

# Paramètres de délimitation des collocations du français courant

J. Bossé-Andrieu et G. Mareschal  
*Université d'Ottawa*

Bien que, depuis une dizaine d'années, plusieurs dictionnaires de collocations aient vu le jour — surtout pour des langues autres que le français —, la délimitation des collocations reste entourée de flou. En effet, si, par collocation, on entend une combinaison lexicale de deux éléments sémantiquement autonomes unis par un lien d'affinité contraignant, il n'est pas toujours aisé de déterminer les limites de l'affinité. Entre les combinaisons totalement libres et les locutions totalement figées s'étend une zone qui comprend des combinaisons qui correspondent exactement à cette définition. Toutefois, de chaque côté de cette zone se trouvent deux zones floues qui contiennent des combinaisons dont on peut se demander si elles doivent entrer dans la catégorie des collocations. Après avoir montré que l'ensemble de ces combinaisons forme un continuum, nous présentons un certain nombre de paramètres qui facilitent la délimitation des collocations du français courant.

Although a number of dictionaries of collocations have appeared in recent years — mostly in languages other than French — the notion of 'collocation' remains somewhat fuzzy around the edges. A collocation is generally understood to be a combination of two lexemes that are semantically independent but which are linked together by a strong, restrictive association. Nevertheless it is not always easy to determine the boundaries of this type of affinity. Between the extremes of completely free word combinations and completely frozen idioms, collocations (as defined above) clearly occupy an intermediate portion of the spectrum. At each edge of this middle ground, however, there are grey areas where one finds word combinations which might or might not be classed as collocations. After first demonstrating that associative word combinations do indeed form a continuum of this kind, we propose some parameters to help delimit the scope of collocations in everyday contemporary French.

Dans l'enseignement de la langue, que ce soit celui de la langue maternelle ou celui de la langue seconde, ou encore dans l'enseignement de la traduction, l'une des difficultés rencontrées par les professeurs est la méconnaissance des collocations par les étudiants. Comme le disent Heid et Freibott (1991, p. 78), la « lisibilité » d'un texte repose sur « l'apparition de collocations, c'est-à-dire de combinaisons de lexèmes préférées par une communauté linguistique ». Or, cette connaissance primordiale des collocations d'une langue donnée n'est pas facile à acquérir, et les ouvrages de référence usuels, tels que les dictionnaires unilingues ou bilingues, ne donnent pas tous les renseignements nécessaires.

Prenons, par exemple, le nom *lien*. Comment savoir quel verbe signifiait « rendre plus fort ou plus étroit » accompagne ce nom en français? Est-ce « resserrer » ou « renforcer »? Si l'on consulte le *Nouveau Petit Robert (NPR)* à l'entrée *lien*, on trouve, au sens 3, les combinaisons *nouer des liens*, *rompre des liens*, mais pas *resserrer* ni *renforcer des liens*; sur le plan traductionnel, le *Robert et Collins Senior (R & CS)* ne donne pas non plus cette combinaison à l'entrée française de *lien* et il ne la donne pas plus dans le volet anglais-français ni à *strengthen*, ni à *tighten*, ni à *link*, ni à *bond*. Il faudrait donc idéalement que les étudiants puissent consulter un dictionnaire qui, pour une langue donnée, indiquerait, à partir d'un mot-entrée, les mots avec lesquels il se combine. Dans le cas du français, il existe bien quelques dictionnaires unilingues de ce type, mais ils sont soit vieillots (Lacroix), soit restreints dans leur portée puisqu'ils traitent de langues de spécialité (Cohen, Lainé)<sup>1</sup>. Quant au *Dictionnaire combinatoire du français contemporain (DCFC)* de Mel'čuk, s'il représente ce que l'on peut faire de plus complet, il faut reconnaître que la raçon de son exhaustivité en est, d'une part, la lenteur des travaux — et donc le peu d'articles qu'il contient actuellement — et, d'autre part, le manque de « convivialité ». C'est cette absence d'un recueil de collocations simple à utiliser qui a conduit une équipe de professeurs de l'École de traduction et d'interprétation d'Ottawa à entreprendre l'élaboration d'un dictionnaire des combinaisons du français courant (*COMBIDIC*).

Dans ce recueil, nous avons l'intention de répertorier, pour les mots retenus dans la nomenclature, les combinaisons syntaxiques et les combinaisons lexicales — ou collocations — propres à ces mots. Notre travail de dépouillement et l'examen de la littérature dans le domaine nous ont amenées à constater que la notion de collocation était très floue, qu'il était extrêmement difficile d'en dresser les limites et, par conséquent, de déterminer avec précision les combinaisons lexicales qui devaient entrer dans le *COMBIDIC*. Dans cet article, nous nous proposons donc, dans un premier temps, de cerner la notion de collocation et d'en définir les limites et, dans un deuxième temps, de présenter des paramètres pouvant servir à la sélection des collocations.

### **Définition et limites de la collocation**

#### ***Que faut-il entendre par collocation?***

Le mot « collocation » a des acceptions diverses selon les linguistes et les écoles de pensée. Alors que, en anglais, « collocation » peut, dans un sens large, être synonyme de cooccurrence et signifier la présence simultanée de deux ou plusieurs mots dans un même énoncé (“the occurrence of two or more words within a short space of each other in a text” [Sinclair, 1991, p. 171]), en français, le mot « collocation » a un sens plus restreint et désigne une combinaison usuelle de mots unis dans un rapport syntagmatique. Il s'agit d'une paire de lexèmes

qui, dictée par l'usage, se retrouve fréquemment dans le discours d'une communauté linguistique donnée, dont les éléments s'attirent mutuellement et entre lesquels existe un lien grammatical. Ainsi, la combinaison *essuyer un affront* est une collocation. Le verbe *essuyer*, dans son sens figuré, ne se combine qu'avec un nombre limité de noms (*affront, échec, refus, tempête*) qui l'« appellent » et avec lesquels il a des liens privilégiés. Il y a, entre les deux éléments des combinaisons ainsi formées, un lien d'affinité et un lien contraignant : un élément en attire un autre ou en repousse un autre comme le ferait un aimant (par exemple, *succès* repousse *essuyer*, mais attire *remporter*). On comprend donc pourquoi les collocations sont aussi appelées « associations privilégiées » (Béjoint et Thoiron, 1993), ou « combinaisons sous contrainte » (Hausmann, 1979). Elles se composent d'un mot-base — souvent appelé « *node* » en anglais (Sinclair, 1991) — accompagné à gauche ou à droite d'un « collocatif » ou « collocateur » (Heid, 1991) — généralement appelé « *collocate* » en anglais. Ainsi, dans *essuyer un affront*, le mot-base est *affront* et le collocatif est *essuyer*. Le collocatif est le plus souvent un seul mot (par exemple, le verbe *essuyer* dans l'exemple donné, ou encore l'adverbe *immensément* dans la collocation *immensément riche*), mais peut aussi comporter plusieurs mots (par exemple, *à millions*, dans la collocation *riche à millions*). C'est dans ce sens restreint que nous employons le mot « collocation »<sup>2</sup>.

### *Les limites de la collocation*

La collocation doit être distinguée, d'une part, de la locution et, d'autre part, de la combinaison libre.

Ce qui distingue la **collocation** de la **locution**, c'est la différence dans le degré de transparence et le degré de figement. Une combinaison est transparente lorsqu'on peut en saisir facilement le sens. Une combinaison est figée lorsque ses éléments constituent un tout indissociable.

La **collocation**, comme le note Hausmann (1989), se caractérise par la transparence et le non-figement. Les deux éléments d'une collocation gardent leur autonomie sémantique, et le sens de la collocation peut se déduire des éléments qui la composent. Alors que le sens du mot-base ne dépend pas du collocatif, en revanche celui du collocatif est fonction du mot-base. Par exemple, dans la collocation *essuyer un affront*, qui a pour mot-base *affront*, le verbe *essuyer*, sous l'influence d'*affront*, prend le sens de « subir » et n'a plus son sens propre et premier de « enlever l'eau ou la poussière » (comme dans *essuyer les verres*); dans la collocation *un profond respect*, dont le mot-base est *respect*, l'adjectif *profond* joue le rôle d'un intensif : il prend le sens de « grand » sous l'influence du mot-base *respect*, et ne fait pas penser, comme dans *un puits profond*, à une chose dont le fond est éloigné de la surface<sup>3</sup>. Par ailleurs, même si la cohésion des deux éléments d'une collocation est grande — en raison notamment de la fréquence à laquelle ils apparaissent ensemble, elle

n'est pas figée : le collocatif peut être remplacé par un synonyme ou par un antonyme (on peut dire par exemple *infliger un affront*); la collocation se laisse manipuler — par exemple, on peut ajouter un adjectif au mot-base (*essuyer un terrible affront*), mettre un élément au pluriel ou intercaler des éléments entre le mot-base et le collocatif (*les affronts qu'il a essuyés*).

La **locution**, elle, se caractérise par son opacité et son figement. Les éléments de la locution perdent leur autonomie sémantique au profit du sens global. Le sens global d'une locution n'est pas la somme du sens des éléments qui la composent et peut donc difficilement être déduit du sens de ceux-ci. Ce sens global — le plus souvent métaphorique — est opaque et doit être appris (ex.: *rater le coche* [= laisser passer une occasion favorable (*Lexis*)]; *trancher le mot* [= parler brutalement, sans ménagements], *tirer une carotte* [= tromper]). La locution est aussi figée. Ses éléments, comme le dit Zaréba (1977) « font bloc ». Elle ne se laisse pas facilement décomposer et elle ne peut subir de modification; ainsi, dans les exemples précédents, il est impossible de mettre au pluriel les mots *coche*, *mot* et *carotte* sans détruire le sens de la locution, ou de manipuler ces expressions (\**les coches qu'il a ratés!*).

Ce qui distingue la **collocation** de la **combinaison libre**, c'est l'affinité qui unit les éléments. Dans la **collocation**, il y a affinité entre les deux éléments; le collocatif est appelé par le mot-base en raison d'une convention imposée par l'usage d'une communauté linguistique donnée<sup>4</sup>. Comme on l'a vu, le mot-base va de préférence avec un collocatif plutôt qu'un autre et parfois en exclut certains. Dans l'exemple classique *gravement malade* et *grièvement blessé*, l'adjectif *blessé* « appelle » l'adverbe *grièvement* et peut être combiné aussi avec *gravement*, alors que l'adjectif *malade* « repousse » *grièvement*. C'est ce que Hausmann (1989) appelle « la combinabilité restreinte ». Cette affinité, qui se révèle par la fréquence dans le discours, fait que, pour un locuteur natif, les collocations ou « combinaisons restreintes » vont d'elles-mêmes; elles font partie de son bagage linguistique : un francophone choisit, sans savoir pourquoi, l'adverbe qui convient pour accompagner « *blessé* » et « *malade* ». Ce sont les caprices de l'usage qui font que, alors qu'en anglais on « *pave* » [*pave*] *la voie à quelque chose*, on « *achète* » [*buy*] *une assurance*, on peut « *créer* » [*create*] ou « *faire* » [*make*] *une bonne impression*, en français — en principe — on ouvre *la voie à quelque chose*, on *contracte une assurance* et on ne peut que *faire une bonne impression*. Rien ne permet à un étranger de « produire » ces combinaisons, qui résultent d'une convention, et de savoir quel collocatif convient au mot-base<sup>5</sup>.

Dans la **combinaison libre**, les deux éléments sont syntaxiquement et sémantiquement indépendants. Aucune affinité particulière n'existe entre les deux éléments. Le mot qui accompagne le mot-base est imprévisible et n'est choisi qu'en fonction de contraintes syntaxiques ou extralinguistiques. Aucun lien conventionnel n'unit les deux éléments. Par exemple, le mot-base *mot* peut

se combiner avec des verbes tels que *imaginer, préférer, prendre, regarder, rencontrer*, etc. La cooccurrence (c'est-à-dire l'apparition simultanée des deux éléments dans le même énoncé) est fortuite sur le plan linguistique. À part le cas des combinaisons forgées avec la volonté de créer un effet de style, les combinaisons libres sont totalement transparentes et leur sens global est la somme du sens des deux éléments qui les composent.

### *Les collocations se situent au centre d'un continuum*

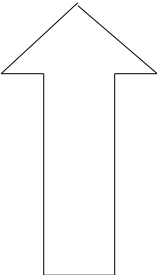
En fait, on peut considérer que l'ensemble des combinaisons lexicales d'un mot-base constitue un **continuum**. Au pôle inférieur se trouvent les combinaisons libres, qui se caractérisent par leur caractère transparent et fortuit. Au pôle supérieur se trouvent les locutions, qui se caractérisent par leur caractère opaque et figé.

Entre les deux pôles se trouvent les collocations. Ces combinaisons, transparentes ou quasi-transparentes, sont conventionnelles et donc prévisibles pour le locuteur natif<sup>6</sup> : l'affinité qui existe entre les deux éléments et la fréquence de ces combinaisons dans le discours font qu'un mot-base évoque instantanément un certain nombre de collocatifs, mais non tous les mots avec lequel il pourrait se combiner.

Les combinaisons qui se trouvent aux deux extrémités « extrêmes » sont faciles à déterminer. Mais entre, d'une part, l'« opaque » et le « transparent » et, d'autre part, le « prévisible » et le « fortuit » existent des zones floues. Ainsi, entre l'opaque et le transparent existe une zone dans laquelle entrent par exemple *manger ses mots* (c'est-à-dire prononcer des mots de façon indistincte; en anglais, *to swallow one's words*) ou encore *ne pas souffler mot* (*not to breathe word*). Il y a là un chevauchement entre la collocation et la locution; bien que chaque élément de la combinaison garde un sens — ici figuré pour le verbe — et que chaque élément garde une certaine autonomie sémantique (la preuve en est que l'anglais emploie un verbe de sens légèrement différent), la transparence n'est pas totale.

D'autre part, entre le fortuit et le prévisible existe aussi une zone floue; en effet, il n'est pas toujours facile de déterminer si la cooccurrence est due à une convention linguistique ou à une réalité extralinguistique. Que fait-on d'un « mot » sinon *le lire, l'écrire, l'épeler*, etc ? Il est donc logique que des cooccurrences telles que *lire un mot, écrire un mot, épeler un mot* soient fréquentes. Étant donné cette fréquence, dictée par des raisons extralinguistiques, on peut hésiter et se demander si le lien qui unit les éléments des combinaisons de ce type est assez contraignant pour en faire des collocations « à part entière ».

On peut donc représenter ainsi le continuum constitué par les combinaisons :

locutions		opaque/figé
xxxxxxxxxx		xxxxxxxxxx
collocations		transparent/prévisible
xxxxxxxxxx		xxxxxxxxxx
combinaisons libres		transparent/fortuit

La flèche indique le passage graduel du transparent et fortuit à l'opaque et figé; la zone située en bas de la flèche correspond aux combinaisons libres (transparentes et fortuites); la zone située en haut de la flèche correspond aux locutions (opaques et figées); la zone du milieu correspond aux collocations, dont la transparence peut ne pas être totale et auxquelles la combinabilité restreinte et le caractère prévisible confèrent un aspect partiellement figé. Aux croix (x) correspondent les zones floues.

### Paramètres de délimitation

Pour que les combinaisons retenues soient bien des collocations, il faut s'assurer qu'elles ne sont ni des combinaisons figées (locutions) ni des combinaisons libres, mais que ce sont bien des combinaisons restreintes, transparentes et prévisibles. Nous proposons, pour ce faire, une série de trois paramètres permettant de repérer les collocations parmi les combinaisons données par un corpus : fréquence dans le discours, affinité entre les éléments et degré de transparence et de figement. Les deux premiers paramètres servent à faire le tri entre les collocations et les combinaisons libres; le troisième, à faire le tri entre les collocations et les locutions.

En ce qui concerne le corpus, il est possible, comme le montre le dictionnaire des collocations adverbiales de Kozłowska (1991) de se limiter à un corpus textuel informatisé<sup>7</sup>. Cependant, nous sommes d'avis qu'il est préférable de combiner la consultation d'un corpus textuel avec celle d'un corpus dictionnaire. L'exploitation d'un corpus textuel à l'aide d'un concordancier permet de relever un grand nombre de combinaisons et d'en mesurer la fréquence dans le discours. Elle permet aussi de repérer des combinaisons récentes. Le corpus dictionnaire, pour sa part, donne également un grand nombre de combinaisons; toutefois, celles-ci sont diverses et éparpillées sous différentes entrées et les collocations sont parfois difficiles à trouver. L'utilisation, idéalement combinée, des deux corpus offre un outil valable pour la délimitation des collocations.

**Sélection des collocations et rejet des combinaisons libres**

Pour juger du caractère prévisible et conventionnel des combinaisons et donc retenir les collocations et rejeter les combinaisons libres, il faut comparer les occurrences présentes dans les deux corpus. La répétition d'une même combinaison dans des sources différentes en confirme, à notre avis, le caractère conventionnel et révèle l'affinité qui unit les deux éléments. Selon nous, une combinaison doit être classée parmi les collocations à partir du moment où elle figure dans deux sources différentes, c'est-à-dire :

- soit dans le corpus textuel et dans une source dictionnaire (sous le mot-base ou sous le collocatif);
- soit à plusieurs reprises dans le corpus textuel, mais à condition qu'elle se trouve sous des plumes différentes;
- soit dans des dictionnaires publiés par des éditeurs différents (ex. : *Lexis* et *Nouveau Petit Robert*), sous le mot-base ou sous le collocatif.

La première situation (présence dans le corpus textuel et dans une source dictionnaire) est de loin la plus satisfaisante puisqu'elle confirme la collocation dans le discours et dans la langue. La sélection d'un grand nombre de collocations résulte de l'application de ce critère. Le recours à un corpus textuel permet, entre autres, de repérer des collocations non données par les dictionnaires sous le mot-base, c'est-à-dire là où il est logique de les chercher en premier. Ainsi la combinaison *trébucher sur un mot* est donnée par le corpus textuel mais par aucun des dictionnaires à l'entrée *mot*. En fait, cette collocation ne se trouve que dans le *Lexis* à l'entrée *trébucher*. Par ailleurs, la consultation des deux corpus permet de rejeter des combinaisons telles que *mot expressif* ou *un mot arrive* — comme dans « *les mots pour le dire arrivent aisément* » —, données par le *Grand Robert* mais absentes des autres sources, et des combinaisons telles que *un mot modeste*, *agiter un mot* relevées dans le corpus textuel mais non dans les dictionnaires. La comparaison de sources dictionnaires différentes peut, dans certains cas, suffire pour confirmer le caractère conventionnel d'une combinaison. Ainsi, le repérage de la collocation *croissance exponentielle* peut se faire à partir du *Grand Robert* à *croissance* et du *Petit Larousse* à *exponentiel*. Enfin, le corpus textuel peut révéler la fréquence de certaines combinaisons non encore répertoriées par les dictionnaires. Ainsi, la combinaison *renforcer des liens* concurrence, en fréquence, la collocation *resserrer des liens* dans les textes du *Monde* puisqu'on la trouve, sous la forme verbale ou sous la forme nominale (*renforcement des liens*) dans 40 % des cas. Le corpus textuel nous met donc sur la piste d'une nouvelle collocation dont la fréquence d'emploi indique qu'elle doit être retenue<sup>8</sup>.

Pour évaluer le degré d'affinité entre les deux termes des combinaisons et décider du rejet ou de la rétention de ces dernières, nous avons parfois

recours aussi, au besoin, à la traduction et à des dictionnaires bilingues. Le recours à la traduction permet en effet de « peser » le caractère conventionnel d'une combinaison dans le cas où l'on hésite à la retenir. Comme le signale Hausmann (1989, p. 1013), « l'idiosyncrasie de la collocation ne se relève [sic = révèle] définitivement que dans l'optique d'une autre langue qui combine, pour exprimer le même fait, des mots différents. » Ainsi, en français, étant donné la banalité du verbe *faire* qui peut entrer dans un grand nombre de combinaisons du type verbe + nom, on peut hésiter sur celles qui constituent des collocations. Le recours à ce qui se dit dans une autre langue — en l'occurrence l'anglais — aide à décider de leur rétention ou de leur rejet. Par exemple, la traduction fait ressortir le caractère contraignant et conventionnel du lien qui unit en français le verbe *faire* à des noms-bases tels que *promenade*, *voyage*, *récit* et indique que ces combinaisons sont bel et bien des collocations<sup>9</sup>. En effet, avec les mots-bases donnés, l'anglais emploie des verbes différents : *to take (a walk, a trip)*, *to tell (a story)*.

### **Sélection des collocations et rejet des locutions**

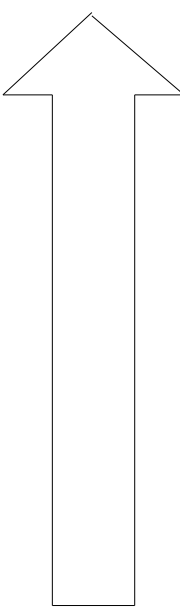
Pour mesurer le degré de transparence et de figement d'une combinaison repérée dans les corpus, nous avons défini deux critères qui vont souvent de pair : possibilité de décomposition et possibilité de substitution.

Parmi les combinaisons données par le corpus, celles qui sont indécomposables, c'est-à-dire impossibles à manipuler grammaticalement, sont évidemment rejetées puisque ce sont des locutions. Le figement de ces combinaisons se manifeste aussi, en général, par l'impossibilité de substituer un élément à un autre. C'est le cas, par exemple, de *se payer de mots*, *trancher le mot*. Pour les combinaisons situées dans la zone floue qui sépare les locutions des collocations, celles qui sont décomposables et formées d'éléments remplaçables par d'autres sont retenues, même si elles ne sont pas totalement transparentes. Par exemple, les combinaisons *souffler un mot* et *sabler le champagne* sont des collocations. Comme le signalent Thoiron et Béjoint (1989, p. 664) à propos de *sabler le champagne*, le sens global de ces combinaisons peut « être partiellement dérivé de la somme des sens des éléments » (*sabler* signifiait autrefois, selon *Le Nouveau Petit Robert*, dans son sens figuré *avalier d'un trait*). Puisque chacun des éléments garde son sens propre ou figuré, on peut remplacer le collocatif par un synonyme (ex. : *dire un mot*, *boire le champagne*). La substitution possible confirme que nous avons affaire à des collocations et non à des locutions figées.

À titre d'illustration, nous donnons ci-dessous quelques-unes des combinaisons du type verbe + nom dans lesquelles entre le nom-base *mot*, combinaisons relevées dans notre corpus textuel et dictionnaire. Nous les avons placées dans le continuum qui va des combinaisons libres aux locutions figées.



Les combinaisons en caractères gras constituent, d'après nos critères, des collocations.

locutions		<i>avoir le dernier ~, prendre au ~ trancher le ~, se payer de ~ s</i>
<b>collocations</b>		<i><b>ne pas souffler ~, glisser/souffler/toucher un ~</b> <b>manger ses ~ s</b> <b>estropier un ~</b></i>
combinaisons libres		<i>écrire, tracer un ~; biffer, raturer un ~; articuler, dire, lire, épeler un ~; créer, forger un ~</i> <i><b>chercher, choisir, définir, trouver un ~</b></i> <i>bannir, changer, citer, comprendre, traduire agiter, aimer, réhabiliter un ~</i>

### Conclusion

En conclusion, il faut reconnaître que les paramètres et les critères que nous avons établis pour délimiter les collocations sont relativement simples; mais ils sont fonction de l'objectif que nous nous sommes fixé. Nous entendons en effet élaborer un ouvrage d'encodage de consultation rapide et aisée, destiné à des rédacteurs et à des traducteurs qui désirent retrouver facilement un mot qui leur échappe et qu'il serait long de repérer dans les dictionnaires usuels. Il est donc évident que les paramètres choisis ne conviendraient pas nécessairement à des recherches ayant une finalité autre que la nôtre. Dans notre cas, il fallait que nos paramètres nous permettent de repérer un large éventail de combinaisons tout en éliminant le caractère subjectif possible de la sélection. Il faut reconnaître que ces paramètres n'éliminent pas complètement l'indécision et l'arbitraire. Certaines des collocations retenues seront jugées inutiles car évidentes et, inversement, on pourra nous reprocher notre manque d'exhaustivité puisque certaines combinaisons attendues par l'utilisateur, mais ne correspondant pas à nos paramètres, n'auront pas été relevées. C'est le lot de tout recueil de ce genre<sup>10</sup>.

## Notes

- <sup>1</sup> Il existe aussi quelques dictionnaires bilingues de collocations du français (pour le français et l'allemand, celui de Reum et celui d'Ilgenfritz), mais leur optique est traductionnelle et leur conception orientée vers la langue d'arrivée.
- <sup>2</sup> Le schéma suivant, inspiré du classement de Hausmann (1989), décrit les différentes structures des collocations selon les différents mots-bases (indiqués en caractères gras):
1. **nom** + adjectif (*profond **sommeil***)
  2. nom + **nom** (*les phases du **sommeil***)
  3. **nom** + verbe (*le **sommeil** gagne quelqu'un*)
  4. verbe + **nom** (*vaincre le **sommeil***)
  5. **verbe** + adverbe (***dormir** profondément/à poings fermés*)
  6. **adjectif** + adverbe (*à moitié **endormi***)
- <sup>3</sup> Mel'čuk (1995, p. 46) appelle aussi ce type de combinaison « un semi-phrasème » et donne l'exemple de *piquer un somme*, dans lequel « le signifié de *piquer un somme* (faire un somme) inclut le sens de *somme* (sommeil court et léger), mais pas celui de *piquer* (entamer avec une pointe) ».
- <sup>4</sup> Mentionnons d'ailleurs que, même dans des communautés linguistiques qui parlent essentiellement la même langue, il existe des variantes régionales. Ainsi, *passer un examen* veut le plus souvent dire, en français du Canada, « réussir à un examen », et, en français européen, « subir les épreuves d'un examen ».
- <sup>5</sup> C'est pourquoi dans les « textes à trous », qui évaluent la connaissance d'une langue, il est souvent demandé de trouver le collocatif d'un mot-base.
- <sup>6</sup> Hausmann, qui avait défini la collocation comme la combinaison probable et usuelle de deux mots (1979, p. 190), précise dans un autre article (1989, p. 1010) que la collocation, bien que transparente, est « imprédictible » et que, chaque langue ayant des combinaisons qui lui sont propres, l'apprenant étranger doit les apprendre pour les connaître.
- <sup>7</sup> Kozłowska n'indique pas le nombre de mots du corpus utilisé, mais mentionne que c'est la seule source d'extraction des collocations. Quant à notre corpus textuel informatisé, il se compose de quelque trois millions de mots de textes journalistiques (*L'Actualité/Québec* et *le Monde*). Notre corpus dictionnaire se compose des dictionnaires de collocations existants (ceux de Lacroix et de Reum) et des principaux dictionnaires de langue : *Trésor de la langue française*, *Grand Robert*, *Nouveau Petit Robert* et *Lexis*; nous consultons aussi, à l'occasion, des dictionnaires encyclopédiques (*Petit Larousse* et *Grand Larousse encyclopédique*) et des dictionnaires bilingues (*Robert et Collins* et *Oxford-Hachette*).
- <sup>8</sup> Il faut ajouter que la comparaison des sources dictionnaires et du corpus textuel permet aussi d'éliminer les collocations désuètes ou rares parfois données par les premières. Les collocations, comme tout autre élément du langage, évoluent avec le temps.
- <sup>9</sup> Même si, en fait, les combinaisons du type *faire* + nom donnent souvent lieu à des locutions verbales (ex. : *faire peur*, *faire honte*) et qu'on peut contester leur statut

de collocations, il nous paraît essentiel qu'elles figurent dans un dictionnaire de combinaisons lexicales.

- <sup>10</sup> Ainsi, Benson (1989) affirme qu'un certain dictionnaire des collocations anglaises "is [...] flawed by the inclusion of many free combinations", en signalant qu'il contient, pour le mot-base *dictionary*, des collocatifs tels que "*excellent, good, useful, useless*", mais omet de vraies collocations telles que "*abridged, bilingual, biographical* [...] *dictionary*."

## Bibliographie

- Béjoint, H. et P. Thoiron. 1993. « Macrostructure et microstructure dans un dictionnaire de collocations en langue de spécialité. » *Terminologie et traduction*, pp. 513–522.
- Benson, M. 1989. « The structure of the Collocational dictionary. » *International Journal of Lexicography*, 2, pp. 1–14.
- Benson, M., E. Benson et R. Ison. 1997. *The BBI Combinatory Dictionary of English: A Guide to Word Combinations*. Revised edition. Amsterdam/Philadelphia, John Benjamins Publishing Company.
- Cohen, B. 1986. *Lexique de cooccurrents — Bourse, conjoncture économique*. Montréal, Linguatex.
- Hausmann, F.J. 1979. « Un dictionnaire des collocations est-il possible? » *Travaux de linguistique et de littérature*, 17, pp. 187–195.
- Hausmann, F.J. 1989. « Le dictionnaire de collocations. » *Dictionnaires. Dictionnaires. Wörterbücher. International Encyclopedia of Lexicography*, Hausmann, et al. (réd.). Berlin, De Gruyter, pp. 1010–1019.
- Heid, U. et G. Freibott. 1991. « Collocations dans une base de données terminologique et lexicale. » *META*, 36, pp. 71–91.
- Ilgenfritz, P. et coll. 1989. *Langenscheidts Kontextwörterbuch Französisch — Deutsch*. Berlin, Langenscheidt.
- Kozłowska, C.D. 1991. *English Adverbial Collocations*. Warsaw, PWN.
- Lacroix, U. 1969. *Les mots et les idées : Dictionnaire des termes cadrant avec les idées*. Édition nouvelle, revue et corrigée [1931]. Paris, Nathan.
- Lainé, C. 1993. *Vocabulaire combinatoire de la CFAO mécanique*. Secrétariat d'État du Canada/Rint.
- Mel'čuk, I., A. Clas et A. Polguère. 1995. *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve, Duculot.
- Reum, A. 1953. *Petit dictionnaire de style à l'usage des Allemands*. Leipzig, VEB Bibliographisches Institut.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford, Oxford University Press.
- Thoiron, P. et H. Béjoint. 1989. « Pour un index évolutif et cumulatif des cooccurrents en langue techno-scientifique sectorielle. » *META*, 34, pp. 661–671.
- Zareba, L. 1977. « Sur la méthode d'élaboration des dictionnaires phraséologiques. » *Terminologies 76 : Colloque international, Paris, La Défense, 15–18 juin 1976*. Paris, Maison du dictionnaire, pp. II25–II35.