ne sont pas simplement contrôlées par des restrictions de sélection.

Les restrictions de sélection sémantique, de la même façon que les collocations, portent sur la combinatoire lexicale puisqu'elles servent à formaliser le type sémantique des arguments attendus d'une lexie. Le sujet de *voir* (dans son acception centrale) sera plutôt animé, l'adjectif *vert* portera sur un nom renvoyant à un référent concret. Ce concept a été introduit dès les premières versions de la grammaire générative (Chomsky, 1965) pour contrôler dans la phase d'« insertion lexicale » les phrases qui seraient stylistiquement anormales mais grammaticales du type *le lait boit le chat*. Pour éviter de générer ce type d'énoncé, un contrôle est effectué dans la phase d'« insertion lexicale », en vérifiant la cohérence des traits inhérents sur les noms (*lait* reçoit le trait [+concret], *chat* le trait [+animé]) et des traits de sélection [*boire* prend un sujet de type [+animé] et un objet de type [+concret]]. Les difficultés d'application de ce système – binarité des traits,

²⁵ Les tests, par ailleurs, varient fortement d'une catégorie syntaxique à l'autre. Les « locutions verbales » sont toujours beaucoup plus souples du fait de l'insertion presque toujours possible d'adverbes.

²⁶ La typologie proposée par Mel'cuk semble par ailleurs avoir été éprouvée de façon convaincante sur un nombre non négligeable d'entrées lexicales des *DEC* même si on ne peut pas nier qu'une zone d'incertitude demeure pour quelques cas délicats.

redondance de traits comme [+ animé] et [+ humain] – ont été souvent soulignées²⁷. Des adaptations plus souples, transposées au niveau sémantique, ont été proposées par exemple chez Jackendoff (1990) qui associe aux arguments des structures conceptuelles des catégories conceptuelles spécifiques. Par exemple, la structure conceptuelle de ‘aller’ prendra comme agent une catégorie conceptuelle de type ‘chose’ et non de type ‘propriété’. Cette contrainte bloquera l’interprétation de **la sincérité entrera dans la chambre*.

Les restrictions de sélection permettent donc de restreindre la cooccurrence de séries de lexies répondant à des critères sémantiques spécifiques, décrits à l’aide de traits ou de catégories conceptuelles. Les collocations, elles, mettent en relation des expressions linguistiques et non des classes d’expressions linguistiques. Cependant, dans certains cas, les collocations apparaissent fortement prédictibles et semblent répondre à des critères sémantiques réguliers.

Dans les langues de spécialité, en particulier dans les domaines technologiques, les associations lexicales, parfois appelées *combinaisons lexicales spécialisées* (L’Homme & Bertrand, 2000) semblent souvent répondre à des patrons sémantiques réguliers. Cela a été montré pour les combinaisons V-N dans le domaine de l’informatique (L’Homme, 1998), dans le domaine de l’aéronautique et même de la philosophie (L’Homme & Bertrand, 2000), dans le domaine des CD-ROMs (Meyer & Mackintosh, 1994). Par exemple, le verbe *charger* dans le domaine de l’informatique prendra comme objet direct un nom qui appartient à la classe conceptuelle des logiciels. Dans ce cas, la collocation peut être dite « conceptuelle » (Heid, 1994), dans la mesure où le collocatif peut être associé à une série de lexies partageant des caractéristiques sémantiques. Décrire les associations lexicales revient alors à associer au collocatif une catégorie conceptuelle donnée, pouvant être décrite dans une taxonomie. Des tentatives de codage en ce sens ont été proposées par L’Homme & Bertrand (2000), mais des études en grandeur réelle restent à faire dans ces domaines. Ces associations peuvent donc être décrites comme étant régies par des restrictions de sélection sémantique. On peut alors se demander si le terme de collocation (et encore plus celui de « collocation conceptuelle ») est encore adapté. On pourra préférer l’appellation plus neutre de *combinaison lexicale spécialisée*.

Cependant, même en langue de spécialité, dans d’autres domaines, les régularités ne sont pas aussi manifestes. Heid (1994) reprenant les exemples de B. Cohen dans le sous-langage de la bourse (Cohen, 1986) étudie les associations N-V. S’il note que certains verbes courants (*augmenter*, *s’accroître*) sont clairement associés à des classes ou à des sous-classes de nom, des collocatifs plus métaphoriques et moins courants comme *bondir* ou *s’effondrer* ont un comportement plus instable et semblent davantage sélectionner un ensemble de noms partageant des propriétés sémantiques communes mais ne constituant pas de façon évidente une classe. Ces faits se rapprochent des données observées dans les études effectuées en « langue générale » qui montrent qu’il existe souvent des tendances à la régularité, sans règle absolue. Par exemple, dans le cas des verbes supports, on relève que le verbe *avoir* tend à sélectionner des noms d’état (*patience*, *peur*, *envie*) alors que le support *faire* est plutôt choisi avec des noms d’action comme *éloge*, *arbitrage*, *lavage* ... Il serait alors partiellement possible de prédire le verbe support à partir de la classe sémantique du nom. Mel’cuk et Wanner (1996) montrent qu’il existe une forte corrélation

²⁷ Le système des restrictions de sélection sémantique est souvent considéré comme normatif. Des aménagements peuvent cependant être imaginés pour autoriser les « transgressions ».

entre certains traits sémantiques des noms d'émotions et les collocatifs verbaux les accompagnant. Dans une étude similaire sur les noms de « joie », nous avons montré qu'un adjectif comme *affreux* avait une fonction intensive en cooccurrence avec des noms exprimant une polarité négative comme *désespoir* (Grossmann & Tutin, à paraître). Dans ce dernier cas, le principe d'association entre base et collocatif ne met pas en jeu une classe conceptuelle de noms (les noms d'émotion de polarité négative ne constituent pas naturellement une classe), mais des dimensions sémantiques spécifiques associées au nom.

Selon Heid (1994, 239), « il semble y avoir une différence graduelle entre les collocations « conceptuelles », prédominantes en terminologie, et les simples collocations « lexicales », prédominantes dans la langue générale ». Pour lui, le sous-langage de la Bourse constituerait un stade intermédiaire entre ces deux extrémités. En fait, plutôt que d'attribuer la différence de comportement des collocatifs à leur appartenance à la « langue générale » ou aux langues de spécialité, on peut se demander si certains domaines de l'expérience, en particulier ceux dans lesquels les référents concrets sont prédominants, ne se prêtent pas plus facilement aux « collocations conceptuelles » que les domaines plus abstraits. Le domaine de la bourse présente en effet de nombreux points communs avec le champ sémantique des émotions. Les noms y sont majoritairement abstraits (*hausse*, ...) et les métaphores n'y sont pas absentes (*bondir*, *s'effondrer*). En « langue générale », des champs sémantiques concrets comme celui des légumes illustrés dans le DEC 4 (Mel'cuk et al., 2000) présentent des régularités semblables à celles que l'on observe dans la terminologie de l'informatique : le nom *tomate* hérite naturellement des collocatifs associés à *légume*. Dans des domaines plus abstraits comme la bourse ou le champ des émotions, les concepts sont moins bien structurés, et ne sont pas clairement organisés en taxonomies.

En bref, les régularités sémantiques entre base et collocatif semblent relativement systématiques dans certains domaines de l'expérience, bien structurés, comme c'est souvent le cas en langue de spécialité. Dans les domaines abstraits, on relève davantage d'idiosyncrasie, et le partage entre combinaisons libres et collocations semble nettement plus délicat. On peut souvent mettre en évidence des traits sémantiques communs aux bases associées à un collocatif donné, mais la mise au jour de ces régularités suppose une analyse approfondie. Dans la pratique, la différence a une incidence sur le codage lexicographique des combinaisons lexicales. Dans le premier cas, les combinaisons seront décrites autour de classes conceptuelles, comme des restrictions de sélection (cf. L'Homme, à paraître). Dans le second cas, le codage des régularités devra s'accompagner d'une prise en compte des exceptions et des anomalies, ce qui alourdira considérablement le traitement, rendant même parfois la faisabilité de l'opération difficile. Le repérage des régularités nous paraît néanmoins indispensable pour le décodage des collocations, utile tant dans une perspective d'apprentissage que pour l'analyse en TAL.

6. Conclusion

Les expressions que l'on regroupe sous l'étiquette un peu fourre-tout de « collocation » sont donc de natures diverses, et il n'apparaît pas pertinent de leur appliquer un traitement homogène. Il nous semble préférable, pour le codage, de nous appuyer sur une typologie des combinaisons lexicales en recourant à des critères sémantiques, suivant la tradition mel'cukienne. Le but visé à terme est de proposer un traitement lexical

différencié selon les différents profils de collocations. Nous avons distingué trois types différents :

- dans le cas des collocations *opaques*, l'association paraît arbitraire et non transparente sur le plan sémantique ;
- lorsque la collocation est interprétable mais peu prédictible (le collocatif n'ayant pas de statut lexical), on a affaire à des collocations *transparentes* ;
- enfin, nous avons montré qu'il existait un nombre de cas non négligeables dans lesquels le phénomène collocatif suivait des règles régulières d'association sémantique : nous avons parlé alors de collocations *régulières*²⁸.

Il ne s'agit là évidemment que d'une esquisse de typologie, qui devra être affinée sur la base de critères sémantiques et syntaxiques. En l'état, elle nous permet cependant déjà de questionner certaines approches classiques du phénomène collocatif. Les lexicologues, principalement Mel'cuk et Hausmann prônent généralement un traitement axé sur l'encodage lexicographique, où la description est effectuée au cas par cas, à partir de l'article de la base. Si ce traitement se justifie pleinement pour les collocations arbitraires, il nous semble plus discutable pour les collocations transparentes et régulières, pour lesquelles il est difficile de prédire, mais plus facile d'expliquer. Dans des applications de linguistique appliquée axées sur le décodage, par exemple la compréhension en FLE ou FLM, ou encore les phases d'analyse en TAL, il paraît en effet pertinent d'analyser ce qui motive l'association, tant au niveau du sens du collocatif qu'au niveau des procédés sémantiques (métaphores, comparaisons) mis en jeu. Encodage et décodage nous paraissent donc deux perspectives complémentaires plutôt qu'antagonistes dans le traitement des collocations.

A. Tutin et F. Grossman
LIDILEM, UFR Sciences du Langage
Université Stendhal-Grenoble-3
BP 25, Grenoble Cedex 09
{tutin,grossmann}@u-grenoble3.fr

Références

- Bally, C. (1909) : *Traité de stylistique française*. Paris, Klincksieck.
- Benson, M., Benson, E. & Ilson R. (1990) : *The BBI Combinatory Dictionary of English. A Guide to Word Combinations*. Amsterdam, Benjamins.
- Benninger, C. (2001) : Une meute de loups / une brassée de questions : collection, quantification et métaphore. *Langue française* 129, 21-34.
- Chomsky, N. (1965) : *Aspects de la théorie syntaxique* (trad.). Paris, Seuil.
- Cohen, B. (1986) : *Lexique de cooccurrents - Bourse et conjoncture économique*. Montréal, Linguatex.
- Corbin, D. (1997) : Locutions, composés, unités polylexématiques : lexicalisation et mode de construction. In Martins-Baltar, M. (ed.), 53-101.
- Cruse, D.A. (1986) : *Lexical Semantics*. Cambridge, Cambridge University Press.

²⁸ En ayant conscience du caractère provocant de cette proposition, puisque les collocations sont souvent considérées comme irrégulières par définition.

- Flaux, N. (1999) : A propos des noms collectifs. *Revue de Linguistique romane* 63, 471-502.
- Giry-Scheider, J. (1987) : *Les prédicats nominaux en français : les phrases simples à verbes supports*. Genève, Droz.
- Gross, M. (1981) : Les bases empiriques de la notion de prédicat sémantique. *Langages* 63, 7-52.
- Gross, G. (1988) : Degré de figement des noms composés. *Langages* 90, 57-72.
- Gross, G. (1997) : Du bon usage de la notion de locution. In Martins-Baltar M. (ed.), 201-224.
- Halliday, M.A.K. & Hasan, R. (1976) : *Cohesion in English*. London, Longman.
- Fontenelle, T. (1997) : *Turning a Bilingual Dictionary into a Lexical-Semantic Database*. Tübingen, Max Niemeyer Verlag.
- Hausmann, F. J. (1989) : Le dictionnaire de collocations. In Hausmann, F.J., Reichmann, O., Wiegand, H.E., Zgusta, L. (eds), *Wörterbücher : ein internationales Handbuch zur Lexicographie. Dictionaries. Dictionnaires*. Berlin, De Gruyter, 1010-1019.
- Hausmann, F. (1997) : Tout est idiomatique dans les langues. In Martins-Baltar M. (ed.), 277-290.
- Heid, U. (1994) : On Ways Words Work Together. Topics in Lexical Combinatorics. In *EURALEX'94 Proceedings*. Vrije Universiteit Amsterdam, 226-257.
- Iordanskaja, L., Kim, M. & Polguère, A. (1996) : Some Procedural Problems in the Implementation of Lexical Functions. In Wanner L. (ed), 279- 297.
- Jackendoff, R. (1990) : *Semantic structures*. Cambridge, MIT Press.
- L'Homme, M.-C. (1998) : Le statut du verbe en langue de spécialité et sa description lexicographique. *Cahiers de Lexicologie* 73, 61-84.
- L'Homme, M.-C. & Bertrand, C. (2000) : Specialized Lexical Combinations : Should they be described as Collocations or in Terms of Selectional Restrictions ? In *Proceedings of Euralex 2000*, 497-506.
- L'Homme, M.-C. (à paraître) : Combinaisons lexicales spécialisées (CLS) : Description lexicographique et intégration dans les banques de terminologie. In Grossmann, F. & Tutin, A. (à paraître), *Modélisation des collocations*.
- Martin, R. (1997) : Sur les facteurs du figement lexical. In Martins-Baltar, M. (ed.), 291-307.
- Martins-Baltar, M. (ed.) (1997) : *La locution entre langue et usage*. Fontenay, ENS Editions.
- Mel'cuk, I. (1988) : *Dependency syntax : Theory and Practice*. Albany, State University of New York Press.
- Mel'cuk, I., Clas, A. & Polguère, A. (1995) : *Introduction à la lexicologie explicative et combinatoire*. Louvain, Duculot.
- Mel'cuk, I. et al. (1984, 1988, 1992, 2000) : *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques*. Vol. I, II, III, IV. Montréal, Presses de l'Université de Montréal.
- Mel'cuk, Igor (1998) : Collocations and Lexical Functions. In A. P. Cowie (ed.), *Phraseology. Theory, Analysis and Applications*. Oxford, Clarendon Press, 23-53
- Meyer, I. & Mackintosh, K. (1994) : Phraseological Analysis and Conceptual Analysis: Exploring a Symbiotic Relationship in the Specialized Lexicon. In *EURALEX'94 Proceedings*, Vrije Universiteit Amsterdam.
- Riegel, M., Pellat, J.C. & Rioul, R. (1997) : *Grammaire méthodique du français*. Paris, PUF.
- Sinclair, John (1991) : *Corpus, Concordance, Collocations*. Oxford, Oxford University Press.
- Smadja, F. & McKeown, K. (1991) : Using collocations for language generation. *Computational Intelligence* 7, 229-239.

- Vivès, R. (1993) : *Avoir, prendre, perdre: constructions à verbes supports et extensions aspectuelles*. Thèse de 3ème cycle, Université Paris VIII.
- Wanner, L. (ed.) (1996) : *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam, Benjamins.
- Williams, G. (1999) : *Les réseaux collocationnels dans la construction et l'exploitation d'un corpus dans le cadre d'une communauté de discours scientifique*. Thèse de doctorat, Université de Nantes.
- Williams, G. (à paraître) : Les collocations et l'école contextualiste britannique. In Grossmann, F & Tutin, A. (à paraître), *Modélisation des collocations*.