

Histoires de mots : affinités (s)électives

Peter Blumenthal

Université de Cologne

peter.blumenthal@uni-koeln.de

1. Affinités, connexions et « forme interne de la langue »

Les « affinités » entre mots évoquées ici ont été dénommées de façons fort diverses depuis Charles Bally, disciple de Ferdinand de Saussure, qui parlait dans son *Traité de stylistique française* (1909) de « locutions phraséologiques ». Cette notion générale englobait dans sa terminologie les « unités phraséologiques » (que nous qualifierions aujourd'hui de « locutions idiomatiques ») et les « groupements usuels », lesquels correspondent *grosso modo* aux « collocations »¹ et que certains subsument sous la notion générale de « figement », bien pratique en raison de son caractère volontairement vague. Pour ne pas nous égarer dans la jungle terminologique où se meut depuis plusieurs décennies la discussion sur les syntagmes réputés non « libres », nous privilégierons d'entrée de jeu la « combinatoire des mots ». Celle-ci s'étend, selon les avis concordants de nombreux lexicologues (que nous ne partageons pas entièrement, cf. *infra* 2.), sur un continuum allant d'un pôle de liberté à un pôle caractérisé par des contraintes d'ordre idiomatique, voire par l'opacité totale du groupement. Nous placerons au cœur de nos réflexions les combinaisons préférentielles² – des cooccurrences dont la fréquence relativement élevée ne saurait être due au hasard et que les calculs probabilistes, effectués sur la base de corpus étiquetés, révèlent comme étant « spécifiques ». Force nous sera de constater que la liberté en matière de combinatoire risque d'être partiellement illusoire, à l'instar de celle que croient pouvoir nous accorder maints philosophes à propos du libre arbitre (cf. note 6). Mais avant de nous aventurer dans une remise en cause de l'existence réelle de la liberté combinatoire, esquissons le cadre théorique du thème qui nous occupe.

L'héritage terminologique du XX^e siècle contient une notion fondamentale restée quelque peu énigmatique : la connexion de Tesnière, « concept central de la théorie de la dépendance » (Neveu, 2004, sous **connexion**). La connexion, « indispensable à l'expression de la pensée » (Tesnière, 1969 : 12), constitue le lien invisible entre les mots dont elle transforme la succession en une unité supérieure. C'est à elle que correspond, dans l'esprit de Tesnière, « la *innere Sprachform* 'forme intérieure de la langue', de Guillaume de Humboldt » (*ibidem*). La « forme intérieure », sur la nature exacte de laquelle Humboldt s'étend peu, est censée représenter la catégorisation du monde par la langue³. Or, cette notion de catégorisation nous conduit au chapitre 24 du livre de Tesnière (« Les catégories »), où nous apprenons que « par les catégories de la pensée, l'esprit humain amène le monde à sa propre mesure » et que « par les catégories grammaticales, le langage amène la pensée à sa propre mesure ». Que pouvons-nous tirer de ce rapprochement entre connexion et catégories – sans faire violence à la pensée de Tesnière qu'on ne saurait embrigader pour une problématique qui ne fut pas forcément la sienne ? Avançons une réponse sous forme d'hypothèse de travail : puisque la combinatoire préférentielle, représentée par des groupements de mots caractéristiques d'une langue, constitue (pour des raisons statistiques) un cas privilégié de connexion, il peut s'avérer particulièrement intéressant de l'analyser sous l'angle des catégories mises en œuvre, qu'elles soient sémantiques et/ou grammaticales.⁴ Notre problématique peut par conséquent se formuler ainsi : les combinaisons de mots caractérisés par une affinité particulière nous renseignent-elles de façon pertinente sur les catégories par lesquelles « le langage ramène la pensée à sa propre mesure » ? Autrement dit, la collocation et le figement en général offrirait-ils un accès privilégié à ce que Humboldt voulait désigner par son concept fort stimulant de « forme intérieure »⁵ ? Cette idée, à l'origine peu claire du fait même de Humboldt, vouée ensuite aux gémonies par des courants formalistes de notre discipline, ne trouve-t-elle pas, de nos jours, une nouvelle actualité grâce aux recherches en neurosciences ? Ainsi, Edelman (2000 : 378), lauréat du prix Nobel de physiologie ou médecine en 1972, écrit-il la phrase suivante que Humboldt n'aurait guère reniée : « Le langage [...] crée un monde, et non

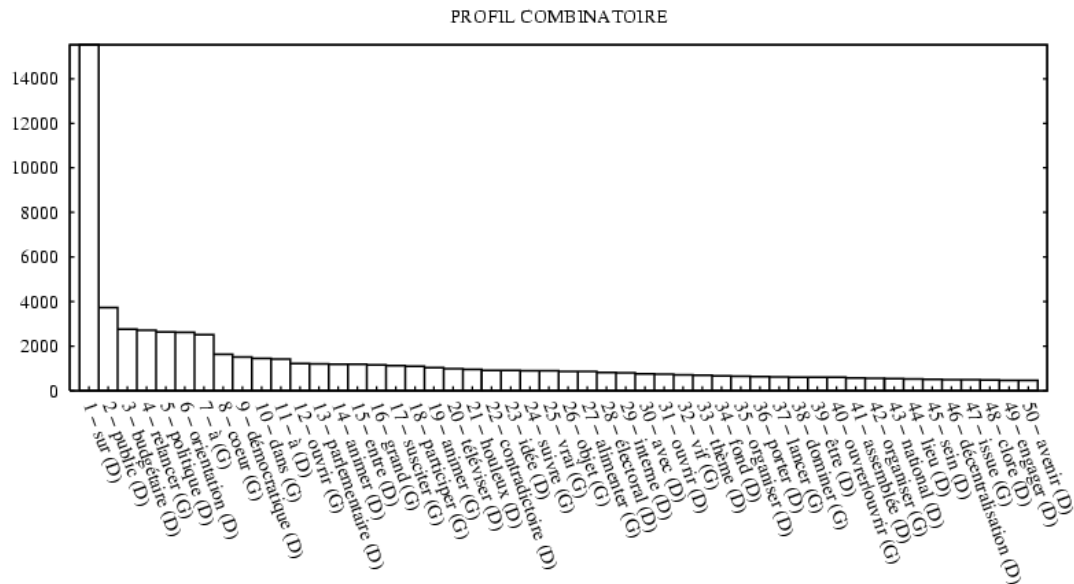
pas seulement un environnement. Ce monde regorge d'intentionnalité, de projections, de sentiments, de préjugés et d'affection. » Dans le même ordre d'idées, Rey (2008 : 192) considère dans son récent livre sur la « science du mot » que l'analyse des locutions, collocations et cooccurrences fréquentes devrait éclaircir leurs « tendances sémio-culturelles » et tenter de contribuer à « combler l'un des abîmes qui séparent la linguistique pure de son objet inéluctable : la réalité globale d'une fonction anthropologique ». Si cet objectif ne manque certes pas d'attraits, la voie qui y conduit est néanmoins semée d'embûches et peut nécessiter des recherches fouillées sur des détails parfois minutieux.

Notons au passage que de grands esprits du passé ont eu recours, pour symboliser le propre d'une affinité (entre personnes du sexe opposé ou entre mots), aux comparaisons avec les éléments de la chimie : Goethe en parle dans une page mémorable du quatrième chapitre de son roman *Les affinités électives*. Tesnière, qui rapproche d'ailleurs Goethe de Humboldt (1969 : 13), évoque dès le premier chapitre des *Éléments* le cas de la combinaison du chlore et du sodium (le sel de cuisine), laquelle « présente de tout autres caractères que le chlore Cl d'une part et le sodium Na d'autre part » ; tout ceci pour illustrer la vertu de la connexion dans sa phrase favorite *Alfred chante*. Prenons un instant au sérieux ce discours métaphorique, qui nous incite à examiner de très près ce qui peut arriver lors de la rencontre (non fortuite !) de deux ou plusieurs mots.

2. Ô liberté, que d'erreurs on commet en ton nom !

Passons du sel de cuisine de Tesnière à la « cuisine » de notre laboratoire de l'Université de Cologne. Celui-ci s'est spécialisé dans l'analyse automatique du voisinage de mots-pivots (ou « mots de base ») dans différents types de corpus électroniques (littéraires, journalistiques, appartenant à divers technolèctes, etc.). Nous appellerons « collocatifs » les accompagnateurs qui s'avèrent préférentiels sur la base du calcul probabiliste *log likelihood*. À titre d'exemple, nous présentons ici ce que nous dénommons (comme d'autres équipes, telle celle de Verlinde, e. a. à Louvain) le « profil combinatoire » – en l'occurrence, celui du mot *débat* dans un corpus d'environ 106 millions de mots (G = à gauche du mot pivot, D = à droite) :

DIAGRAMME 1



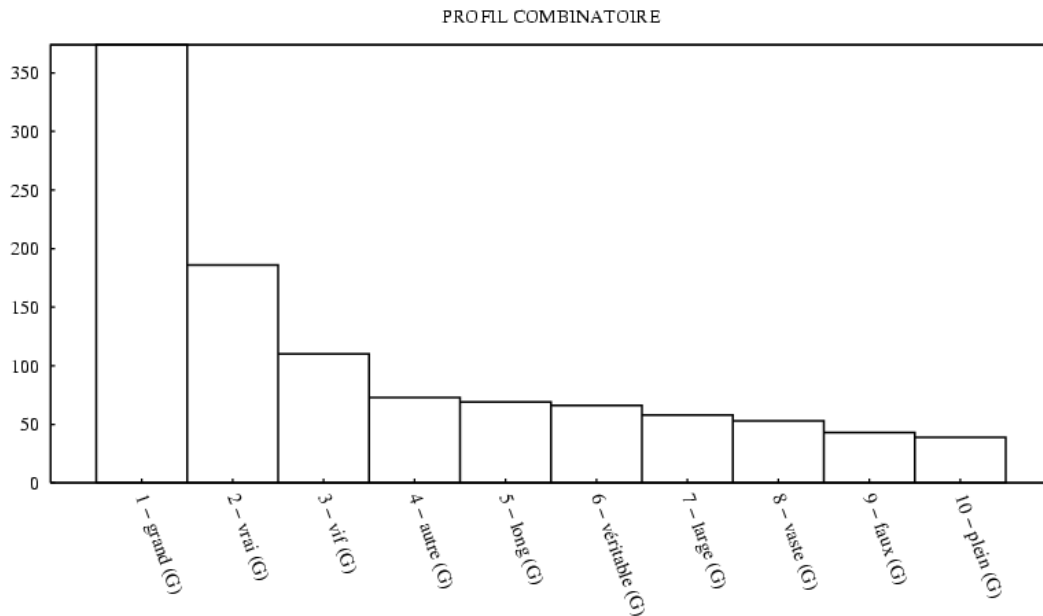
Collocatifs de débat dans quatre journaux de 2002 (*Le Figaro, Le Monde, Sud-Ouest, L'Est Républicain*), fenêtre de 5 mots à gauche et à droite (noms, verbes, adjectifs, prépositions sans *de*) ; type de calcul : *log likelihood*

L'accompagnateur de loin le plus spécifique de *débat* est la préposition *sur* (au 1^{er} rang en abscisse) à droite de *débat*, comme le fait apparaître le degré de spécificité selon *log likelihood* (plus de 14.000) indiqué en ordonnée. On en tirera la conclusion, certes sans grand effet de surprise, que le sens de *débat* se conçoit essentiellement en fonction d'une thématique particulière, amenée par *sur*.

Remarque : L'intérêt de cette conclusion, à première vue banale, se dégage d'une comparaison avec les synonymes de *débat*. Le mot le plus semblable à *débat* est, selon le dictionnaire des synonymes de l'Université de Caen (laboratoire CRISCO) le nom *discussion* qui, lui, engendre un profil très différent : son collocatif le plus spécifique est *avec* (à droite). De nombreux autres faits distributionnels (par exemple, le rang élevé de *entre*) montrent clairement que dans le cas de *discussion*, l'aspect interactionnel l'emporte sur l'orientation thématique – pourtant également sensible. Loin de nous écarter de notre sujet, cette petite comparaison entre deux synonymes sert surtout à donner un avant-goût de notre principal but : sonder les orientations sémantiques profondes (appelées plus bas « catégorielles », cf. 3.) d'un mot à travers ses emplois.

L'intention première de notre examen de *débat* est cependant une réflexion sur le concept de liberté combinatoire du mot. Revenons au diagramme ci-dessus, où se trouve au rang 16 l'adjectif antéposé *grand* ; *grand(s) débat(s)* est attesté 374 fois dans notre corpus. En nous focalisant sur les épithètes antéposées à *débat*, nous obtenons le diagramme suivant, sorte d'agrandissement d'un détail du diagramme précédent :

DIAGRAMME 2



Adjectifs antéposés à débat dans quatre journaux de 2002 (*Le Figaro*, *Le Monde*, *Sud-Ouest*, *L'Est Républicain*), fenêtre d'un mot à gauche ; type de calcul : nombre d'occurrences

Faisons un autre agrandissement pour nous concentrer sur la première des épithètes antéposées, *grand* (qui garde ce rang même sur la base de *log likelihood*), en nous posant la question suivante : la combinaison de *grand* et *débat* peut-elle passer pour 'libre' – au sens de non déterminée par des contraintes provenant du système de la langue ou de l'usage ? L'excellent *Dictionnaire des combinaisons de mots* (DCM, 2007), paru l'an dernier aux Éditions Le Robert, affirme dans sa préface une telle liberté (p. VI) en ces termes : « Près de 162 000 combinaisons sont décrites, des plus libres (ex. : *grand débat*, *obtenir une aide*) aux plus idiomatiques (ex. : *colère noire*, *comptes d'apothicaire*). » *Grand débat* donc comme modèle de la combinatoire dite libre – qu'en est-il dans la perspective de notre approche statistique ? Le statut à l'évidence non idiomatique de ce syntagme garantit-il son caractère libre ? Tout dépend de ce que l'on entend par « liberté ». Si ce mot prend le sens d'indétermination⁶ que la tradition philosophique occidentale lui prête souvent, *grand débat* relève non pas du règne de la liberté, mais d'un entrelacs de déterminations réciproques que seules les méthodes quantitatives et en particulier l'analyse probabiliste parviennent à démêler. Ces interdépendances transparaissent à travers une extension du champ contextuel de l'observation, passant de la combinaison binaire (cas classique de la collocation simple) à des contextes plus étendus. Lorsqu'on utilise un outil informatique permettant de détecter les mots les plus fréquents (et/ou spécifiques) dans le contexte de *grand débat*, on se rend vite compte que l'environnement de cette combinaison tend à être plus ou moins stéréotypé – mais attention : les observations de ce genre ne valent strictement que par rapport à un corpus déterminé, constitué par un certain type de textes. Dans notre corpus des quatre quotidiens de 2002 (cf. diagramme 2), les résultats suivants se dégagent assez nettement : un *grand débat* est typiquement *national* et l'on en parle la plupart du temps comme d'une tâche ou d'un défi de l'avenir, ce qui peut favoriser le recours à des expressions modales ou finales ainsi qu'à des verbes comme *exiger* et *lancer*. Voici un exemple statistiquement représentatif, tiré du *Figaro* (2002) : « Le choix des Pays de la Loire pour lancer ce grand débat national n'était pas neutre. » De nature fortement préconstruite et souvent réglées comme du papier à musique, ces

« collocations complexes » ne se modifient pas de façon inconsidérée. Ainsi, le syntagme *relancer + débat* (rang 4 au diagramme 1), bien plus spécifique encore que *lancer + débat* (rang 37), se présente normalement dans un contexte totalement différent. Comment qualifier l'instance qui prédétermine les arrangements de mots du type cité ? Du moins à première vue, il ne s'agit pas là de tendances profondes inscrites dans le système catégoriel de la langue, mais du jargon, des clichés ainsi que des modes d'argumentation stéréotypés⁷ et topoï qui en découlent, de l'air du temps et éventuellement de la langue de bois en usage dans certains milieux et certains médias. Rastier (2006) parle à ce propos de « stéréotypie textuelle » qui relève de la « doxa », terme défini par Schapira (1999 : 160) comme « le savoir collectif, l'ensemble des idées et des opinions communes et leur expression, ayant cours dans une société déterminée, à un moment donné de son développement ». Concluons que la fameuse liberté, qui serait inhérente à ce que Sinclair appelle – non sans mettre en doute cette notion⁸ – l'*open-choice principle* (opposé à l'*idiom principle*), peut s'avérer rapidement illusoire pour peu qu'on aille au-delà des apparences de la prétendue créativité du langage. Malgré l'impact de la doxa sur la combinatoire, nous avons cependant pressenti, dans cette section, l'existence d'une autre force contribuant à orienter la combinatoire du mot : les éléments de type sémantique qui poussent *débat* à privilégier le lien avec la préposition *sur*. Nous entrevoyons donc dès maintenant les contours d'une théorie de la combinatoire du mot, soumise à (au moins – v. note 33) deux champs de forces, la doxa d'une part, les catégories sémantiques de l'autre.

Retenons que la méthode permettant de se prononcer sur le degré de liberté d'une combinaison se tient dans un cadre en partie saussurien. Elle opère en deux étapes : on se rapproche d'abord, comme par zoom, des accompagnateurs qui, à l'intérieur de leur paradigme (par ex. celui des épithètes antéposées), se signalent par un intérêt statistique particulier ; ensuite, en partant de la combinaison qui s'est avérée préférentielle, on passe de l'axe paradigmatique à l'axe syntagmatique pour filtrer d'éventuelles collocations complexes – celles, en quelque sorte contraintes, qui s'assimilent à des combinaisons stéréotypées. Nous ne cessons de souligner que la validité des résultats de cette méthode dépend du corpus choisi. Car le style de « la vie des mots », pour reprendre le titre d'un livre célèbre de sémantique française, diverge souvent fortement d'un type de textes à l'autre.

3. En-deçà de la doxa : le potentiel catégoriel

Comme nous venons de le voir, il paraît plausible qu'avant de devenir l'enjeu de tel cliché ou jargon à la mode qui figent leur emploi dans des cooccurrences plus ou moins stéréotypées, les combinaisons préférentielles du mot soient influencées par les orientations sémantiques générales (« catégorielles ») qui lui sont propres. Ces orientations font que nous associons spontanément son référent à certaines qualités, situations et relations plutôt qu'à d'autres. Il s'agit donc, selon cette hypothèse, de caractéristiques du mot (subsumables sous le concept générique de traits pour les partisans de ce type d'analyse, cf. note 34) qui préexistent à son éventuelle exploitation par la doxa.

Pour illustrer ces idées sur le rapport entre combinatoire et valeurs catégorielles sous-jacentes, nous nous pencherons sur un détail quasiment microscopique dans le comportement du nom *amour*⁹. Notre problématique est la suivante : quelle différence combinatoire et sémantique sépare les deux syntagmes *en amour*¹⁰ et *dans l'amour* ? Et pourquoi l'antonyme *haine* ne se prête-t-il pas au même type d'opposition ? Choisissons au hasard deux citations dans le *Petit Robert* pour exemplifier ces emplois : « « Il ne joue pas le jeu » disent les Anglais d'un homme qui triche en amour, en affaires, en politique » (Maurois, sous **tricher**) et « Même dans l'amour, même en étant deux, on ne veut pas être deux, on veut rester seul » (Montherlant, sous **deux**). Le locuteur natif expliquera sans doute qu'*en amour* signifie 'dans le domaine de l'amour', 'en matière d'amour', alors que la gamme des valeurs de *dans l'amour* paraît plus large et peut impliquer la référence à des situations tout à fait concrètes, comme c'est le cas des 32 positions « des corps dans l'amour physique » mentionnées par le *Petit Robert*, sous **position**. Pour étayer ces premières intuitions que l'on peut avoir sur les constructions avec *amour*, nous appliquons notre méthode habituelle, en calculant les mots les plus fréquents dans l'environnement des deux syntagmes.¹¹ Les noms à droite des deux syntagmes prépositionnels (*en amour* et *dans l'amour* dans une fenêtre de 20 mots) se distinguent fortement sur le plan lexical et assument des fonctions syntaxiques totalement

différentes : *en amour* est suivi de noms comme *amitié*, *politique*, *affaires*, qui se trouvent syntaxiquement (ou du moins logiquement) sur le même plan qu'*amour* ; cf. « En amour comme en politique, l'angélisme, c'est l'ennemi. » (*Frantext*, d'Ormesson) ou « Fidèle en amour, G. L. l'est aussi en amitié. » (*Le Figaro*, 2002). Sur le plan textuel, le syntagme *en amour* suscite donc l'apparition de noms qui semblent se trouver dans une relation paradigmatique avec *amour*. À gauche du syntagme n'apparaissent pas de mots saillants. Quant à *dans l'amour*, les noms à droite ne constituent pas des éléments coordonnés, mais des compléments déterminatifs d'*amour* (*femme*, *homme*, *famille*, *musique*, etc.) ; les mots les plus typiques de la fenêtre gauche (également de 20 mots) sont les verbes *trouver* et *élever*. Voici quelques exemples caractéristiques :

« trouvant l'énergie dans son amour de la famille » (*L'Est Républicain*, 2002) ;
« Élevé dans l'amour de la musique par ses oncles pianiste et violoniste, il forme sa voix dans les chœurs d'église. » (*Encyclopædia Universalis*, 2005) ; « Il y vécut dans l'amour de la musique et des longues randonnées solitaires. » (*ibidem*) ; « La prise de conscience de l'autre en tant qu'être humain trouvera sans doute un épanouissement dans l'amour réciproque de l'homme et de la femme [...]. » (*ibidem*).

Dans ces dernières citations, *dans l'amour* ne représente pas un complément circonstanciel « scénique » (ou « cadratif »), comme c'est le cas de la citation ci-dessus de Montherlant (« Même dans l'amour, ... »), mais un complément prépositionnel du verbe,¹² difficilement antéposé. C'est dans ces constructions-là qu'apparaît le plus clairement la capacité de *dans l'amour* à nommer le milieu ou les conditions d'un processus qui se déroule dans l'ordre des faits évènementiels (quelqu'un vit ou est élevé *dans l'amour*). En revanche, celui qui « triche en amour », pour reprendre l'exemple plus haut, ne triche pas à l'intérieur d'une situation constituée par un amour préexistant, mais par rapport aux attentes que l'on peut légitimement avoir dans le domaine de l'amour. Cette valeur sémantique, à la fois plus abstraite que celle de *dans l'amour* et bien plus limitée en extension, car moins polysémique (*en amour* ne réfère qu'à l'amour érotique,¹³ qui représente en français l'amour tout court), se manifeste encore plus clairement dans les nombreuses phrases à teneur générique ; exemples : « Hélas, en politique comme en amour, les ménages à trois, quatre, ou plus, donnent rarement satisfaction. » (*Le Monde*, 2002) ; « Dans la corruption comme en amour il faut être au moins deux, ici le corrupteur et le corrompu. » (*ibidem*). La dernière citation montre clairement qu'avec *corruption*, *dans* se charge du rôle que joue *en* devant *amour*. Cette fonction est aussi celle de *dans* devant *haine* dans la phrase suivante, non générique : « Rancunier ? Disons plutôt fidèle. Dans la haine. Comme en amour. » (*ibidem*).

Nous pouvons désormais proposer un premier bilan des constructions discutées ici : *en amour* se caractérise par deux qualités – des tendances plutôt que des règles strictes : 1) la capacité à véhiculer une vision abstraite et distanciée, voire générique de la sphère de l'amour ; 2) la propension à la confronter à d'autres domaines paradigmatiquement comparables. *Dans l'amour* se signale par des valeurs concrètes et situationnelles.¹⁴

Remarque : Toutefois, dans la réalité des corpus, il semble y avoir intersection entre les emplois d'*en amour* et de *dans l'amour*, dans une zone intermédiaire entre les pôles abstrait et concret – avec des incursions occasionnelles de l'une des constructions dans le domaine de l'autre. Quant aux noms comme *haine* que l'usage exclut de la combinatoire avec *en*,¹⁵ il faut sans doute considérer que *dans* + nom constitue, sur le plan sémantique, une neutralisation entre les valeurs distinguées par l'opposition (partielle) entre *en amour* et *dans l'amour*.

Nous approchons du problème central de notre investigation : dans la combinatoire préférentielle des mots (leurs « affinités électives »), il existe apparemment des liens qui permettent l'expression explicite et lexicalisée d'une certaine catégorisation (ici l'amour en tant que « domaine »¹⁶) que d'autres mots, sémantiquement très proches, tendent à exprimer – et souvent ne sauraient exprimer que – de façon ambiguë (*dans l'amitié*) ou périphrastique (*dans le domaine de l'amitié*, *en matière de haine*).

Ces phénomènes de mise en perspective du référent¹⁷ par la combinatoire du mot de base possèdent souvent une dimension diachronique passionnante. Les banques de données comme *Frantext* offrent de merveilleuses possibilités de recherche sur le cheminement de ces structures et leurs péripéties au cours

d'un millénaire d'histoire du français. Le syntagme *en amour*, avec la gamme de valeurs sémantiques et les différentes fonctions syntaxiques que nous lui connaissons aujourd'hui, était déjà parfaitement en place au XVI^e siècle. Le grand intérêt d'une enquête diachronique poussée à travers les siècles réside dans l'observation du voisinage d'*en amour*, comprenant un petit nombre de noms attribués contextuellement au même paradigme (tels *amitiés*, *guerre*, *jeu*, *affaires*, *chasse* et parfois quelques autres, moins stéréotypés). Dans une perspective de longue durée, l'historien de la langue sera sensible aux tentatives, périodiquement relancées, d'uniformiser les syntagmes prépositionnels, de construire les autres noms de façon parallèle à *en amour*, de dire donc *en guerre*, *en chasse*, etc. Tendances auxquelles la majorité des noms finissent par résister, pour reprendre une préposition plus conforme à leur sémantisme (essentiellement *dans* ou *à*). Voici quelques citations, trouvées dans *Frantext*, qui émaillent la longue histoire d'une construction :

- « [...] autres maux et inconveniens qui sçauroient advenir à l'homme en ceste vie, de peu de pris au respect du tourment d'esprit que sent celuy qui a esté mal fortuné **en amour** et malheureux **en femme**. » (Poissenot, 1583).
- « si bien que nous n'avons à faire qu'à des femmes, qui, à la verité, sont redoutables **en amour**, mais grandement inutiles **en la guerre**. » (d'Urfé, 1627).
- « et qui pensoit que le plus grand plaisir **en amour** estoit comme **à la chasse**, de prendre à force ce que l'on poursuit » (d'Urfé, 1627).
- « **en amour** comme **en chasse**, la vieillesse routièrè éventa bien la trace » (de Schéalandre, 1628).
- « on s'en pare **en amour**, **en amitié**, **en affaires**, **dans le commerce** du monde, **dans les compliments** » (Scudéry, 1667).

Si les jeux semblent faits pour l'essentiel à partir des années 1660 – l'identité des noms pouvant se construire avec *en* dans cette construction ne changeant plus fondamentalement – on relève dans le français actuel comme à chaque époque de sa formation des escapades de tel ou tel auteur essayant d'échapper à une norme qui a toujours peiné à s'imposer.¹⁸ Puisque la notion de norme paraît par ailleurs peu adéquate dans ce domaine, le linguiste est réduit à parler plutôt de tendances statistiques caractérisant les emplois de tel nom – tendances variant parfois au gré des types de textes et des registres stylistiques. Cependant, une fois cette conception probabiliste de l'usage acceptée, de nouvelles possibilités de description et d'analyse se profilent. Ainsi, en comparant les comportements d'*amour* et d'autres noms appartenant au même champ sémantique comme *amitié* et *haine*, l'on observe d'intéressantes dissemblances à propos d'une structure aussi réduite que *en* + nom. Nous avons déjà relevé qu'*en amour* se présente souvent (et de façon 'spécifique' au sens défini du terme) avec une fonction cadrative (en antéposition segmentée, à titre d'arrière-plan thématique de l'énoncé) avec une certaine autonomie par rapport à la suite ;¹⁹ *amitié* reste rare dans cette fonction, dont *haine* est exclu (**En haine, il faut...*). En revanche, les occurrences d'*en amitié* dépassent celles d'*en amour* dans la grande majorité des corpus (et sur *google.fr*) lorsque ce groupe prépositionnel dépend d'un adjectif ou d'un nom – ce qui diminue son autonomie au sein de la phrase. Les adjectifs et les noms régissant *en amitié* ne sont pas quelconques, mais s'intègrent à ce que l'on peut appeler une collocation ternaire (par exemple, adjectif + préposition + *amitié*). Il s'agit surtout de (*in*)*fidèle en amitié* (mais **fidèle en haine*), de *fidélité en amitié* et de *jaloux en amitié* – combinaisons qui n'excluent pas une coordination avec *en amour*, comme le montrent un exemple déjà cité ci-dessus et la phrase suivante, trouvée sur *google.fr* : « Êtes-vous jaloux en amour ou en amitié ? Répondez à cette question sur Yahoo ! » Soulignons que, dans les constructions de ce type, l'antéposition d'*en amitié* semble possible (*En amitié il est fidèle*, tournure toutefois moins souvent attestée sur *google* qu'*En amour il est fidèle*) et peut contribuer à revaloriser le degré d'autonomie du syntagme prépositionnel. Alors qu'*en amour* et *en amitié* peuvent donc se présenter en semi-autonomie, *amour* est exclu d'un syntagme figé (avec perte totale d'autonomie au sens défini) dans lequel *amitié* et *haine* ne sont pas rares : *prendre quelqu'un en amitié/haine*, mais **prendre quelqu'un en amour*.

Comment interpréter ces quelques constatations sur le comportement combinatoire de trois noms appartenant à la même sphère onomasiologique ? Convenons d'abord de l'extrême exigüité de notre

champ d'observation, analysé à la loupe selon le seul critère syntaxique des degrés d'autonomie. Les faits combinatoires relevés permettent cependant d'entrevoir une hypothèse explicative, qu'il conviendrait de recouper avec d'autres observations : le nom *amour*, bien plus qu'*amitié* et à l'opposé de *haine*, possède cette qualité de pouvoir présenter son référent comme un 'domaine'. Évidemment (et heureusement), tout amour n'est pas 'domaine', *i. e.* champ d'une réflexion ou d'un comportement circonscrit et virtuellement opposé à d'autres champs. Les calculs et les analyses des profils combinatoires démontrent que, parmi les dispositions (définissant général d'*amour* dans le *Petit Robert*), l'*amour* occupe un rôle à part quant à ses « connexions » – très différentes de celles de son prétendu antonyme *haine*. De quoi méditer sur les relations entre un mot, les choses et l'univers du discours auquel ressortit l'*amour* dans l'histoire de la civilisation française. Rappelons que, pour Montaigne (*Essais*, livre III, chap. 3), l'*amour*, l'*amitié* et les livres sont les « trois commerces » qui nous aident à vivre. On peut affirmer, sans faire tort à la combinatoire de *livre*, que la « forme intérieure » de la langue tend à honorer la particularité du premier de ces commerces, laquelle paraît s'être progressivement affirmée au fil des siècles dans le sens d'une clarification catégorielle.

Nos considérations sur le comportement combinatoire et syntaxique (problème de l'autonomie du syntagme prépositionnel) du nom *amour* ne manquent pas de soulever des questions sur l'intérêt des notions de « profil combinatoire » et de « potentiel catégoriel ». Ces notions n'ont d'intérêt que lorsqu'on prend soin de les réserver à l'analyse soit de la même acception d'un mot, soit d'acceptions proches les unes des autres²⁰ susceptibles de passer pour différentes facettes d'un seul sens global. Sinon, ces données ne reflèteraient que la richesse polysémique du mot de base, problématique peu pertinente quand on s'interroge sur la conceptualisation d'une certaine réalité par les possibilités combinatoires d'une langue.

4. Combinatoire et catégorisation des noms d'affect

Nous avons jusqu'ici suivi en premier lieu un problème purement syntagmatique, la combinatoire de *débat* (section 2.), pour aborder ensuite des questions liées aux deux axes syntagmatique et paradigmatique : la combinatoire prépositionnelle d'*amour*, les relations d'équivalence approximative et partielle (donc de synonymie) entre *en amour* et *dans l'amour* ainsi que la polysémie du nom en fonction de son voisinage. Nous poursuivrons en développant le problème (paradigmatique) de la synonymie dans la perspective de la combinatoire privilégiée entre verbes et compléments d'objet nominaux ; nous nous intéresserons en particulier aux affinités de certains verbes (surtout *éprouver* et *ressentir*)²¹ avec les noms d'affect.

Dans un article sur les robots qui « apprennent les émotions humaines » (projet de recherche financé par l'Union Européenne), le journaliste (*Le Monde*, 23/6/2007) attire l'attention sur les corrélations entre affects et phénomènes linguistiques (prosodie, sémantique, syntaxe, etc.) et explique à propos de deux des « états émotionnels basiques » que « le vocabulaire de la joie diffère de celui de la colère ». Qui l'eût cru ? Cependant, au-delà des différences, il existe certainement aussi des affinités importantes à l'intérieur de ce champ conceptuel – même en ce qui concerne la combinatoire des deux affects nommés, *joie* et *colère*. Un bref coup d'œil sur les matériaux du Web (*google.fr*) suffit pour montrer que l'on peut « éprouver », mais aussi « ressentir de la joie » et « de la colère ». Inversement, « sentir de la colère » (en soi-même) se présente bien rarement ; quant à *sentir de la joie*, il est inexistant dans notre corpus littéraire (cf. *supra*). Mais on peut parfaitement « sentir la colère monter » ou « sentir la joie » d'une autre personne. Lafaye (1869 : 108), éminent synonymiste du XIX^e siècle, voyait la différence entre *sentir* et *ressentir*, au sein de la langue classique qui l'intéressait avant tout, en ceci que « *ressentir* est relatif à une cause étrangère dont on reçoit l'action »²², alors que *sentir* lui paraissait plus « intime » et « subjectif ». En français contemporain, le critère de la cause externe (dite aujourd'hui aussi « exogène », cf. Anscombe, 1995) ou interne (« endogène »), qui devait connaître un regain d'intérêt dans la sémantique de la fin du XX^e siècle, s'applique probablement moins à la relation entre *ressentir* et *sentir* qu'à celle entre *ressentir* et *éprouver*.²³ Retenons toutefois la pertinence du critère de la causalité externe, candidat parfait au statut d'élément du potentiel catégoriel d'un mot. Quant à la concurrence de *ressentir* et *éprouver* devant les noms d'affect, nous nous bornerons à résumer ici les résultats d'analyses probabilistes développées en détail ailleurs (Blumenthal, 2009). Les faits quantitatifs, d'abord : les calculs

sur la base de *log likelihood* dans un corpus littéraire (*Frantext.romans*) montrent que, si les deux verbes peuvent se combiner en principe avec la très grande majorité des noms considérés comme référant à des émotions ou des sentiments, le degré d'affinité combinatoire entre verbe et nom présente néanmoins parfois d'importantes différences. Ainsi, *ressentir* est bien plus spécifique qu'*éprouver* devant *choc*, *douleur*, *émotion* et *trouble*, alors qu'*éprouver* l'emporte de loin sur *ressentir* devant *admiration*, *amour*, *crainte*, *passion*, *plaisir*, *tendresse* et *sentiment*. À partir de ce constat empirique (qui ne vaut, insistons bien là-dessus, que pour un certain corpus), plusieurs méthodes ouvrent la voie aux conclusions auxquelles nous aspirons et dont l'enjeu peut se formuler comme suit : que nous apprend la combinatoire préférentielle du mot de base (nom d'affect) sur son potentiel catégoriel ? Sur le plan méthodologique, on peut :

(a) chercher des dénominateurs communs de noms ayant une préférence patente (exemple : les noms se construisant de préférence avec *éprouver* comportent souvent une connotation positive) ;

(b) analyser de façon détaillée toutes les phrases contenant un nom d'affect régi soit par *éprouver*, soit par *ressentir* afin d'y détecter des facteurs sémantiques et combinatoires corrélés à l'un de ces deux verbes (exemple : le nom accompagnant *éprouver* est plus souvent qualifié par une épithète ou un complément, il tend donc à véhiculer une information d'ordre qualitatif) ;

(c) étendre l'étude d'*éprouver* et de *ressentir* à des emplois en dehors de la classe d'objets des affects, dans l'espoir de découvrir des dénominateurs communs à l'intérieur de la polysémie de chacun des verbes ; partant de la valeur générale du verbe, on peut ensuite revenir au potentiel catégoriel du nom d'affect caractérisé par son affinité avec *éprouver* ou *ressentir* (le corpus *Frantext.romans* est quantitativement insuffisant pour une telle recherche).

Illustrons brièvement cette dernière approche (c). Dans les journaux et sur Internet, *ressentir* se rapporte avec une fréquence très élevée à l'enregistrement pur et simple d'un fait objectif, en faisant abstraction de tout jugement de valeur ou de l'aspect qualitatif du phénomène (par exemple, d'un tremblement de terre), alors qu'*éprouver* est exclu de ces emplois (**un tremblement de terre a été éprouvé...*). Retournons à présent aux noms d'affect : le verbe *ressentir* se combinerait-il de préférence avec les noms qui, de par leur potentiel catégoriel, ont vocation à appréhender un affect sous l'angle de son existence : *trouble*, *émotion*, *douleur*, *choc* ? Inversement, *éprouver* serait-il l'accompagnateur attitré des noms ayant tendance à focaliser un affect du point de vue de sa qualité : *admiration*, *amitié*, *passion*, *plaisir*, *tendresse*, *sentiment* ? De nombreux autres indices concourent à corroborer cette hypothèse. Contentons-nous d'un seul exemple. Dans le cadre de l'approche (b), on constate qu'*éprouver* et *ressentir* se combinent tous deux de manière spécifique avec *besoin* – souvent dans des phrases *grosso modo* semblables, mais qui contiennent aussi des divergences caractéristiques dont le nombre ne semble pas être le fruit du hasard. Après *éprouver* + *besoin* vient en général un infinitif exprimant une action que le sujet désire accomplir (exemple : « au bout de quelques phrases, j'ai éprouvé le besoin de me lever et de marcher dans la pièce. » ; *Frantext*, Dutourd, 1967). Après *ressentir* + *besoin*, on trouve assez souvent un complément nominal désignant l'objet matériel du désir (« ils ressentaient le besoin d'une chambre un peu plus grande, d'eau courante, d'une douche, de repas plus variés, ou simplement plus copieux que ceux des restaurants universitaires, d'une voiture peut-être, de disques, de vacances, de vêtements. » ; *Frantext*, Pérec, 1965). En fait, cette différence, pertinente bien que non régulière, correspond largement à la différence que fait le *Petit Robert* sous **besoin** (I.2.) entre le sens « subjectif » et « objectif »²⁴ de ce mot.

Que conclure de ces observations concernant la combinatoire étendue d'*éprouver* et de *ressentir* ?

– Primo, les deux verbes se distinguent par le fait qu'*éprouver* tend à porter un éclairage « subjectif »²⁵ sur l'affect, alors que *ressentir* favorise le centrage « objectif » de la sphère du complément d'objet – centrage qui a pour corollaire l'idée que la réalité exprimée par ce complément n'est enregistrée qu'objectivement, sans appréciation qualitative ou jugement de valeur.

– Secundo, l'hypothèse exprimée ci-dessus à propos du potentiel catégoriel de deux groupes de noms d'affect trouve une certaine confirmation : les noms liés fortement à *éprouver* sont de nature à fournir une

vision subjective (au sens d'appréciative) d'un phénomène émotionnel, les noms privilégiant nettement *ressentir* focalisent l'attention sur le caractère factuel de ce phénomène, éventuellement conditionné par des facteurs externes au sujet. Les deux points de vue sont toutefois possibles pour la majorité des noms d'affect, même si l'on peut démontrer des affinités entre tel nom et l'un ou l'autre des points de vue.

Après l'analyse minutieuse des différences combinatoires entre *éprouver* et *ressentir* (dans un seul corpus, il ne faut pas l'oublier), il convient de relativiser les résultats dans une perspective statistique. Car si les deux verbes sont loin du statut de synonymie parfaite, les ressemblances combinatoires l'emportent. En voici la preuve : en calculant dans *Frantext.romans* le taux de similarité entre les voisinages nominaux des deux verbes (dans une fenêtre de 5 mots à gauche et à droite du mot pivot, limitation aux 50 premiers collocatifs)²⁶, on obtient une valeur étonnamment élevée : 0,559. Bien entendu, ce chiffre n'informe que sur l'identité des noms et leurs rangs dans les deux profils combinatoires comparés, sans nullement exclure les différences de perspective véhiculées par *éprouver* et *ressentir*. On appréciera ce taux à sa juste valeur en sachant que le même calcul fournit seulement 0,097 pour *ressentir/sentir*,²⁷ verbes apparemment plus proches dans la langue ancienne, et 0,016 pour *éprouver/sentir*. *Sentir* évolue donc aujourd'hui dans un univers du discours très différent de celui des deux autres verbes – non seulement dans *Frantext.romans*, comme le prouvent nos analyses faites dans le journal *Le Monde* (2002), où *sentir* se trouve encore bien plus éloigné de *ressentir* (taux : 0,043) et d'*éprouver* (taux : 0,003) que ce n'est le cas dans les textes littéraires ; par contre, le taux de similarité reste relativement élevé pour *ressentir/éprouver* (0,338), le seul des couples où les deux verbes peuvent de nos jours être considérés comme synonymes dans leurs acceptions les plus fréquentes.

Au XIX^e siècle, Lafaye et Littré avaient une vision sensiblement différente des relations entre ces trois mots. Seuls des calculs sur la base de corpus représentatifs des siècles passés sauraient nous renseigner avec précision sur les évolutions historiques au sein de ce petit champ sémantique. Histoire à suivre. En tout cas, les « histoires de mots » du titre ne peuvent se comprendre sans l'histoire des mots, à son tour largement déterminée par leurs affinités combinatoires au cours de plus de mille ans de langue française.

5. Synthèse : affinités, profils et potentiels

Après les différents gros plans sur le comportement de quelques mots et l'analyse fouillée de leurs tropismes (leurs affinités combinatoires à première vue parfois un peu énigmatiques), nous devons au lecteur un plan d'ensemble. Car le risque est grand que les arbres cachent la forêt, les facteurs influant sur les possibilités d'emploi des diverses acceptions d'un mot s'étant avérés nombreux et variés. Le principe selon lequel nous avons analysé ces facteurs n'est pas celui d'une logique binaire (par exemple, combinaison +/- grammaticale) ou d'un déterminisme monocausal omniprésent, mais celui d'un système ouvert régi par des tendances statistiques, plus précisément des lois de probabilité – système produisant²⁸ par définition des « contre-exemples ». Le concept de base utilisé ci-dessus est clair et simple : il s'agit du « **profil combinatoire** » d'un mot, structure descriptive illustrée par le diagramme 1 (cf. *supra*). Ce profil reflète le comportement stéréotypé²⁹ d'un mot (ou, après désambiguïsation, d'une de ses acceptions) dans un corpus, et relève ainsi d'une certaine « parole ». Peut-on escompter s'approcher, en combinant les analyses de nombreux corpus, de l'asymptote que représenterait le profil du même mot au niveau de la « norme » telle que la concevait Coseriu ?³⁰ L'espoir de parvenir un jour (grâce à une exploitation intelligente des énormes ressources du Web ?) à un tel panorama compréhensif pêche actuellement par excès d'optimisme. En attendant, il paraît plus réaliste de s'accommoder des profils qui se dégagent de certains registres, variétés ou types de textes (romans, presse, ouvrages d'une discipline scientifique, langue familière, etc.). Dénommons « **potentiel combinatoire** » la vision quelque peu utopique de la combinatoire caractéristique du mot dans la « norme ». À partir de l'exemple de *débat*, nous nous sommes aperçus que la liberté de choisir les accompagnateurs d'un mot n'est parfois que doux semblant, car beaucoup de facteurs concourent à téléguidier la sélection des mots du contexte, et ce, largement au-delà des simples collocations binaires que nous proposent les dictionnaires spécialisés. Celles-ci ne constituent souvent que la partie visible de l'iceberg que forme le réseau des associations syntagmatiques d'un mot.³¹ Comment définir les forces déterminant les caractéristiques du profil combinatoire ? C'est à partir de cette interrogation que notre parcours méthodologique passe du volet descriptif, ancré dans des

données chiffrables, à celui des hypothèses explicatives, nécessairement provisoires. Dans les corpus étudiés ici, il nous a semblé possible de ramener les principaux éléments du profil combinatoire soit à la « **doxa** » (opinions consacrées par l'air du temps, clichés, prêt-à-penser médiatique, idéologies, etc.), soit à des structures moins superficielles, parce qu'enracinées dans le système du lexique, le « **potentiel catégoriel** »³² du mot.³³ Nous entendons par là l'ensemble des critères permettant à un mot de présenter son référent sous un certain angle (exemples : l'amour non seulement comme une situation concrète, mais aussi en tant que « domaine » – cf. *en amour...* ; l'affect comme objet d'une observation factuelle ou bien d'une appréciation subjective – cf. *ressentir vs éprouver*). Certaines **composantes** du potentiel catégoriel du mot exercent sans doute une contrainte stricte sur la sélection de sa combinatoire, mais celles détectées dans le cadre de nos profils établis sur des bases probabilistes semblent être essentiellement des invitations à choisir telle perspective plutôt qu'une autre pour penser et exprimer une réalité donnée. Ces différences de point de vue instillées par le potentiel catégoriel correspondent certainement aux majeures nuances entre synonymes – idée brillamment défendue par Lafaye voici un siècle et demi. Précisons toutefois que, par nature, de telles invitations ou offres n'ouvrent pas la porte à une gamme infinie de possibilités, mais établissent un système d'ordre souvent fait de choix alternatifs, auquel le locuteur se soustrait difficilement. Dans cette perspective, la « liberté » combinatoire ressortit en principe à un processus de sélection, et non de création. En l'état actuel de nos recherches, il ne semble pas exclu que le nombre des options que présentent les potentiels catégoriels au niveau du lexique français soit très limité – éventuellement à quelques douzaines d'items³⁴ dans le cas des noms abstraits, auxquels appartiennent les noms traités ci-dessus. Le décryptage systématique du potentiel catégoriel propre à un mot et à ses acceptions, qui met en évidence les rapports entre leurs contenus et leurs vocations sur le plan combinatoire, ne manque pas d'intérêt pratique en lexicographie.

Le concept de 'potentiel catégoriel' lié à la combinatoire des mots nous semble susceptible d'éclairer d'un jour nouveau la « connexion ». La tentative de Tesnière de la rattacher à des données d'ordre cognitif a laissé ses lecteurs sur leur faim. Or, il a été montré ici dans quelle mesure les connexions correspondant à des combinaisons spécifiques reflètent ce que Tesnière appelle « catégories ».

Sur le plan méthodologique, nous accédons au type d'appréhension de la réalité que nous fournit un mot seulement par l'analyse de ses emplois stéréotypés. Mais qu'est-ce qui précède quoi sur le plan des réalités cognitives : la structure linguistique ou l'expérience et l'interprétation du référent ? Cette question est des plus traditionnelles au sein de notre vieille discipline, et déjà Humboldt..., n'est-ce pas ? Certes. À notre époque, il est d'usage de traiter de tels problèmes – analogues à celui de l'œuf et de la poule – selon les principes épistémologiques de l'« émergence », *i. e.* de l'apparition de nouvelles propriétés dans un système devenant plus complexe par l'interaction de ses facteurs (*bootstrapping*). L'histoire de la combinatoire des mots serait-elle donc faite, au moins partiellement, de complexifications cognitives croissantes (ou bien du contraire, selon les domaines) ? D'une part, de ce que nous avons appelé des clarifications catégorielles, d'autre part, d'obscurcissements ? Même si la réalité historique devait s'avérer plus différenciée encore,³⁵ la question mérite réflexion. Tout, ou presque, reste à faire en la matière.

Références bibliographiques

- Anscombre, J.-C. (1995). Morphologie et représentation événementielle : le cas des noms de sentiment et d'attitude. *Langue Française*, 105, 40-54.
- Anscombre, J.-C. (2001). Le rôle du lexique dans la théorie des stéréotypes. *Langages*, 142, 57-76.
- Bally, C. (1909). *Traité de stylistique française*, vol. 1. Paris : Klincksieck.
- Bischof, B.-B. (2007). *Französische Kollokationen diachron. Eine korpusbasierte Analyse*. Thèse, Université de Stuttgart [http://elib.uni-stuttgart.de/opus/volltexte/2008/3418/pdf/Franzoesische_Kollokationen.pdf].
- Blumenthal, P. (2009). Les noms d'émotion : trois systèmes d'ordre. In Novakova, I., & Tutin, A. (s. l. d.), *Le lexique des émotions et sa combinatoire*, Grenoble : Presses Universitaires de Grenoble (à paraître).

- Blumenthal/Diwersy/Mielebacher = Blumenthal, P., & Diwersy, S., & Mielebacher, J. (2005). Kombinatorische Wortprofile und Profilkontraste. Berechnungsverfahren und Anwendungen. *Zeitschrift für romanische Philologie*, 121, 49–83. [<http://www.romanistik.uni-koeln.de/home/blumenthal/publications/wortprofil-zrph121.pdf>].
- Carel, M. (2001). Argumentation interne et argumentation externe au lexique : des propriétés différentes. *Langages*, 142, 10–21.
- Chabrolle-Cerretini, A.-M. (2008). *Histoire d'un concept linguistique : la vision du monde de Wilhelm von Humboldt*. Lyon : ENS Editions.
- Coseriu, E. (1952). *Sistema, norma y habla*. Montevideo : Universidad de la República.
- CRISCO = *Dictionnaire des synonymes* [<http://elsap1.unicaen.fr/cgi-bin/>].
- Croft/Cruse = Croft, W., & Cruse, D. A. (2004). *Cognitive Linguistics*. Cambridge : Cambridge University Press.
- DCM, 2007 = Le Fur, D., e. a. (2007). *Dictionnaire des combinaisons de mots. Les synonymes en contexte*. Paris : Le Robert.
- Ducrot, O. (2001). Critères argumentatifs et analyse lexicale. *Langages*, 142, 22–40.
- Edelman, G. M. (2000). *Biologie de la conscience*. Paris : Jacob.
- Encyclopædia Universalis. DVD Version 11.00*, Paris : Encyclopædia Universalis.
- Flaux, N., & Van de Velde, D. (2000). *Les noms en français, Esquisse de classement*. Paris : Ophrys.
- François/Manguin = François, J., & Manguin, J.-L. (2006). *Dispute théologique, discussion oiseuse et conversation téléphonique : les collocations adjectivo-nominales au cœur du débat*. *Langue française*, 150, 50–65.
- Frantext = CNRS, & ATILF (2008). *Base textuelle Frantext*. Nancy : UMR CNRS-Nancy 2 [<http://atilf.atilf.fr/categ.htm>].
- Gil, F. (2005). Catégories. In *Encyclopædia Universalis. Grand dictionnaire de la psychologie*. Collectif. Nouvelle édition (2002). Paris : Larousse.
- Gross, G. (1996). *Les expressions figées du français*. Paris : Ophrys.
- Halliday/Matthiessen = Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An Introduction to Functional Grammar*. London : Arnold.
- Hausmann, F. J. (2007). Collocations, phraséologie, lexicographie. Études 1977-2007 et Bibliographie. Éditées par Elke Haag (Wortschatz Lernen Französisch – WLF, 4). Aachen : Shaker.
- Hoey, M. (2005). *Lexical Priming. A new theory of words and language*. London/New York : Routledge.
- Humboldt, W. v. (2003). Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts. Über die Sprache. Wiesbaden : Fourier.
- Lafaye, P.-B. (1869 [1858]). *Dictionnaire des synonymes de la langue française*. Paris : Hachette.
- Lalande, A. (1972). *Vocabulaire technique et critique de la philosophie*. Paris : PUF.
- Leeman, D. (1998). *Les circonstants en question(s)*. Paris : Kimé.
- Libet, B. (1999). Do We Have Free Will ?. *Journal of Consciousness Studies*, 6, No. 8/9, 47–57.
- Littré = *Dictionnaire Le Littré. Texte intégral sur Cédérom (Windows 95/98/Me)* (2002). Paris : Hachette éducation.
- Luhmann, N. (1990). *Amour comme passion*. Paris : Aubier.
- Lyons, J. (1977). *Semantics. 2 volumes*. Cambridge : CUP.
- Martin, R. (1983). *Pour une logique du sens*. Paris : PUF.
- Mejri, S. (1997). *Le figement lexical. Descriptions linguistiques et structuration sémantique*. Tunis : Publications de la Faculté des Lettres de la Manouba.
- Mel'čuk/Clas/Polguère = Mel'čuk, I./Clas, A./Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve : Duculot.

- Mel'čuk, I. (1998). Collocations and Lexical Functions. In Cowie, A. P., *Phraseology : theory, analysis, and applications*. Oxford/New York : Oxford University Press.
- Ménard, Y. (1976). *Manuel du français du moyen âge*. Bordeaux : Sobodi.
- Neveu, F. (2004). *Dictionnaire des sciences du langage*. Paris : Colin.
- Petit Robert = Rey, A., & Rey-Debove, J. (2004). Le Petit Robert. Dictionnaire de la langue française. Paris : Le Robert.
- Prévost, S. (2007). Adverbiaux temporels et structuration textuelle au 15^{ème} siècle. In Vanderheyden, A., & Mortelmans, J., & de Mulder, W., & Venckeleeer, T. (éds.), *Texte, Codex, Contexte : Actes du XIII^e colloque sur le moyen Français*, Turnhout : Brepols, 95–108 [cf. <http://www.lattice.cnrs.fr/IMG/doc/prevost-biblio15.doc>].
- Rastier, F. (2000). De la sémantique cognitive à la sémantique diachronique : les valeurs et l'évolution des classes lexicales. *Mémoires de la Société de Linguistique de Paris*, t. IX, Leuven : Peeters, 135–164.
- Rastier, F. (2001). *Sémantique et recherches cognitives*. Paris : PUF.
- Rastier, F. (2006). Doxa et lexique en corpus – pour une sémantique des 'idéologies'. In Dietrich, W., e. a., *Lexikalische Semantik und Korpuslinguistik*, Tübingen : Narr, 251–279.
- Rey, A. (2008). *De l'artisanat des dictionnaires à une science du mot*. Paris : Colin.
- Schapira, C. (1999). *Les stéréotypes en français : proverbes et autres formules*. Paris : Ophrys.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford : Oxford University Press.
- Sinclair, J. (2004). *Trust the Text*. London/New York : Routledge.
- Tesnière, L. (1969). *Éléments de syntaxe structurale*. Paris : Klincksieck.
- TLFi = CNRS, & ATILF. *Trésor de la Langue Française Informatisé (TLFi)*. Nancy : UMR CNRS/Université Nancy 2 [<http://atilf.atilf.fr/frantext.htm>].
- Tobler-Lommatzsch = Tobler, A., & Lommatzsch, E. (1925–2002). *Altfranzösisches Wörterbuch*. Stuttgart : Steiner.
- Tutin, A., & Novakova, I., & Grossmann, F., & Cavalla, C. (2006). Esquisse de typologie des noms d'affect à partir de leurs propriétés combinatoires. *Langue Française*, 150, 32–49.
- Verlinde, S., & Selva, T. *DAFLES. Dictionnaire d'apprentissage du français langue étrangère ou seconde* [<http://www.kuleuven.be/dafles/acces.php?id=>].
- Vigier, D. (2005). Les adverbiaux praxéologiques détachés en position initiale et leur portée. In Vigier, D., & Terran, E. (éds.), *Les adverbiaux cadratifs et l'organisation des textes (Verbum, XXVII, 3)*, Nancy : Presses Universitaires, 293–312.
- Violi, P. (2001). *Significato ed esperienza*. Milano : Bompiani.
- Wildgen, W. (2008). *Kognitive Grammatik. Klassische Paradigmen und neue Perspektiven*. Berlin : Walter de Gruyter.

¹ Terme dont la définition proposée par Hausmann (2007) semble appelée à faire école ; cf. Rey, 2008, chap. 7.3.1., Gross, 1996 et Mejri, 1997 sur le figement.

² Soit entre lexèmes (éventuellement des « collocations »), soit entre lexèmes et morphèmes grammaticaux (« colligations »). Cf. Croft/Cruise, 2004 : 249, qui précisent aussi le modèle du continuum.

³ Cette idée de Humboldt passe aujourd'hui pour l'une des bases de la linguistique cognitive (Wildgen, 2008 : 11 ; cf. Chabrolle-Cerettini, 2008).

⁴ Dans la perspective du modèle lexique-grammaire qui est la nôtre, les différences entre ces deux types de catégories s'estompent. Mais, même indépendamment de cette théorie particulière, on peut constater avec Gil (2005) que « les catégories sont des notions < stratégiques > ; susceptibles d'usages divers, elles circulent entre le syntactique et le sémantique et témoignent d'une pensée constructive. »

⁵ Une méthode totalement différente et complémentaire de la nôtre permettant de conclure de la combinatoire des mots à leurs caractéristiques sémantiques s'intéresse non pas au degré de spécificité d'une combinaison, mais à sa grammaticalité (cf. Flaux & Van de Velde, 2000).

⁶ Cf. Lalande, 1972, sous **LIBERTÉ (F), ARBITRE (libre), INDÉTERMINATION, INDIFFÉRENCE**. La problématique linguistique ressemble à celle du libre arbitre telle qu'elle est discutée depuis les années 1980 sous l'influence des résultats expérimentaux obtenus par le neurobiologiste Libet (1999), qui tendent à limiter fortement la portée du libre arbitre, sans nier totalement son existence.

⁷ Cf. Anscombre, 2001. Notre description des combinatoires spécifiques d'un mot de base pourrait offrir la possibilité de confronter à l'épreuve des faits statistiques les aperçus, extrêmement intéressants en soi, de la théorie de l'argumentation dans la langue (ADL) telle qu'elle a été développée par Carel et Ducrot (cf. Ducrot, 2001). On observera dès maintenant que de légères modifications du contexte du mot de base (par exemple, *relancer le débat* au lieu de *lancer un débat*) peuvent faire basculer l'argumentation dans un tout autre sens.

⁸ Cf. Sinclair, 1991, chap. 8 (« idiom principle ») et Sinclair, 2004 : 29 : « Complete freedom of choice, then, of a single word is rare. » La « semantic prosody » de Sinclair (2004 : 34) représente une partie de ce que nous attribuons à la « doxa ». Mel'čuk (1998 : 28), qui ne semble guère intéressé par l'approche probabiliste, ne pose pas dans les mêmes termes les problèmes de « freedom of selection » et « freedom of combination ».

⁹ Il semble presque superflu de préciser qu'*amour* a déjà fait l'objet d'une kyrielle d'analyses linguistiques, littéraires, sociologiques, etc. ; cf. Luhmann, 1990.

¹⁰ Combinaison spécifique dans le corpus *Frantext.romans* (correspondant à 230 romans tirés de la base *Frantext* 1950–2000, 16,4 millions de mots), que nous exploiterons par la suite (section 4.). Dans le même corpus, *dans* + *amour* se situe au-dessous du seuil de spécificité. Pour les combinaisons entre *amour/amitié* et *en/dans* en tant qu'« adverbiaux praxématiques », cf. l'étude excellemment documentée de Vigier 2005, section 1. Utilisant une autre grille statistique (et d'autres corpus) que Vigier, nous sommes cependant davantage sensible aux différences distributionnelles entre les deux noms. Pour les valeurs de la préposition *en*, cf. Leeman, 1998.

¹¹ Puisque le voisinage de *dans l'amour* et d'*en amour* se révèle assez semblable dans différents types de textes, nous nous sommes permis de constituer, pour cette enquête, un grand corpus thématiquement assez hétérogène d'environ 256 millions de mots (*Frantext.romans* ; *L'Est Républicain*, 2002 ; *Le Figaro*, 2002 ; *Le Monde*, 1999, 2000, 2002 ; *Sud-Ouest*, 2002 ; *Encyclopædia Universalis*, 2005) ; c'est seulement ainsi que nous avons pu obtenir un nombre statistiquement suffisant d'attestations : 229 occurrences d'*en amour* et 491 occurrences de *dans l'amour*.

¹² Cette fonction syntaxique est également accessible à *en amour* (cf. *transformer en amour*).

¹³ C'est en général le sens d'*amour* employé absolument (cf. *Petit Robert*, sous **amour** I.3.), emploi qui s'accorde par ailleurs avec la syntaxe habituelle de la préposition *en*.

¹⁴ Dans les termes de Lyons (1977, chap. 11.3), *dans l'amour* renvoie en général à une « second-order entity » (situation ou phénomène évènementiel, partant relativement concret), alors que *en amour* peut se rapprocher d'une « third-order entity » (phénomène abstrait, sphère du savoir).

¹⁵ Nous faisons évidemment abstraction ici de locutions comme *prendre en haine*.

¹⁶ Défini si bien par Lalande (1972) comme « partie de l'Univers du discours à laquelle s'appliquent une idée, une relation, une fonction, une faculté » (sous **DOMAINE**).

¹⁷ Reprécisons que selon les contextes, les emplois d'*amour* dans *en amour* et *dans l'amour* peuvent, mais ne doivent pas avoir le même référent.

¹⁸ L'emploi de la préposition *en* au sens de 'en matière', 'en ce qui concerne' existait dès l'ancien français (Ménard, 1976 : 283 ; Tobler-Lommatzsch sous **en**, t. III.1, p. 173) ; il perpétue l'une des valeurs de la préposition latine *in*.

¹⁹ Cf. Vigier, 2005, section 2. et Prévost, 2007 : 2 : « Nous faisons l'hypothèse que, plus un élément est autonome, plus il a un pouvoir cadratif important pour ce qui suit. ». Par 'autonomie', nous entendons ici tout d'abord la non-dépendance par rapport à un autre mot de la phrase (analysée selon le modèle dépendanciel) ; sur le plan du sens, l'autonomie correspond à la possibilité de concevoir le référent du mot antéposé comme une réalité (situation, domaine, point de vue) existant même en dehors du fait que rapporte la phrase. Critère formel d'autonomie : grammaticalité en tant que cadre d'une interrogation. Sont « autonomes » *en amour* et à *Paris* dans *En amour/À Paris, que faut-il faire ?* – à la différence de *souvent* ou *rapidement* (**Souvent*/**Rapidement, que faut-il faire ?*).

²⁰ Selon les critères développés par Martin, 1983 : 64–74 pour les noms.

²¹ Verbes régulièrement paraphrasés l'un par l'autre dans les dictionnaires (cf. *Petit Robert*, sous **éprouver** et **ressentir**). Leur possibilité d'emploi à titre de verbes supports devant des noms en fonction de complément d'objet est souvent considérée comme indice du fait que ces noms désignent des émotions ou des sentiments (cf. Tutin, e. a., 2006 : 32).

²² Même conception dans le *TLFi* (sous **ressentir** I. B.) : « Éprouver vivement dans son âme ou dans son esprit l'effet d'une cause extérieure. »

²³ L'article de Lafaye (1869, *Supplément*, p. 283) sur *sentir/éprouver* ne correspond pas à la situation en français actuel. Sur la concurrence entre *éprouver* et *ressentir* depuis l'ancien français, cf. Bischof, 2007, section 7.3.2.2.

²⁴ Comme le montre le menu « recherche avancée » du *Petit Robert électronique* (« texte intégral », « rechercher dans le plan »), la mention « sens subjectif » apparaît 7 fois, « sens objectif » 13 fois dans le dictionnaire – sans que la signification exacte de ces mentions soit toujours évidente.

²⁵ Au double sens d'« émanant du référent du sujet » et de « contenant une appréciation qualitative ».

²⁶ Pour le modèle mathématique de ce calcul, cf. Blumenthal/Diwersy/Mieiebacher, 2005 ; le taux 1 équivaldrait à l'identité de la combinatoire, le taux 0 signifierait l'absence d'éléments communs.

²⁷ Verbes rapprochés dans l'une de leurs acceptions (la première indiquée par l'article **ressentir**) dans Littré et le *Petit Robert* ; en revanche, le *TLFi* ne renvoie à *sentir* que dans la partie historique de l'article **ressentir**.

²⁸ Tant dans la réalité de la langue que dans les discussions entre linguistes après une conférence sur le sujet...

²⁹ Sur le plan psychologique, cette stéréotypie correspond à la prégnance telle que la conçoivent les gestaltistes : « Force, et par suite stabilité et fréquence d'une organisation psychologique privilégiée, parmi toutes celles qui sont possibles. » (définition par P. Guillaume, citée par le *Petit Robert*, sous **prégnance**). Cf. *Grand Dictionnaire de la psychologie*, 2002, sous **PRÉGNANCE** à propos de la « prégnance empirique, qui caractérise des bonnes formes significatives et découle de l'expérience acquise ». Dans le même ordre d'idées, mais en englobant la perception (« perception sémantique »), Rastier (2001 : 221) établit un parallèle entre les principes de la *Gestalt* et ceux présents dans l'isotopie, condition préalable à toute affinité sémantique entre mots associés sur l'axe syntagmatique.

³⁰ Coseriu (1952) entend par « norme » la réalisation usuelle du système de la langue dans une communauté linguistique (et non pas la norme prescriptive).

³¹ Cf. Hoey (2005, chap. 2), qui subordonne par ailleurs, dans ce domaine, le linguistique au psychologique en faisant des collocatifs un sous-ensemble des mots liés au mot de base par « association sémantique », qui est due à l'« amorçage associatif » (*priming*, 2005 : 24). Rastier (2001 : 218) considère que « l'amorçage sémantique et l'amorçage associatif ne font qu'un ».

³² On pourrait parler de « **profil catégoriel** » en se restreignant aux catégories sous-jacentes à un profil combinatoire, effectivement mobilisées dans le corpus correspondant. Aux différents potentiels définis ci-dessus, il conviendrait d'adjoindre le « **potentiel argumentatif** » (Carel, 2001 : 10) si d'ultérieures recherches viennent à confirmer le fondement empirique de cette notion.

³³ Il existe au moins une troisième dimension pouvant déterminer la combinatoire, dimension que la place attribuée à la présente contribution nous interdit de développer : les données du domaine référentiel (Halliday/Matthiessen, 2004 : 39, 577), objets du **savoir encyclopédique** et points de mire du savoir scientifique (frère ennoblé de la doxa ?); exemple : les collocatifs les plus spécifiques de *cellule* dans des journaux comme *Le Monde* sont presque tous des adjectifs (souvent relationnels) reflétant les découvertes biochimiques sur les types de cellules. Entre ces trois champs de forces, potentiel catégoriel, doxa et savoir encyclopédique, qui agissent sur l'« amorçage associatif » (cf. note 31), on relèvera non pas des cloisons étanches, mais de nombreuses passerelles et interactions.

³⁴ En ce qui concerne les noms d'affect, les « paramètres sémantiques » établis par Tutin, e. a. (2006 : 36-49) seraient d'excellents candidats à ce statut. Nous entendons restreindre notre vision componentielle de la signification à la méthode d'analyse (cf. Violi, 2001 : 82), sans nous engager dans une théorie postulant la réalité des traits sémantiques. Cette réserve ne nous dispense pas de la tâche de nous interroger sur les rapports entre le potentiel catégoriel et les nombreuses autres composantes hypothétiques que diverses théories inscrivent dans la signification du mot (cf. Violi, 2001 : 98 et Mel'čuk/Clas/Polguère, 1995 : 111 sur les « connotations »). L'une des spécificités des composantes du potentiel catégoriel réside dans le fait qu'il s'agit de vecteurs démontrables uniquement de façon

inductive à travers leurs effets sur la combinatoire et que leur détection nécessite le plus souvent des outils statistiques. Elles se distinguent en cela des « fonctions lexicales » syntagmatiques de Mel'čuk (1998) qui se présentent, dans la théorie Sens-Texte, comme une taxinomie à laquelle contribue grandement la part hypothético-déductive. En effet, cette théorie reconnaît dans chaque collocation l'instanciation d'une des fonctions lexicales, conçues comme des liens de nature universelle et des outils conceptuels de base.

³⁵ Cf. le modèle d'analyse diachronique esquissé par Rastier, 2000.