

Franz Josef Hausmann

Université d'Erlangen-Nürnberg

Peter Blumenthal

Université de Cologne

Présentation : collocations, corpus, dictionnaires

I. QU'EST-CE QU'UNE COLLOCATION ?

Ce cahier de *Langue française* désire promouvoir l'étude d'un type de signe linguistique qui est connu sous le nom de *collocation*, terme utilisé en linguistique de deux façons.

La première acception, quantitative, relève de la description statistique de corpus. Elle désigne l'apparition soit fréquente, soit statistiquement significative (compte tenu de fréquences pondérées) de deux unités lexicales données dans un contexte plus ou moins étroit. Un dictionnaire qui réunit des cooccurrences fréquentes, mais souvent peu pertinentes pour qui veut se renseigner sur la combinatoire spécifique des mots, existe pour l'anglais et s'appelle *A Dictionary of English Collocations* (Kjellmer 1994). L'article *hotel* de ce dictionnaire donne, entre autres, les cooccurrences suivantes : *hotel at/hotel room/hotel suite/a hotel/either hotel/Hilton Hotel/my hotel/of hotel*.

La seconde acception, essentiellement qualitative, de « collocation » peut être définie par « cooccurrence lexicale restreinte ». Comme l'explique très bien Steinlin (2003 : 3), l'intérêt qu'on trouve à ce type de cooccurrences tient davantage aux contraintes qu'elles font peser sur l'utilisation du lexique qu'à leur fréquence, car il s'agit d'un savoir lexical, d'une propriété du lexique. Et le même auteur de donner comme exemples : on ne *pousse* pas un son, on l'*émet* ; on n'*émet* pas un hurlement, on le *pousse* ; qqch *inflige* une déception, mais *inspire* une crainte. C'est cette seconde acception qui se trouvera au centre de la majorité des contributions présentées ici ; nous verrons par ailleurs qu'elle ne s'oppose pas radicalement à la première (*cf. infra* 4.). Un dictionnaire qui réunit les collocations dans leur seconde acception existe pour l'anglais et s'appelle *Oxford Collocations Dictionary for Students of English* (2002). L'article *hotel* de ce dictionnaire fournit, parmi d'autres, les collocations suivantes : VERB + NOUN *stay at/in | book in at/intو, check in at/intو, check out of | own | manage, run.*

Regardant de plus près l'unité syntagmatique en question, nous pouvons décrire la **collocation** (exemples : *passer un examen, un célibataire endurci, grièvement blessé, une bouffée de colère*) comme la combinaison phraséologique (codée en langue) d'une **base** (*examen, célibataire, blessé, colère*) et d'un **collocatif** (*passer, endurci, grièvement, bouffée*). La base est un mot (plus précisément l'acception d'un mot, appelée aussi « lexie ») que le locuteur choisit librement parce qu'il est définissable, traduisible et apprennable sans le collocatif. Le collocatif est un mot (ou l'acception d'un mot) que le locuteur sélectionne en fonction de la base parce qu'il n'est pas définissable, traduisible ou apprennable sans la base.

Étant une unité polylexicale codée en langue, la collocation se distingue premièrement de la **combinaison libre**, non combinée en langue, qu'elle soit banale (*oublier un examen, un célibataire rancunier*) ou originale (*lait noir, la route se rabougrit, le jour se fissure*). Mais elle se distingue aussi d'autres unités polylexicales codées en langue, notamment de la **locution idiomatique** ou **figurée** (*prendre la mouche, casser les pieds à qqn*) qui est codée en langue mais n'a pas de base (ni, par conséquent, de collocatif), puisqu'elle signifie et est sélectionnée par le locuteur en bloc (le locuteur ne parlant pas de mouche ou de pieds). Elle semble également se distinguer des mots composés polylexicaux terminologiques du type *arrêt cardiaque, équipement portuaire* ou *feu rouge* qu'on a proposé d'appeler **phraséotermes** (G. Gréciano, d'après González Rey 2002 : 92) et qui, malgré leur transparence, sont codés, sans pour autant présenter une structure base-collocatif (le locuteur ne choisit pas *cardiaque* en fonction d'*arrêt*). Ces distinctions se résument de la façon suivante : la **phraséologie** (unité : phrasème) se compose de **phraséotermes** (*arrêt cardiaque/équipement portuaire/feu rouge*), de **collocations** (*administrer une gifle/un célibataire endurci/grièvement blessé*) et de **locutions** (*casser les pieds à qqn/prendre la mouche/un cordon bleu*).

La hiérarchie base-collocatif (qu'on peut appeler **l'orientation de la collocation**), hiérarchie qui fait qu'en formulant, le locuteur progresse en partant de la base pour arriver au collocatif et non inversement, est une réalité psychologique observable en discours, notamment de la part d'un locuteur étranger qui perçoit l'apprentissage des bases comme celui d'équations simples (angl. *exam* = fr. *examen*), alors que l'équation du collocatif varie en fonction de la base (pour dire *passer un examen*, l'anglais dit *take* < prendre > et le collocatif anglais *pass* veut dire en français < réussir >). Cette variation pouvant prendre des proportions considérables, l'étranger même très avancé dans la maîtrise de la langue peut se retrouver dans une véritable situation de détresse d'encodage où, connaissant la base, il cherche désespérément le collocatif, l'inverse étant improbable : on ne cherche pas désespérément la base *célibataire* à partir du collocatif *endurci*. C'est une détresse dont même le locuteur natif n'est pas toujours à l'abri, notamment quand il s'agit de variation synonymique du collocatif. Tout le monde n'a pas toujours à sa disposition les différentes façons d'exprimer la réussite ou l'échec à l'examen : *réussir, réussir à, être reçu à, être admis à/échouer à, être refusé à, recalé à, collé à, blackboulé à, manquer, rater, etc.*

La collocation est donc une entité à prendre en compte en linguistique théorique et descriptive, en psycholinguistique, en linguistique cognitive, en linguistique contrastive, en lexicographie et dans la pédagogie des langues,

maternelles comme étrangères. Prenons le cas, particulièrement significatif, de la **lexicographie monolingue**. À l'article du collocatif (par exemple *administrer*), le bon dictionnaire général signalera que la définition (en l'occurrence, « donner ») est restreinte à la collocation du mot (ou de l'acception) avec un nombre réduit de bases (en l'occurrence, *gifle*). La mention de la collocation constitue ainsi le prolongement naturel de la définition et est indispensable pour le traitement lexicographique. Tel n'est pas le cas à l'article de la base (*gifle*) où le collocatif (*administrer*) est sans importance pour la définition. Il est en revanche indispensable si le dictionnaire veut également remplir une fonction d'encodage et apprendre à l'utilisateur à contextualiser correctement la base (*gifle*). On peut dire que la collocation a une fonction sémasiologique à l'article du collocatif et une fonction onomasiologique à l'article de la base. Comme les dictionnaires généraux privilégient par nature la fonction sémasiologique (de décodage) et négligent souvent la fonction onomasiologique (d'encodage), il faut pour cette dernière un dictionnaire spécialisé qu'on appellera dictionnaire de collocations et dont la macrostructure sera faite de bases de collocation sur lesquelles il adressera en microstructure la totalité des collocatifs. Le modèle (forcément imparfait) d'un tel dictionnaire de collocations est pour le moment le *Oxford Collocations Dictionary*, cité ci-dessus. Pour le français, quatre projets sont en cours (cf. Mel'äuk & Polguère et Verlinde, Binon & Selva dans ce fascicule, ainsi que Blumenthal 2005 ; Dominique Le Fur, des Dictionnaires Le Robert, travaille également sur le concept de « collocation »).

En linguistique syntagmatique, les deux perspectives (de la base aux collocatifs / du collocatif aux bases) ont leur intérêt. On peut donc prendre comme **point de départ** (anglais *node*) de la description le collocatif et inventorier une classe de bases. Par exemple, le collocatif *appeler* se combine en français avec les bases *taxis* et *ascenseur* (contrairement à l'allemand *rufen* qui va pour *taxis* mais non pour *ascenseur* où on dit *holen*). On peut aussi prendre comme point de départ de la description la base (*taxis*) et faire l'inventaire des collocatifs (*appeler*, *héler*, etc.). Dans les deux cas, on établit ainsi le **profil combinatoire** du mot – ou de l'acception – (Blumenthal 2002, 2005). Et dans les deux cas le profil combinatoire peut se mettre au service de la synonymie, de l'apprentissage systématique du vocabulaire, de la traduction automatique ou d'autres activités. Bosque (2004) est pour l'espagnol un dictionnaire (admirable de richesse) consacré au profil combinatoire des collocatifs (et accessoirement des bases).

Certaines écoles, refusant l'orientation de la collocation, ne connaissent par conséquent que des points de départs (*nodes*) et utilisent le terme anglais *collocate* pour désigner n'importe quel partenaire collocationnel. À la recherche d'un équivalent de *collocate*, on pourra proposer *collocateur* (qu'on prendra soin de distinguer de *collocatif*). De même pourra-t-on proposer pour l'anglais *node* le terme de *noyau*. Dans ce sens, *noyau* et *collocateur* (point de départ de la description et partenaire d'une collocation) seraient à distinguer de *base* et *collocatif*. En aucun cas, on ne pourra accepter que le collocatif soit appelé « base » si on veut éviter la confusion totale et la fin de la communication entre les écoles.

La théorie de la collocation, telle qu'elle est présentée ici, sans être absente de la linguistique française (cf. Cusin-Berche 2002, González-Rey 2003 : 89 sqq, Grossmann & Tutin 2003, Polguère 2003 : 134-142, Steinlin 2003),

reste relativement confidentielle. Cela a de quoi surprendre, étant donné l'universalité du phénomène et sa présence massive dans la lexicographie en France depuis la Renaissance.

2. LEXICOGRAPHIE ET COLLOCATIONS EN FRANCE – PASSÉ ET PRÉSENT

La description et l'étude des collocations a en France une longue tradition qui s'exerce dans l'activité lexicographique spécialisée et générale du latin, dans l'activité lexicographique spécialisée et générale du français et dans le discours métalexicographique des lexicographes.

Le premier dictionnaire français monolingue est en 1571 une sorte de dictionnaire de collocations, *Les Épithetes* de Maurice de La Porte, dont le sous-titre affirme son utilité « pour illustrer toute composition Françoise » (Bierbach 1997 : 59-62, 363-371). Ce dictionnaire, qui relève de la lexicographie spécialisée, s'insère logiquement dans la tradition des dictionnaires rhétoriques collocationnels du latin qui accompagnent, à l'époque de la Renaissance, les grandes réalisations bien connues de Robert Estienne, également riches en exemples de type collocationnel (Lindemann 1994). L'histoire des dictionnaires collocationnels du latin, du français, ainsi que d'autres langues est esquissée dans Hausmann (1989). La place nous manque pour revenir là-dessus. Il faudra, en revanche, répondre rapidement à deux questions peu posées : quelle place occupent les collocations au sein de la lexicographie française générale et comment y ont-elles été nommées ?

Dans la lexicographie française générale, les collocations ont toujours eu une place de premier choix, et ceci grâce à l'Académie française et son refus de citer. Regardons l'article DÉBAT du premier *Dictionnaire de l'Académie française* en 1694 :

DÉBAT. Différend, contestation.

- (1) *Être en débat de quelque chose*
- (2) *mettre en débat*
- (3) *vider un débat*
- (4) *terminer un débat*
- (5) *avoir débat avec quelqu'un*
- (6) *ouvrir un débat*
- (7) *apaiser un débat*
- (8) *débat d'un compte*

On dit prov. À eux le débat/entre eux le débat pour dire qu'on ne se veut point mettre en peine de leur différend.

L'Académie est restée fidèle à sa méthode rhétorique et collocationnelle jusqu'à la 9^e édition en cours. De par son modèle « réaliste », l'Académie s'était opposée au modèle philologique et autoritariste de l'Accademia della Crusca. Or, c'est ce modèle premier, privilégiant la citation, qui est repris à son compte par le plus grand lexicographe du XIX^e siècle, Émile Littré. Par conséquent, le dictionnaire de Littré est relativement pauvre en collocations. Qui cherche la double richesse, citationnelle et collocationnelle, la trouve pour

la première fois dans le « Nouveau Littré » que veut être *Le Grand Robert* et après lui dans le *Trésor de la langue française*.

Comment les lexicographes parlent-ils, depuis la Renaissance, des unités phraséologiques ? Les dictionnaires de collocations latines du XVI^e siècle ne manquaient pas de termes pour les nommer : N + A : Epitheta / Adiuncta Nominibus / N + V : Phrases / Locutiones / Formulae / V + Adv : Adiuncta (verbis) / X et Y : Coniuncta (Hausmann 1982 : 191). L'examen attentif des premiers dictionnaires français généraux monolingues (préfaces, articles concernés, pratique de la catégorisation) révèle une théorie sous-jacente qui, pour floue et implicite qu'elle soit, n'est pas moins existante et surtout concordante (*cf.* aussi Leroy-Turcan 1997 et Grimaldi 1997). Certains emplois de *phrase* semblent indiquer que ce terme vise davantage les collocations que les locutions. L'Académie place deux fois le terme à côté des *épithètes* (Quemada 1997 : 30, 39) qui sont des collocatifs, et ceci apparemment pour signifier les collocations autres qu'adjectivales. Furetière (1690) nomme *phrase* l'unité collocationnelle « Reparations locatives » (sous **locative**). Dans ses *Factums*, il oppose aux définitions quatre composants syntagmatiques de l'article (Hausmann 1982 : 197 *sq*) : épithètes, phrases (propres), phrases figurées et proverbes – ce qui confirme que *phrase* (propre) < collocation > est utilisé à la fin du XVII^e siècle.

Pour nommer des unités phraséologiques, le terme de *phrase* est en recul au moins depuis le XIX^e siècle (*cf.* TLF sous **phrase I, Rem.**, et Grimaldi 1997). Il a cédé au profit de *phrase*, terme de syntaxe, ou a subi une détérioration pour signifier < lieu commun >, < parole creuse >. Par quoi le terme de *phrase* a-t-il été remplacé pour signifier < collocation > ? Le *Dictionnaire de l'Académie* de 1835 (préface par A.-F. Villemain) l'utilise encore, à côté de *locution* et *idiotisme* (Quemada 1997 : 346, 362). Après le texte de Villemain, la notion est diluée dans la vaste catégorie des *exemples* (*op. cit.* : 404, 446).

En conclusion, on peut dire que, dans les premiers dictionnaires, grâce au terme de *phrase*, les collocations sont plus clairement distinguées que plus tard où elles seront noyées dans la catégorie peu spécifique des exemples. Pour l'époque moderne, on en arrive à ce constat sidérant : les dictionnaires sont bourrés de collocations, mais les lexicographes ne ressentent pas – ou peu – le besoin de les nommer.

3. DISCOURS LINGUISTIQUE ET COLLOCATIONS

La linguistique du XIX^e siècle ignore les collocations. Lorsque Hermann Paul (Paul 1970 [1880] : § 73) parle de « groupes de mots » (*Wortgruppen*), de « combinaisons » (*Verbindungen*) et de « tours » (*Wendungen*), ses exemples montrent qu'il s'agit uniquement de locutions idiomatiques ou figurées. Et Saussure ? Quand il en vient à parler des syntagmes non libres, appartenant à la langue, il emploie deux termes : *expressions* et *locutions toutes faites*. Pour les dernières, il donne les exemples *à quoi bon ?* et *allons donc !* (L'édition du CLG par Engler ajoute *s'il vous plaît*, appelé « locution » ; Engler 1968 : 284). *Expression* est chez Saussure une catégorie floue qui comprend les locutions toutes faites et un autre type auquel il ne donne pas de nom spécifique mais que les exemples

(prendre la mouche, forcer la main à qqn, rompre une lance, avoir mal à, à force de, que vous en semble ?, pas n'est besoin de) permettent d'identifier également comme des locutions (Saussure 1965 : 172). Faute d'exemples concrets, tout ce que dit Saussure au sujet des syntagmes, groupes de mots, combinaisons, groupes d'associations, groupements et groupements syntagmatiques (Engler 1968 : 283-294) reste trop vague pour qu'on lise ces termes dans le sens des collocations.

Tout autre est la situation pour Charles Bally, disciple de Saussure certes, mais ayant au moment du *Cours de linguistique générale* près de 45 ans et une solide expérience de l'enseignement du français langue étrangère. C'est cette vue contrastive sur la langue qui lui fait découvrir qu'il y a deux types de « locutions phraséologiques » [= phrasèmes] : les « unités phraséologiques » [= locutions idiomatiques] et les « séries phraséologiques » (Bally 1951, 1 [1909] : 68) [= collocations] qu'il appelle aussi « groupements usuels » (*op. cit.* : 70) et pour lesquels il donne de nombreux exemples, entre autres : *grîvement blessé, gravement malade, désirer ardemment, aimer éperdument, intimement lié, refuser catégoriquement, chaleur tropicale, chaleur torride*. (Cf. aussi Bally 1951, 2 : 43-53, avec des exercices sur des collocatifs verbaux). Voici donc la terminologie de Bally : la locution phraséologique [phrasème] est ou bien une série phraséologique / un groupement usuel (séries d'intensité/séries ou périphrases verbales) [collocation] ou bien une unité phraséologique [locution].

Cette analyse n'a laissé aucune trace immédiate dans la linguistique de son temps. Trois raisons semblent être à l'origine de cette réception ratée. D'abord, le genre de publication choisi, un traité de stylistique pratique pour étrangers. Cette linguistique appliquée avant la lettre passait pour artisanale et n'était pas prise au sérieux par les théoriciens. Bally, lui-même, ne semble pas lui attribuer une valeur extraordinaire, puisque, dans sa célèbre *Linguistique générale et linguistique française* de 1932 (Bally 1965), il n'en parle pas. Ensuite, la terminologie choisie, trop périphrastique, insuffisamment identifiante et individualisante. Pour finir, l'éloignement de ce sujet des grands courants de la linguistique de la première moitié du XX^e siècle.

C'est ainsi que dans ce petit livre, destiné surtout à l'enseignement secondaire (Bally 1909, 1 : VIII), les chercheurs, dont Bally avait dit : « Peut-être les chercheurs y trouveront-ils aussi leur compte » (*op. cit.* : IX), ont mis très longtemps à découvrir le pactole qui y était enseveli. Les *lexies complexes* de Pottier renferment des unités collocationnelles, sans pour autant clairement les analyser (Pottier 1964 : § 290 *sqq*). Les *synapsies* de Benveniste sont seulement un certain type de collocations, de forme N + prép + N (Benveniste 1966). Quant à André Martinet, il ne parle pas des collocations (oubli réparé par Builles 1998 : 166).

Lecteur de Bally et de Marouzeau (qui avait dit : « On peut constater que la langue est faite de formules autant que de mots » (Marouzeau 1959 [1941] : 145)), le grand linguiste Eugenio Coseriu reprend dans une conférence retentissante de 1964 lors du Premier Colloque International de Linguistique Appliquée (*Actes* 1966) des exemples du type *un gros chagrin, une grande douleur, de graves soucis* qu'il appelle « clichés lexicaux » et « syntagmes lexicaux » pour les situer dans la norme, c'est-à-dire dans la langue (Coseriu 1970 : 42). Malheureusement, il les oublie lorsqu'il fait quelques pages plus tôt l'inventaire du « discours répété » (*op. cit.* : 27-32, oubli réparé par Thun (1978 : 65

et 1995 : 670) qui compte ces « analyses préférées » (collocations) parmi les combinaisons figées de mots [« fixierte Wortgefüge »]). Dans la discussion intervienint des noms aussi prestigieux que Baldinger, Culoli, Gougenheim, Isaäenko, Larochette, Ullmann, mais seul K. V. Sinclair, médiéviste manifestement au courant des nécessités de l'enseignement scolaire, relève cette partie de la conférence et insiste sur l'importance fondamentale des « **high-frequency collocations** », relayé par P. J. Wexler (*cf.* Coseriu 1970 : 92, 103).

Il a donc fallu que l'école britannique découvre le rôle du contexte et invente le terme de *collocation* (Williams 2003) et que ce terme, par le truchement de Halliday *et al.* (1964 : 33 *sqq*), parvienne à Paris et soit repris par certains (Mitterand 1965 : 79 ; Galisson 1971 [1965] : 6 ; Wagner 1967 : 140 ; Quemada 1967 : 508 *sqq* ; Imbs 1971 : XXXII), sans pour autant s'imposer. Car on lui a longtemps préféré le terme de « *syntagme* », pourtant trop ambigu (libre ? codé ?) pour imposer un concept nouveau.

Et puis il restait à décrire la structure interne des syntagmes/collocations. Il revenait en effet à un chercheur à orientation pédagogique, Robert Galisson, d'aller plus loin que Bally en analysant les collocations en un terme-noyau [= la base] et un terme-satellite qu'il appelle collocatif (Galisson 1971 [1965] : 14). Pour Galisson, qui s'inspire des travaux britanniques, le verbe *fumer* est un collocatif limité, puisqu'il n'admet guère que *cigarette*, *pipe* et quelques autres comme termes-noyaux. Galisson s'était mis à faire un petit dictionnaire de collocations où on trouvait *allumer*, *frotter* et *éteindre* dans l'orbite d'*allumette*, dictionnaire circulant au BELC depuis 1965, mais accessible à la communauté des chercheurs seulement en 1971. C'est de Bally et de Galisson que s'inspire Hausmann (1979 ; *cf.* aussi 1996, 1997, 1999, 2004).

Mais parallèlement s'est développée une autre réflexion qui, à ses débuts, ignore Bally et ne connaît comme modèle de collocations que les dictionnaires d'Albrecht Reum (1953, 1961), c'est celle d'Igor Mel'äuk. En 1962, dans les collines pierreuses du Kazakhstan, ce jeune chercheur russe (né en 1932, ayant étudié l'espagnol et le français, travaillant depuis 1956 à la traduction automatique), tout en portant le barda d'un géologue, réfléchit au problème de l'automatisation de la traduction des collocations anglaises (*heavy rain*, *to launch an attack*) et à la géniale idée des fonctions lexicales, publiée pour la première fois en 1965 avec A. K. Îolkovskij (Mel'äuk *et al.* 1984 : 3 et communication personnelle du 27 mars 2005). Cette recherche devait mener au *Dictionnaire explicatif et combinatoire du français contemporain* (Mel'äuk *et al.* 1984-1999) dont l'analyse collocationnelle est certainement la plus poussée en théorie et la plus universellement productive.

Les dictionnaires de terminologie linguistique sont avant 2000 d'une affligeante pauvreté en matière de phraséologie. Pour que cela change, il faut attendre le dictionnaire de Charaudeau & Maingueneau (2002) et son article « cooccurrence » par la regrettée F. Cusin-Berche. Les introductions à la lexicologie avant 2000 (à l'exception de Niklas-Salminen 1997 : 112 *sq*) ignorent également les collocations. La notion de « collocation » n'est sortie de la confidentialité en France que dans les années 1990 (Liang 1991, Béjoint & Thoiron 1992, Descamps 1994, Candel 1995, Tutin 1995). On peut constater un début de prise de conscience à l'occasion du colloque international de Saint-

Cloud : « La locution : entre lexique, syntaxe et pragmatique. Identification en corpus, traitement, apprentissage » (Martins-Baltar 1995, 1997 ; Fiala, Lafon & Piguet 1997). Mais la claire distinction entre collocation et locution est souvent encore refusée. *Le Petit Robert* définit depuis 1993 la « phraséologie » dans le corps du dictionnaire comme l'« ensemble des expressions, locutions, collocations et phrases codées dans la langue générale ». Mais la définition de la collocation dans le même dictionnaire est bien maigre : « Position (d'un objet, d'un élément) par rapport à d'autres ; proximité dans une chaîne. →cooccurrence. » (inchangée en 2003). Pour une première initiation au phénomène de la collocation, on peut s'adresser maintenant à González Rey (2003 : 89 *sqq*), Grossmann & Tutin (2003), Polguère (2003 : 134-142), Steinlin (2003).

4. COLLOCATIONS, CORPUS, DICTIONNAIRES

« Dans la règle, nous ne parlons pas par signes isolés, mais par groupes de signes, par masses organisées qui sont elles-mêmes des signes » (Saussure 1965 : 177). Si on appelle *phraséologie*, comme nous le proposons, l'ensemble des unités polysegmentales codées en langue, que ce soit dans le système ou dans la norme, il faudra reconnaître que la collocation est un certain type de *phrasème* à côté de bien d'autres (locutions figurées, phraséotermes, énoncés codés, proverbes, citations célèbres, jusques et y compris les phrasèmes grammaticaux, polymorphématiques ou morpholexématiques, transparents ou opaques). Et bien qu'il n'y ait « pas de limite tranchée entre le fait de langue, marque de l'usage collectif, et le fait de parole, qui dépend de la liberté individuelle » (Saussure 1965 : 173), il est de notre devoir de linguiste d'identifier dans les corpus et de décrire dans les dictionnaires ces mêmes collocations. Voilà pourquoi les auteurs de ces lignes ont organisé les 1^{er} et 2 juillet 2005 le Colloque de Cologne sur les Collocations¹ dont les sept communications, connues auparavant dans leur version provisoire, figurent ici dans une version qui a pu profiter des débats menés en toute convivialité².

Plus haut, nous avons évoqué l'existence de deux conceptions dans les recherches actuelles sur la collocation, l'une purement quantitative, l'autre plutôt qualitative. Il convient toutefois de ne pas surestimer l'intérêt méthodologique de cette opposition. Car pour le moment, ces conceptions partagent toutes deux l'approche descriptive d'un phénomène dont l'explication scienti-

1. Nous remercions l'Université de Cologne pour son hospitalité, La Deutsche Forschungsgemeinschaft, ainsi que le Bureau pour la collaboration universitaire, Service culturel de l'Ambassade de France à Berlin (CCCL Düsseldorf) pour le soutien financier. Le comité d'organisation était composé de Sandra Lhafi, Sascha Diwersy et Ludwig Fesenmeier.

2. Ont présidé les séances et/ou participé aux débats : Danielle Leeman, Dominique Le Fur, Ulrich Heid et Franz Schneider. Qu'ils en soient remerciés. Gaston Gross et Igor Mel'äuk, vivement sollicités, n'ont pas pu venir pour cause d'empêchement. La contribution de F. J. Hausmann, intitulée *Lexicographie française et phraséologie* et dédiée « à la mémoire de Charles Bally (1865-1947), métacollocationniste avant la lettre, et à Igor Mel'äuk (né en 1932), déchiseur d'horizons en linguistique », n'a pas pu être insérée, étant trop longue. Toutefois, certaines de ses idées-force se retrouvent dans cette présentation.

fique (pour quelles raisons profondes les mots se combinent-ils d'une certaine façon ?) en est encore à ses débuts. Avouons incidemment que les auteurs des articles réunis ici s'estimerait heureux s'il s'avérait qu'ils avaient pu contribuer, ne serait-ce que dans de bien modestes proportions, à faire évoluer cette situation : le temps semble mûr pour passer des classifications descriptives aux tentatives de mieux comprendre le pourquoi et le comment des faits dûment étiquetés et catalogués.

Un second facteur plaident en faveur d'une vision plutôt complémentaire qu'antagoniste des deux approches tient au fait que tous les chercheurs en ce domaine, quel que soit par ailleurs leur credo théorique, travaillent aujourd'hui sur de vastes corpus informatisés, souvent catégorisés et lemmatisés ; le dépouillement automatique de ceux-ci s'impose presque irrésistiblement et permet au lexicologue, même s'il n'est pas un aficionado de méthodes quantitatives, soit de se faire une première idée de l'environnement du mot de base, soit de contrôler empiriquement les résultats d'un travail théorique sur les contraintes combinatoires.

En effet, corpus numérisés et outils lexicométriques ont profondément révolutionné la recherche linguistique et, plus particulièrement, lexicographique, en décuplant les possibilités d'observations. Vu l'impact de cette révolution, il ne paraît pas exagéré d'affirmer que l'avant et l'après s'opposent autant que l'œil nu s'oppose aux jumelles, télescope et microscope réunis. Longtemps pionnière dans le domaine lexicographique de par l'immense entreprise du *Trésor de la langue française*, la France a un rang à tenir, désormais menacé par la Grande-Bretagne, où l'amélioration annuelle des dictionnaires pour étudiants avancés n'est plus imaginable sans le soutien des grands corpus accessibles à tous.

En matière de dictionnaires spécialisés, l'*Oxford Collocations Dictionary*, cité à plusieurs reprises, représente la dernière preuve du haut niveau lexicographique généré par l'exploitation électronique des corpus. Pourquoi ne disposons-nous pas d'un outil comparable en français, capable de nous fournir pour chaque mot de base une liste de mots-satellites présélectionnés par la langue française – de mots justes, disait-on autrefois – pour exprimer une idée, articulée tout à fait différemment dans d'autres langues ?

Dans ces conditions, l'objectif général du présent projet est de réunir un certain nombre de tentatives en matière de description syntagmatique des mots pour créer une synergie profitable à tous, au niveau de la recherche pure comme à celui de la recherche appliquée, celle qui mène à la confection de dictionnaires. Tous les contributeurs ne comprennent pas le terme de « collocation », point de mire de nos réflexions, exactement de la même façon, mais tous sont convaincus de l'extrême utilité des corpus et des concordances pour donner une description réaliste de la dimension combinatoire des mots et de la langue.

Les contributeurs se sont efforcés de faire apparaître la diversité des buts et des méthodes qui caractérise cette vaste thématique, mais de garantir aussi une certaine comparabilité des approches – d'où la focalisation sur les mêmes lexèmes ou champs sémantiques dans plusieurs articles (deux exemples : *débat* et ses synonymes, les noms d'affects).

Peter Blumenthal soumet à la discussion l'hypothèse que rien dans la combinatoire stéréotypée du mot ne relève du hasard des rencontres fortuites ; dans cette perspective, tout dépend en dernière instance des structures sémantiques ou cognitives profondes que reflète le contenu du mot. À l'intérieur du contenu, il s'agit de dégager la conceptualisation (d'une réalité) à laquelle donne lieu l'unité lexicale et qui amène le locuteur à imaginer le référent d'une certaine manière, différente d'un mot à l'autre. Puisqu'il n'existe pas non plus deux unités lexicales à environnement identique, l'analyse du profil combinatoire (basée sur des outils lexicométriques) permet de cerner l'identité sémantique du mot et de le distinguer de ses synonymes. D'un point de vue méthodologique, l'article part donc de données distributionnelles, passe par la recherche de schémas explicatifs et aboutit à des observations sur les différences entre synonymes.

L'objectif d'Agnès Tutin, Iva Novakova, Francis Grossmann et Cristelle Cavalla est d'esquisser une typologie des noms d'affect à partir de leurs propriétés sémantiques et syntaxiques. Ils partent pour cela de l'hypothèse que les dimensions sémantiques peuvent être mises en évidence par la combinatoire lexicale et syntaxique. Ayant recensé dans une étude de corpus (utilisant Frantext) les contextes linguistiques des 40 noms d'affect les plus fréquents du français, ils étudient à travers ces cooccurrences un ensemble de dimensions : la structure actancielle, l'aspect (ponctuel *vs* duratif, phases), les manifestations physiques extérieures, le contrôle de l'affect, la causation, la verbalisation. La méthode combinatoire employée permet d'établir six classes de noms d'affect, basées en premier lieu sur la structure actancielle, le caractère ponctuel *vs* duratif et le contrôle, et plus accessoirement sur les manifestations extérieures et la verbalisation. Si les dimensions < causation > et < phases > se révèlent peu discriminantes, d'autres paramètres sémantiques jugés *a priori* non pertinents (la polarité, l'intensité, la dimension axiologique) apparaissent finalement essentiels.

Le propos de l'article de Jacques François et Jean-Luc Manguin est en premier lieu de comparer les terminologies qui se sont développées autour de la notion de « collocation » dans deux optiques, fonctionnelle et statistique. En second lieu, les auteurs testent l'intuition que les noms exprimant des types de dialogues ou de débats s'attachent des adjectifs particuliers, de type relationnel comme de type qualificatif, le rattachement adjectivo-nominal étant pour partie du type « quasi-phrasème ». À partir du nom *débat*, et en s'appuyant sur la liste des synonymes de *débat* fournie par le dictionnaire électronique des synonymes du CRISCO, ils repèrent toutes les cooccurrences < N + Adj > dans la base de données Frantext catégorisée de 1830 à nos jours. Trois types de profils collocatifs (au sens statistique) en ressortent : les collocations étroites (ex. : *joute oratoire*), asymétriques (ex. : *controverse théologique*) et lâches (ex. : *conciliabule interminable*). La notion fonctionnelle de « collocation » ne s'applique qu'au premier type, les deux autres ne représentant que des cooccurrences préférentielles.

La contribution d'Alain Polguère et Igor Mel'ăuk présente une « approche globale » de la description des dérivations sémantiques et des collocations, celle du DiCo/LAF. Cette approche se conforme au postulat suivant : les collo-

cations ne doivent pas être décrites de façon isolée. Les relations base-collocatif – liens lexicaux syntagmatiques – doivent être modélisées avec les dérivations sémantiques – liens paradigmatiques –, au moyen d'un même outil descriptif : les fonctions lexicales. De plus, il faut considérer d'autres propriétés lexicales, comme l'étiquette sémantique de l'unité lexicale, sa structure actancielle et son schéma de régime, tout cela en tenant compte d'une stricte séparation des différentes acceptations. Les auteurs expliqueront comment toutes ces propriétés sont encodées dans la base de données *DiCo*, en illustrant leur approche à partir du cas particulier du vocable polysémique français DÉBAT.

Dans leur contribution, **Serge Verlinde, Jean Binon et Thierry Selva** illustrent le traitement du mot *débat* dans le *Dictionnaire d'apprentissage du français langue étrangère ou seconde (Dafles)*, un dictionnaire d'apprentissage semi-multilingue du français général, accessible en ligne et destiné à des apprenants de niveau intermédiaire et avancé. Ils étudient successivement l'établissement des lexies et la rédaction des définitions, les relations syntagmatiques (la syntaxe du mot et ses cooccurrents) et les relations paradigmatisques (les corrélés).

Pour chacun de ces aspects, ils mettent en évidence la nécessité d'un dépouillement de corpus, en l'occurrence un corpus de textes journalistiques de plus de 75 millions de mots, provenant de trois communautés francophones importantes.

Ils procèdent également à une comparaison exhaustive de leurs données avec celles relevées dans les principaux dictionnaires et inventaires de collocations disponibles sur le marché à l'heure actuelle. La comparaison porte non seulement sur la sélection des collocations, mais également sur leur modélisation.

Dirk Siepmann nous propose une étude richement documentée qui aborde le phénomène collocatif du point de vue de l'apprenant de langues étrangères et du lexicographe pédagogique. Une première partie théorique se penche sur un certain nombre de difficultés inhérentes aux définitions actuelles de la notion de « collocation », telles que le caractère supposé « binaire » des collocations et les différences supposées de statut sémiotaxique entre la base et le collocatif. Une attention toute particulière est prêtée aux collocations entre sèmes. Sur la base de ces réflexions, l'auteur propose une nouvelle division du continuum phraséologique et une nouvelle définition du phénomène collocatif.

La deuxième partie de l'étude donne une vue d'ensemble de la façon dont les collocations peuvent être enregistrées dans un manuel de vocabulaire de fort volume. Dirk Siepmann argumente en faveur d'une méthodologie onomasiologique bilingue ou multilingue, seule à même selon lui « de répondre aux besoins de l'apprenant des langues et d'aboutir à un traitement exhaustif du lexique », et illustre les premiers résultats d'une telle méthodologie.

Gageons que ce volume aidera le lecteur à mieux s'orienter face aux différentes approches du rapport entre base et collocatif – lequel rapport constitue la clé de voûte de toutes les contributions. Espérons, de plus, que la notion de « collocation », longtemps négligée en linguistique française, sera encore mieux reconnue comme outil indispensable à une lexicographie moderne, celle qui s'appuie sur les avancées de la linguistique de corpus et qui met à profit certains modèles dictionnaires provenant d'outre-Manche.