



La question de la géométrie du lexique

Alain Polguère

► To cite this version:

Alain Polguère. La question de la géométrie du lexique. SHS Web of Conferences, 2016, Actes du 5e Congrès Mondial de Linguistique Française (CMLF 2016), 27, pp.01002. 10.1051/shsconf/20162701002 . hal-01344814

HAL Id: hal-01344814

<https://hal.science/hal-01344814>

Submitted on 12 Jul 2016

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

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

La question de la géométrie du lexique

Alain Polguère^a

ATILF & Université de Lorraine, 44 av. de la Libération, BP 30687, 54063 Nancy Cedex, France

Résumé. Cet article s'intéresse à la structure globale du lexique, en tant que module fonctionnel de la langue, et à ses possibles représentations formelles. Nous posons tout d'abord le problème de la représentation de la connaissance lexicale, en expliquant l'importance d'adopter une métaphore géométrique du lexique qui soit adaptée à l'étude lexicologique et à ses diverses applications. Cela nous amène préciser la nature même du lexique qui doit faire l'objet d'une telle géométrisation en revenant sur la notion souvent galvaudée de lexique mental et en distinguant deux types de lexiques mentaux – le lexique logique et le lexique physiologique –, dont seul le premier est notre objet d'étude. Après avoir examiné les principales métaphores courantes du lexique logique, nous présentons un modèle formel plausible, sur lequel s'appuie notamment le projet du Réseau Lexical du Français : il s'agit des Systèmes Lexicaux, dont la charpente structurale est constituée par les fonctions lexicales Sens-Texte.

Abstract. The question of the lexicon geometry. This paper questions the global structure of natural language lexicons and their formal representation. We first examine the general problem of representing lexical knowledge, highlighting the importance of adopting a geometric metaphor for lexicons that is suitable for lexicology and its applications. This leads us to specify the very nature of lexicons that have to undergo geometrization by reanalyzing the overused notion of mental lexicon. We distinguish two instances of mental lexicons – the logical lexicon vs. the physiological lexicon –, identifying the former as our sole object of study. After examining the most common metaphors of the logical lexicon, we present so-called Lexical Systems as plausible formal models. Lexical Systems are used in the context of the French Lexical Network project and their structural backbone is based on Meaning-Text lexical functions.

^a alain.polguere@univ-lorraine.fr

1 Enjeux de la représentation formelle du lexique

1.1 Problème abordé

Le discours grand public sur un objet naturel se distingue du discours scientifique notamment par le fait que ce dernier pose de façon primordiale le problème de la représentation formelle de l'objet étudié. La linguistique structurale n'est en cela pas différente des autres sciences dites *dures* et elle a donné naissance à un nombre significatif de formalismes représentationnels pour l'étude de la sémantique, de la syntaxe, de la morphologie et de la phonétique des langues. On notera cependant que la plupart de ces formalismes visent à représenter les énoncés ou parties d'énoncés, c'est-à-dire les produits du système structural de chaque langue. Le problème de la formalisation des deux grands modules linguistiques fonctionnels eux-mêmes – lexique et grammaire – a généré beaucoup moins de propositions. On trouve bien entendu des formalismes permettant d'encoder les règles grammaticales ou lexicales, prises individuellement, mais pratiquement aucune proposition sur la représentation de la structure globale de ces systèmes de règles que sont le lexique et la grammaire. L'approche qui nous semble remarquable de ce point de vue est celle de la linguistique systémique fonctionnelle de M. A. K. Halliday, qui conceptualise formellement la lexico-grammaire (*lexicogrammar*) sous la forme d'un gigantesque graphe réunissant tous les réseaux systémiques de choix expressifs (*system networks*) que la langue offre au Locuteur : « the grammar¹ is seen as a network of interrelated meaningful choices » [1 : p. 31]. Cette perspective sur la langue permet notamment de visualiser la grammaire dans les études linguistiques ou, du moins, des extraits du graphe lexico-grammatical global.

Dans cette section, nous posons la question de la représentation formelle du lexique, comme problème central de la lexicologie et, plus globalement, de la science linguistique. Il s'agit d'un problème étrangement éludé par la lexicologie théorique, bien qu'il soit présumé résolu dans nombre d'applications potentielles de cette dernière : en psycho- et neuro-linguistique, en didactique des langues, en traitement automatique des langues, etc. Nous commencerons par poser la question toute naïve de l'existence d'une forme du lexique, puis nous récapitulerons les principaux postulats sur lesquels reposent les propositions que nous ferons par la suite.

1.2 Le lexique a-t-il une forme ?

La question que nous posons ci-dessus – *Le lexique a-t-il une forme ?* – est en fait un raccourci et ne doit pas être considérée littéralement. Le lexique n'étant pas une entité visuelle, mais informationnelle, il est absurde de demander si littéralement il possède une forme. Par contre, il est important de s'interroger pour savoir si un modèle du lexique peut être visualisé par une entité formelle possédant certaines propriétés qui feront de cette visualisation du lexique une **métaphore** productive en lexicologie, ainsi que dans d'autres disciplines pouvant avoir recours à un modèle du lexique. Notons que le lexique est une entité tellement abstraite et diffuse qu'on tend à le concevoir inconsciemment comme n'étant pas associable à une visualisation formant un tout cohérent. Le lexique serait ainsi quelque chose d'informe échappant à la métaphorisation visuelle.

La notion de métaphore scientifique visuelle est fondamentale. Nous sommes tout particulièrement intéressé par les éventuelles propriétés géométriques des visualisations qui, pour reprendre les termes de R. Hoffman [2 : p. 411–414], s'appliquent au lexique en tant que Topique (*Topic*) et servent de Véhicule (*Vehicle*) au raisonnement métaphorique, les

propriétés formelles du Véhicule permettant de générer de nouvelles entités théoriques en relation avec le Topique.

A metaphor can suggest new theoretical entities or concepts, or reinterpretations of old ones. A metaphor can suggest new structural interrelations or similarities between the theoretical entities, that is, new categories of entities or properties. Not all of the concepts in a theory are translatable into observables, only some are. The metaphor may suggest new functional relations, possibly of a specified mathematical form. [2 : p. 411]

Cette façon de procéder est une évidence pour les sciences dont les objets d'études sont des entités du monde physique, comme en physique ou en chimie [3], et cela est d'autant plus facile que l'entité physique est directement perceptible visuellement, au besoin au moyen de microscopes oculaires. Les choses sont moins évidentes lorsque l'on s'attaque au très petit ou au très grand cependant, et les représentations visuelles des atomes, molécules ou des galaxies utilisées par les scientifiques paraissent être beaucoup plus des métaphores visuelles que des transcriptions de formes véritables. Dans le domaine des entités non physiques ou non visuelles, les sciences se donnent aussi des modes de visualisation, qui sont des supports commodes, bien que non strictement nécessaires, pour certains types de raisonnements. Ainsi, en physique, on pourra représenter une force par un vecteur dessiné dans un espace euclidien ; en mathématique, on utilise aussi la métaphore de l'espace géométrique euclidien pour visualiser les applications algébriques par des droites ou des courbes. On pourrait multiplier les exemples de ce type. Ce sur quoi nous voulons mettre l'accent, c'est que l'intérêt premier de telles métaphores n'est pas de rendre les notions correspondantes plus compréhensibles dans le contexte de l'enseignement des disciplines concernées, mais bien de servir de Véhicules hoffmaniens au raisonnement. Plus particulièrement, comme nous l'avons indiqué dans [4], ces visualisations constituent des métaphores correspondant à des notions *primaires* – par exemple, la visualisation d'une application algébrique – servant de support à des notions *dérivées* – par exemple, la notion d'application algébrique dite *linéaire*².

Concentrons-nous maintenant sur le cas de la linguistique et sur les géométrisations les plus connues dans cette science : celles de la structure syntaxique des énoncés représentée au moyen de graphes arborescents – les fameux arbres syntaxiques. La modélisation arborescente en syntaxe de constituants a servi de support à l'élaboration d'une profusion de notions métaphoriques dérivées de la métaphore primaire de l'arbre pendant l'époque glorieuse de la Grammaire générative-transformationnelle ; par exemple, dans la thèse de J. Ross [6], les notions de *tree pruning* (élagage d'arbre) et de *island* (île), fondées sur la structure des arbres syntaxiques. Il faut cependant se défier de la facilité avec laquelle il est possible de dériver des métaphores à partir d'une visualisation. Une métaphore dérivée semble souvent mettre au jour une vérité tout simplement parce qu'elle se construit naturellement à partir de métaphores primaires et le grand cimetière des publications linguistiques est ainsi peuplé de métaphores sans lendemain parce que plus faciles à produire qu'à rentabiliser. En revanche, certaines métaphores vont s'avérer particulièrement puissantes, parce que, dans leur simplicité, elles permettent de rendre compte élégamment d'un nombre important de faits linguistiques fondamentaux. Pour illustrer un tel cas, nous nous attarderons un instant sur une métaphore dérivée en syntaxe de dépendance [7] : la (non-)projectivité des arbres syntaxiques [8–10]. Il a été démontré que lorsque l'on représente la structure syntaxique d'un énoncé non sous forme d'un graphe d'imbrication de constituants, mais sous forme d'un graphe de dépendances entre mots-formes, la figure arborescente obtenue possède dans la plupart des cas une propriété géométrique remarquable : la projection verti-

cale des sommets (= nœuds) du graphe sur la phrase linéarisée ne présente aucun croisement entre liens de dépendance syntaxique et axes de projection. Ce phénomène est illustré ci-dessous dans la Figure 1, exemple (a), où les lignes continues représentent les dépendances syntaxiques entre mots-formes et les lignes en pointillé représentent les projections des sommets de l'arbre syntaxique sur la phrase linéarisée.

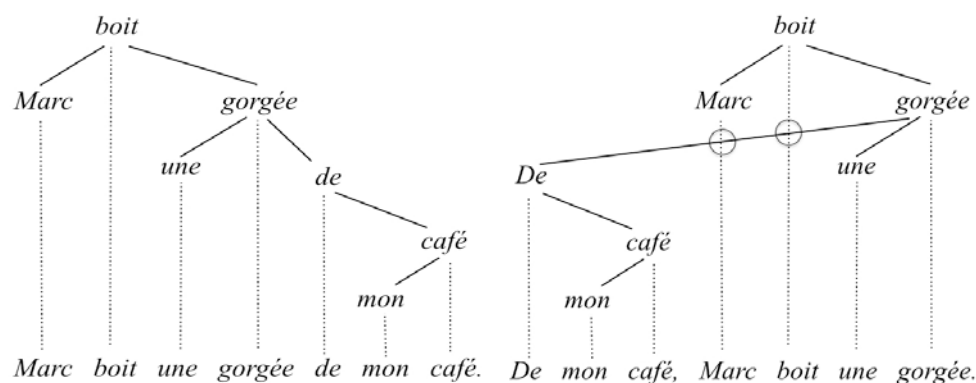


Fig. 1. (a) Phrase projective

(b) Phrase non projective

Les énoncés dont l'arbre syntaxique possède cette propriété de non-croisement des liens de dépendances et des projections des sommets sont ainsi appelés *énoncés projectifs*, et la propriété correspondante *projectivité*. Il existe bien entendu dans toute langue la possibilité de générer des énoncés non projectifs – cf. Figure 1, exemple (b) ; ceux-ci sont cependant marqués et le principe universel gouvernant la linéarisation des structures syntaxiques est, par défaut, la projectivité. La propriété géométrique de non-projectivité donne chair à un cas de non-iconicité grammaticale où, comme en (b) ci-dessus, la linéarisation des mots-formes contredit leur appartenance aux syntagmes. Elle est particulièrement intéressante, car elle permet d'expliquer des phénomènes comme la difficulté relative que peuvent avoir les apprenants à maîtriser certaines structures de phrase, en production ou en compréhension ; elle trouve aussi des applications potentielles en traitement automatique des langues [11].

Ainsi qu'indiqué au début de cet exposé, les métaphores visuelles en linguistique – comme l'arbre syntaxique et ses métaphores dérivées – concernent essentiellement la visualisation des structures d'énoncés. La même démarche est cependant applicable à la langue elle-même, en tant que système de règles. Si la langue est constituée minimalement d'un lexique et d'une grammaire, se poser la question de la forme de ces entités, c'est ouvrir la possibilité de proposer des métaphores visuelles, géométriques, qui vont servir de support au raisonnement linguistique.

Le lexique, s'il n'a pas une forme véritable en tant qu'entité informationnelle, peut avoir un modèle qui possède une « forme logique », déterminée par son organisation interne, et c'est cette forme logique qui est visualisable au moyen d'une entité géométrique. La **géométrisation** du lexique peut ainsi être résumée de la façon suivante :

- (1) Lexique → Modèle [forme logique] → Visualisation [forme géométrique]

En géométrisant le lexique, on pourra disposer d'un modèle prédictif, où les propriétés géométriques de la représentation métaphorisent et anticipent des propriétés fonctionnelles de l'objet étudié. De plus, la forme géométrique peut être également exploitée pour enrichir

la forme logique des modèles. En lexicologie, les bonnes représentations géométriques du lexique seront celles qui permettront de construire des notions dérivées explicatives et productives. De plus, si la géométrie postulée pour le lexique est une bonne métaphore de ce dernier, les propriétés géométriques des structures permettront non seulement de modéliser l'information lexicale de façon statique, mais aussi de rendre compte de sa nature dynamique [12], telle qu'elle se manifeste dans l'évolution diachronique du lexique, dans l'apprentissage du vocabulaire, etc.

1.3 Présupposés

Tout au long de cet exposé, nous tentons de présenter des arguments justifiant la représentation formelle particulière que nous avons adoptée depuis plusieurs années pour la modélisation du lexique – celle des Systèmes Lexicaux [13], représentation qui sera introduite dans la section 4. Nous devons cependant aussi nous appuyer sur des présupposés, tenus pour acquis. Nous résumons brièvement ci-dessous les trois principaux présupposés sur lesquels repose le modèle des Systèmes Lexicaux.

Premièrement, nous présupposons bien évidemment qu'une visualisation formelle du lexique est possible, et même souhaitable. Nous considérons que le lexique, dans son ensemble, peut se géométriser d'une façon qui nous procure un support formel adapté à l'étude lexicologique théorique et descriptive, ainsi qu'à ses diverses applications.

Deuxièmement, nous adoptons le système notionnel de la Lexicologie Explicative et Combinatoire [14] et, plus généralement, de la linguistique Sens-Texte [15]³. Nous considérons donc, en accord avec les principes de notre approche théorique de référence, que l'objet d'étude de la lexicologie n'est pas le signe linguistique lexical – *grosso modo*, le **mot-forme** –, mais bien une entité encore plus abstraite⁴ : le **lexème**, compris comme étant un regroupement de mots-formes que seule distingue la flexion ; par exemple, le lexème adjectival ODORANT, qui regroupe les quatre mots-formes *odorant*, *odorants*, *odorante* et *odorantes*. Chaque lexème est structuré autour d'un sens spécifique, et les lexèmes qui (i) possèdent une intersection de sens significative et (ii) s'expriment par les mêmes signifiants se regroupent au sein de **vocables** polysémiques ; par exemple, le vocable polysémique CAMEMBERT contient les lexèmes CAMEMBERT I 'type de fromage...' et CAMEMBERT II 'graphique circulaire visualisant des données statistiques...'. L'**unité lexicale** – aussi appelée **lexie** – est donc bien le lexème, et non le mot-forme ou le vocable. C'est elle qui sera de fait l'unité de modélisation du lexique. Notons, pour être complet, qu'il existe un autre type d'unité lexicale que le lexème : la **locution**. Le lexique de toute langue contient en effet un nombre très important de syntagmes phraséologiques non compositionnels – 'SE LA COULER DOUCE', 'POMME DE DISCORDE'... –, qui sont des unités lexicales à part entière et qui doivent donc être modélisés en tant que tels, au même titre que les lexèmes. Les entités de base de nos visualisations du lexique seront donc des **représentations de lexies** : lexèmes ou locutions.

Finalement, le dernier de nos présupposés fondamentaux concerne la nature du lexique : celui-ci n'est aucunement une collection, une liste, un répertoire de lexies. C'est un système relationnel, une sorte de réseau social de lexies interconnectées. Cette vision du lexique est une évidence pour certains, sans qu'il soit besoin de faire des recherches poussées en lexicologie ou en neurolinguistique. La vision du lexique en tant que collection de mots – héritée du modèle des dictionnaires (voir plus bas, section 3.1) – a cependant la vie dure et il convient de tout faire pour l'éradiquer définitivement. Voir le lexique comme un réseau de lexies a des conséquences importantes sur la conceptualisation de la notion même de lexie

et des propriétés que chaque lexie possède – c’est-à-dire, des règles qui lui sont associées. Développons brièvement ce point.

Il a mainte fois été souligné en lexicologie que la lexie ne ferait sens qu’au sein des énoncés. Pour parler en termes sémiotiques, la sémiose⁵ lexicale n’aurait lieu que lorsque le signe lexical est enchâssé dans un énoncé, ce que A. Wierzbicka exprime ainsi :

In actual fact, neither lexical items nor grammatical categories have any meaning: only utterances have meaning, because only utterances express thoughts. To speak about the meaning of lexical items or grammatical categories is elliptical: strictly speaking we can only talk about the contribution that particular elements make to the meaning of the whole utterance. [17 : p. 164]

Cette perspective sur la lexie trouve écho dans les recherches en psycholinguistique, notamment, où il est d’usage de mettre l’accent sur l’importance cruciale du contexte d’énonciation dans le fonctionnement sémiotique lexical : « Up to the very moment of usage words are but virtual entities. They are static and without life, because divorced from context. » [18 : p. 356]. Il faut cependant faire attention de ne pas verser dans une vision de contextualisme extrême, qui conduit notamment à nier l’existence même du sens comme propriété inhérente des lexies et à affirmer : « les mots prennent leur sens en contexte ». Les mots ne prennent pas leur sens en contexte : ils possèdent un sens, même s’ils ne **font** sens qu’en contexte [19]. Mais nous irons plus loin : les lexies font sens, non seulement dans les énoncés, mais aussi dans le lexique, car aucune lexie ne s’y trouve isolée. Chaque lexie appartient, dans le lexique, à un système de connexions paradigmatiques et syntagmatiques au sein duquel elle est un véhicule de sens. Le Locuteur accède aux unités lexicales, les manipule, etc., en naviguant dans un système de relations qui activent sémiotiquement chaque lexie de façon au moins aussi efficace que ne le font, dans un énoncé donné, les lexies formant un contexte d’énonciation. La notion de lexie prise isolément est un mythe, qu’il s’agisse de l’étude du contenu lexical d’une phrase ou du lexique dans son ensemble. Les lexies sont donc bien vivantes, non inertes, dans le lexique, pour peu que ce dernier soit effectivement modélisé comme un système connectif, au moyen d’une lexicologie **explicative et combinatoire**. Tant que l’on présente le lexique comme une collection de mots en revanche, la véritable nature de la lexie et de l’information lexicale ne pourra être modélisée adéquatement.

Le lexique relationnel qui doit être l’objet d’une modélisation géométrisable peut être qualifié de *lexique mental*, car il doit être en adéquation avec la structure de l’information lexicale manipulée par les Locuteurs. La notion de lexique mental demande cependant à être éclaircie, notamment dans le contexte du rapport que la lexicologie peut entretenir avec des disciplines connexes comme la psycholinguistique et la neurolinguistique.

2 Les deux faces du lexique mental

Nous faisons ci-dessous l’hypothèse qu’il existe une notion générale de lexique mental, notion qui peut s’instancier dans deux entités bien distinctes : le lexique logique et le lexique physiologique. Nous commençons par présenter la notion générale de lexique mental (2.1), puis nous abordons la distinction lexique logique vs physiologique (2.2), en expliquant pourquoi seul le dernier est – dans l’état actuel de la connaissance scientifique – l’objet d’étude de la lexicologie.

2.1 Du côté de la psycholinguistique

La notion de lexique mental est très directement associée aux recherches en psycholinguistique et, plus récemment, en neurolinguistique [20, 21]. Nous tenons tout d'abord à éclaircir la notion en question, selon la perspective originelle de la psycholinguistique, afin de mieux nous l'approprier dans le contexte de la lexicologie.

La discussion qui suit repose sur une distinction que nous établissons entre concepts et notions [22]. Nous appelons *concept* toute entité informationnelle pouvant être impliquée dans des processus inférentiels⁶ et *notion* une association « concept-terme » donnée. Une notion est donc un concept lexicalisé par un support terminologique et, dès qu'il est question de notion et non de concept, le choix terminologique joue un rôle primordial. Dans ce qui suit, lorsque nous parlons de la notion de lexique mental, nous considérons un concept donné en tant qu'il est désigné par le terme *lexique mental* (ou son équivalent anglais *mental lexicon*). La distinction entre concept et notion est cruciale en science, puisque l'activité scientifique ne peut s'effectuer sans le recours à une délimitation rigoureuse et à une désignation terminologique des concepts manipulés. Seule la proposition d'un terme donné permet, par la définition du terme qui est élaborée par le chercheur, de circonscrire le concept correspondant et de le rendre opératoire.

De nombreuses publications de psycholinguistique font remonter l'introduction du terme *mental lexicon* (lexique mental) à l'article de 1966 de R. Oldfield intitulé « Things, words and the brain » [24]⁷. Une première remarque que l'on peut faire à ce propos est qu'il s'agit d'une assertion parfaitement fausse. Nulle part dans ce texte n'apparaît le terme *mental lexicon*, pas plus que le terme *mental*, sa source dérivationnelle *mind* et, même, *lexicon*. Il est intéressant de regarder quel est exactement le propos du texte en question. Tout d'abord, le propos de [24] est ancré dans l'examen du problème de l'accès à l'information lexicale : comment un Locuteur – un sujet parlant – accède-t-il aux unités lexicales dont il a besoin pour s'exprimer, sachant que R. Oldfield évalue la connaissance lexicale moyenne d'un individu éduqué à 75 000 lexèmes ? (On notera que le point de départ est ici la production linguistique, l'encodage, et non l'accès à l'information lexicale en situation d'interprétation, le décodage.) En fonction de ce problème d'accès, se pose la question de savoir quel est le mode d'encodage de l'information lexicale dans la mémoire du Locuteur, encodage qui doit permettre un accès quasi instantané aux unités lexicales. Dans la suite de son article, R. Oldfield se concentre sur les expérimentations exploitant des tâches de nommage, pour identifier la façon dont se fait l'accès à l'information lexicale. Puis, en cours d'exposé (section « The Brain part in naming »), R. Oldfield aborde le sujet du support physique de l'information lexicale : le cerveau lui-même. Oldfield ne traite donc aucunement d'une *notion* de lexique mental, entité informationnelle dénommée par le terme correspondant ; sa cible n'est pas l'esprit (*mind*) du locuteur, mais son cerveau, comme l'indique le titre même de l'article. Il traite cependant bien d'un concept proche de celui aujourd'hui encapsulé dans la notion de lexique mental et il est légitime de considérer que les travaux de psycholinguistique expérimentale du type de [24] sont à l'origine de la notion qui nous intéresse⁸.

Le fait que la paternité du terme *mental lexicon* soit erronément attribuée à [24], alors même que ce terme n'apparaît pas dans ce texte et que c'est bien la réalité **physiologique** de l'encodage dans le cerveau et non la représentation mentale qui est visée, est révélateur d'une équivalence qui est fréquemment établie entre lexique dans l'esprit et lexique dans le cerveau. Il s'agit pourtant, comme nous allons l'expliquer, de deux notions distinctes, bien qu'intimement liées.

Voyons comment le lexique mental est caractérisé au début des années quatre-vingt par G. Miller, l'initiateur du projet WordNet (cf. section 3.2 ci-dessous), dans une perspective plus fonctionnelle que physiologique :

The mental lexicon is needed most immediately to produce and understand sentences, so lexical information must be represented in a manner compatible with what people know about the syntax and semantics of sentences in their language. [27 : p. 61]

La citation ci-dessus pose deux caractéristiques fonctionnelles du lexique mental, qui méritent d'être considérées du point de vue des contraintes qu'elles imposent à une modélisation formelle du lexique en question.

Tout d'abord, le lexique mental est une ressource organisée pour servir de support aux activités linguistiques de base que sont la production et la compréhension des énoncés. Le modèle formel du lexique mental ainsi caractérisé sera donc spécialisé dans sa finalité. Son design doit répondre à cette exigence première d'adéquation avec les processus linguistiques, même si un modèle formel informatisé de ce type pourrait être utilisé pour des tâches aussi diverses que rechercher l'orthographe d'un mot, lemmatiser ou étiqueter sémantiquement des corpus, classer des textes en fonction de leur contenu, etc.

Ensuite (et G. Miller présente cette seconde propriété comme une conséquence de la première), l'organisation du lexique mental est compatible avec la connaissance qu'ont les locuteurs de la grammaire de la langue et qui leur permet de produire des énoncés sémantiquement et grammaticalement adéquats⁹. En conséquence, un modèle formel du lexique mental doit posséder une structure logiquement organisée en fonction de la nécessaire interconnexion (pour ne pas dire *interpénétration*) des règles lexicales et grammaticales.

Les caractéristiques fonctionnelles proposées ci-dessus imposent sur le lexique mental des contraintes d'adéquation au comportement linguistique du Locuteur que l'on peut qualifier de **logiques**, c'est-à-dire qu'il s'agit de contraintes de compatibilité avec la réalité des inférences effectivement effectuées par le Locuteur en situation de parole. Ce comportement du Locuteur est en tout premier lieu observable dans ses productions langagières orales et écrites. C'est-à-dire qu'elles relèvent d'un observable que l'on peut trouver dans les corpus de textes, dans le discours spontané et, même, dans une activité linguistique introspective du Locuteur. Un modèle formel du lexique mental conçu en fonction de ces contraintes pourra ainsi servir de support à des inférences du type (2), où « $\therefore$ » est l'opérateur d'inférence logique.

- (2)
- | | |
|--------------|---|
| | <i>éprouver</i> + DÉT + <i>désir</i> |
| & | <i>éprouver</i> + DÉT + <i>angoisse</i> |
| & | <i>éprouver</i> + DÉT + <i>amour</i> |
| & | ... |
| $\therefore$ | <i>éprouver</i> + DÉT + <i>doute</i> |

Les inférences de ce type jouent un rôle capital dans le cadre d'une lexicologie du lexique mental et doivent pouvoir être conduites directement à partir des propriétés formelles du modèle lui-même, même s'il est plutôt d'usage actuellement de considérer exclusivement l'extraction de telles règles à partir de l'analyse des corpus [28]. Or, les traitements statistiques de corpus rendent compte des productions effectives, non du potentiel expressif ! C'est pourquoi, par exemple, une collocation comme *éprouver* + DÉT + *doute* inférée en (2) est absente de la liste des cooccurrences de *doute* proposée dans le « Dictionnaire de cooccurrences » de la suite logicielle Antidote [29]. Sa fréquence d'apparition dans les

corpus utilisés est vraisemblablement trop peu significative pour faire émerger la collocation en question comme une régularité de combinatoire lexicale du français¹⁰. De même, la consultation du corpus FrWac – aspiration de l’Internet français effectué en 2008 et qui contient plus de 1,6 milliard de mots-occurrences [31] – retourne seulement 89 occurrences de la suite lemmatisée *éprouver + doute* (dans une fenêtre de 3 mots-occurrences), ce qui correspond à une fréquence négligeable de 0,1 par million. En comparaison, on trouve 21 429 occurrences de la séquence *avoir + doute* (13,3 par million). Contrairement à la plupart des modèles statistiques, un apprenant type du français pourra effectuer une inférence comme (2) ci-dessus, même s’il n’a jamais été confronté à la collocation peu fréquente *éprouver + DÉT + doute*. De la même façon, un modèle du lexique mental pourra, même si la collocation en question en est absente, la générer à la base des motifs collocationnels qui y sont déjà présents.

Le lexique mental que nous visons est donc cette entité informationnelle qui est un lieu de calcul compatible avec les inférences effectuées par le Locuteur.

2.2 Lexique mental logique vs physiologique

Les contraintes d’adéquation logique qui viennent d’être évoquées sont distinctes des contraintes d’adéquation **physiologiques**, qui imposent que le lexique mental soit, dans son contenu et sa structure, compatible avec le comportement physiologique du Locuteur. Ce comportement s’observe non dans l’analyse des énoncés produits ou productibles, mais dans la mesure des différents paramètres d’exécution de tâches linguistiques dans des situations expérimentales : temps de réponse à des stimuli linguistiques ou conceptuels, mouvements oculaires, activité neuronale dans les diverses aires du cerveau, etc.

Nous proposons donc de distinguer explicitement deux facettes du lexique mental :

1. le lexique tel qu’il est structuré dans la langue et dans l’esprit du Locuteur – désigné par le terme *lexique logique* ;
2. le lexique tel qu’il est structuré dans le cerveau du Locuteur – désigné par le terme *lexique physiologique*.

Chacune des entités en question peut être considérée comme une instanciation particulière du lexique mental.

La structure du lexique mental physiologique doit être compatible avec les diverses mesures faites lors d’expérimentations psycho- ou neurolinguistiques et il n’y a aucune raison de penser qu’elle est **identique** à la structure du lexique mental logique. Les deux lexiques doivent cependant bien entendu pouvoir être mis en correspondance, et il faut même admettre qu’une compréhension scientifique véritablement complète du lexique mental ne pourra se faire que par la mise en relation et la synthèse des deux types d’entités, logique et physiologique. Cependant, **dans l’état actuel des connaissances en linguistique et en neurologie**, la notion de lexique mental regroupe deux objets d’études distincts et, donc, deux activités scientifiques distinctes. Établir cette distinction ne signifie pas que nous la cautionnons. Les chercheurs des disciplines concernées doivent œuvrer vers une plus grande interconnexion disciplinaire, jusqu’à ce qu’une jonction véritable puisse s’opérer. Nous espérons que les réflexions présentées ici apportent une contribution en ce sens.

La Figure 2 ci-dessous résume ce qui vient d’être dit à propos de la distinction entre lexique logique et lexique physiologique.

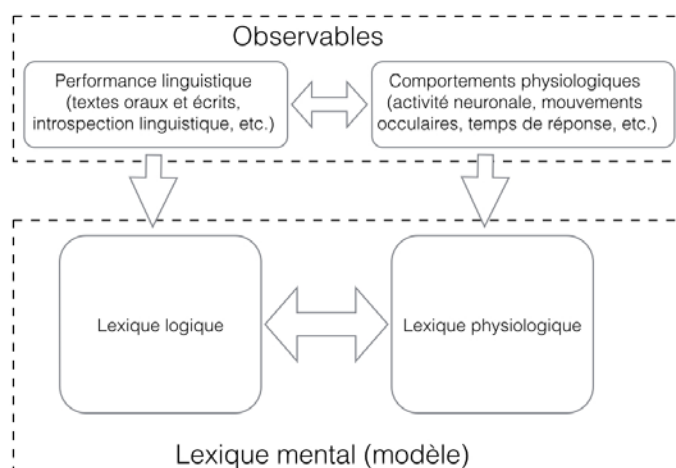


Fig. 2. Distinction entre modèles du lexique mental logique et physiologique

Comme il a été souligné ci-dessus, la construction d'un modèle du lexique physiologique relève d'une démarche avant tout expérimentale, basée sur le recours à la mesure de l'activité neuronale, aux expériences de psycholinguistique, à l'observation de pathologies langagières, etc. Une telle démarche donne accès à la façon dont l'information lexicale est physiologiquement encodée dans le cerveau et conduit à la construction de modèles qui cartographient ce mode d'encodage. Cependant, elle ne donne pas nécessairement accès à ce qu'est l'information en question. Prenons un exemple très spécifique, mais tout à fait représentatif des recherches en neurolinguistique pour illustrer notre propos. Blakely et coll. présentent dans [32] les résultats d'une étude visant à identifier avec une relative précision les localisations corticales activées lors de tâches de verbalisation de phonèmes. Pour simplifier un peu grossièrement, on pourrait dire qu'il s'agit d'expérimentations visant à localiser les phonèmes dans le cerveau. De tels travaux pourraient parvenir à cartographier la connaissance phonologique, mais ils s'appuient sur les résultats de travaux linguistiques beaucoup plus théoriques et descriptifs, non expérimentaux, qui ont mis en évidence la notion même de phonème et ont permis d'identifier pour chaque langue un ensemble de phonèmes donnés. La structure phonémique même de la langue est le produit d'une démarche logique, qui offre à l'étude de la réalité physiologique du langage le matériau notionnel sur lequel cette étude peut se développer. Bien entendu, on peut s'attendre à une rétroaction de l'étude de la réalité physiologique du langage, qui permettra de corriger, revoir, peut-être révolutionner les modèles logiques. Quoi qu'il en soit, il semble raisonnable de postuler qu'une découverte de la structure logique du lexique mental, distincte de sa structure physiologique, n'est pas une tâche absurde. On peut même défendre l'idée qu'elle doit produire des résultats qui permettront d'alimenter l'étude du lexique mental physiologique. Ainsi, dans le cas qui nous intéresse tout particulièrement de la structure relationnelle du lexique, à quoi sert de prétendre localiser expérimentalement les mots et leurs relations dans le cerveau, si l'on n'a pas auparavant établi avec un minimum de rigueur ce qu'il faut entendre par *mot* et par *relations entre mots* ?

Les acquis remarquables des travaux contemporains en neurolinguistique expérimentale poussent malheureusement certains à développer une philosophie assez étrange de l'étude scientifique de la langue :

Ultimately, no linguistic structure should be assumed to exist unless the model can offer an explanation of how it comes about and is implemented in the human brain in the first place. [33 : 505]

L'objet d'étude de la neurolinguistique actuelle n'est pas la langue, au sens saussurien, mais l'information linguistique en tant qu'elle est encodée dans le cerveau. Cependant, la citation ci-dessus va bien au-delà que cette simple vérité. Elle présuppose trois choses : 1) la langue n'est pas une réalité sociale, mais une réalité biologique, 2) la langue n'existe pas en tant qu'entité autonome des cerveaux où elle réside informationnellement et 3) les observables de la langue sont les activités neuronales mesurées par des appareils. Étrangement, cette remise en question de l'existence indépendante sociale (non physiologique) de la langue semble possible alors que personne ne songe à remettre en question l'existence indépendante de l'algèbre, par exemple. Pourtant, on peut tout à fait analyser comment nous encodons dans notre cerveau les concepts algébriques et comment ceux-ci sont physiologiquement mobilisés dans le cadre d'opérations mathématiques que nous effectuons. Il existe certainement une « algèbre mentale » physiologique, une algèbre – avec son lexique et ses règles combinatoires et inférentielles – telle qu'elle est encodée dans nos neurones. Cette algèbre mérite d'être étudiée, mais cela ne justifierait pas que l'on nie l'existence de l'algèbre immanente afin de mener l'étude de l'algèbre telle qu'elle est encodée dans le cerveau de chaque individu. La langue n'est pas différente des autres systèmes informationnels manipulés par les humains et, tout comme eux, elle existe et doit être étudiée du point de vue de sa structure logique.

La lexicologie, en tant qu'étude du lexique logique, n'est pas une discipline primordialement expérimentale, mais descriptive, qui ancre sa méthodologie de modélisation des règles lexicales de chaque langue par l'étude des faits langagiers¹¹. L'application la plus immédiate d'une géométrisation du lexique se trouve donc dans l'activité lexicographique, c'est-à-dire dans la construction de modèles des lexiques des langues. Pour conclure la présente section, nous allons brièvement justifier la focalisation de la lexicologie sur le lexique logique, non physiologique.

2.3 Le lexique mental logique comme seul objet d'étude de la lexicologie

Il existe deux raisons pour lesquelles nous pensons que la cible de la lexicologie **actuelle** est le lexique mental logique et non physiologique.

Tout d'abord, comme nous l'avons mentionné plus haut, on ne peut mesurer que quelque chose dont on sait qu'il existe ou dont on a fait l'hypothèse de l'existence. L'étude du lexique mental physiologique s'appuie donc sur une théorie du lexique logique. Ensuite, la langue n'est pas nécessairement organisée en fonction de la façon dont les individus la stockent dans leur cerveau et, même, l'utilisent procéduralement, mais plutôt en fonction de grands principes formels. Pour justifier cette affirmation qui pourrait sembler un rien métaphysique, nous nous permettrons une analogie. Imaginons qu'un être d'une autre galaxie, où l'informatique est inconnue, se donne pour mission de comprendre ce qu'est un ordinateur personnel dans sa structure en observant la façon dont les individus lambda utilisent ces machines. Il est certain qu'il arriverait à un modèle de l'ordinateur bien différent de celui conçu par les informaticiens et les ingénieurs. Il en va un peu de même avec le lexique. Si l'on fait l'hypothèse que le lexique, comme la langue, existe comme entité informationnelle abstraite et sociale, il est certain que cette entité sociale ne possède pas une structure identique à celle des connaissances lexicales encodées électriquement dans le cerveau de tel ou tel individu, en fonction de contraintes notamment physiologiques.

Nous croyons à la réalité d'un lexique mental logique et nous croyons qu'une des clés de la modélisation de ce lexique est l'identification de bonnes métaphores géométriques, ou plus généralement mathématiques, qui en font des objets calculables et des supports de recherche. Ce sont de tels lexiques mentaux logiques qui sont à la fois les objets d'étude et les buts descriptifs de la lexicologie théorique et descriptive. Cette dernière ne doit aucunement prétendre construire des modèles physiologiquement déterminés. Il est certain, cependant, qu'un modèle logique qui n'éclaire en rien la modélisation physiologique sera ultimement inadéquat. On peut penser que le lexique est logiquement organisé d'une façon qui est compatible avec son encodage physiologique et qu'en cherchant à découvrir cette organisation logique, on œuvre par là même dans le sens de l'exploration de la réalité physiologique.

Récapitulons ce qui vient d'être dit. Nous admettrons le fait que le lexique dont nous devons déterminer la géométrie est bien le lexique mental logique et que ce dernier doit avant tout être construit pour rendre compte des deux activités linguistiques fondamentales de production et compréhension langagières, ainsi que de l'instabilité inhérente à la connaissance lexicale (changements linguistiques diachroniques, apprentissage, perte de compétence, etc.). Cette structure ne doit pas être déterminée en fonction du comportement physiologique du Locuteur et de l'encodage neuronal de l'information lexicale qui, eux, sont en tout premier lieu des indices de la structure du lexique physiologique. Néanmoins, la structure du lexique logique doit être compatible (au sens le plus général du terme) avec celle du lexique physiologique et les deux types d'études doivent être connectés.

L'objet de la modélisation a été circonscrit et il est temps d'aborder le problème de la formalisation proprement dite. Avant de présenter les Systèmes Lexicaux en tant que métaphores formelles du lexique mental logique, nous passerons brièvement en revue les principaux modèles proposés pour rendre compte de la forme globale du lexique mental.

3 Métaphores courantes de la forme du lexique

Nous allons tout d'abord examiner le cas du dictionnaire papier ou électronique, produit de l'activité lexicographique et modèle du lexique, en nous concentrant sur ses particularités formelles qui en font une piètre métaphore du lexique mental (3.1). Nous examinerons ensuite une autre métaphore structurale du lexique : l'arbre taxinomique (3.2), puis nous présenterons ce qui nous semble être l'approximation formelle du lexique mental la plus plausible actuellement à notre disposition : le réseau lexical (3.3).

3.1 Le lexique en tant que dictionnaire

La métaphore traditionnellement proposée pour le lexique mental dans une perspective de vulgarisation est celle du dictionnaire :

The history of science is littered with examples of analogies that do not work. Often, they are simply inappropriate, simply wrong, or simply confusing. But even when inappropriate, they can prove useful. For instance, it is not unnatural to think of our knowledge about the words in our language as residing in some sort of dictionary. [34 : 65]

Pour rendre compte du fonctionnement de la voie lexicale, l'image la plus appropriée est donc celle d'une collection de dictionnaires, dont les pages s'ouvrent successivement à mesure que nous accédons à l'information. Notre cerveau héberge une bibliothèque en plusieurs volumes, depuis le guide de l'orthographe jusqu'au manuel de prononciation et au grand dictionnaire encyclopédique. [35 : 72–73]

Alors que nombre de chercheurs pensent que la métaphore dictionnaire fonctionne jusqu'à un certain point, nous voulons au contraire défendre l'idée qu'il est vital d'abandonner cette référence au dictionnaire lorsque l'on parle du lexique mental, quels que soient les contextes où la notion est mise en jeu. Le simple fait de postuler que la connaissance lexicale est un **dictionnaire** mental permet de se débarrasser de la question de la structure même du lexique, ou du moins de l'ignorer, pour passer directement aux problèmes procéduraux d'activation de la connaissance lexicale, ce qu'illustre parfaitement la citation ci-dessous.

Si l'on accepte le principe que les connaissances lexicales sont représentées sous la forme d'un dictionnaire mental, la question se pose de savoir comment l'on accède aux « entrées » de ce dictionnaire lors de la production et de la compréhension des mots. [36 : 22]

Nous n'acceptons pas l'idée que le lexique puisse être décrit comme un dictionnaire mental et nous refusons de filer la métaphore en parlant d'*entrées*, même entre guillemets. Cette métaphore est dangereuse, car elle nous écarte de ce qui nous pensons être une structure plausible du lexique et elle a donc toutes les chances de nous faire faire fausse route dans l'appréhension des phénomènes lexicaux. Or, la question de la structure du lexique est bien première : se poser la question de savoir comment on accède aux informations lexicales présuppose que l'on dispose au moins d'une hypothèse viable sur la nature et la structuration de cette information lexicale.

Rappelons que ce que l'on nomme *dictionnaire* est en tout premier lieu un livre, c'est-à-dire un long texte structuré en parties et sous-parties enchâssées. Même sous leur forme numérique, les dictionnaires de la fin du XX^e siècle et du début du XXI^e siècle ne sont pas libérés de la linéarité inhérente au texte imprimé. Ainsi, pour ce qui est de la matérialisation informatique, le système prédominant d'encodage des ressources lexicales reste le langage XML et ses extensions, c'est-à-dire un système de balisage de la structure hiérarchique interne des textes [37].

La métaphore dictionnaire textuelle détermine d'autant plus notre vision du lexique qu'elle est entrée dans la langue : en français, par exemple, avec des expressions comme *de A à Z*, *être dans le dictionnaire*, *dictionnaire vivant*, *raier du dictionnaire* et, bien sûr, *dictionnaire mental*. Ce dernier terme, notamment, impose à notre esprit une vision caricaturale du lexique mental, illustrée dans la Figure 3.



Fig. 3. Métaphore caricaturale (et contre-productive) du dictionnaire mental

Le message que cette image cherche à véhiculer est que le fait d'utiliser le terme *dictionnaire* pour désigner le lexique mental n'a absolument aucun pouvoir d'évocation quant à la structure relationnelle multidimensionnelle du lexique en question, que la métaphore nous force à concevoir comme un simple répertoire de « mots ». Ce terme bloque toute visualisation du lexique en laissant croire que le lexique étudié par les lexicologues et les lexicographes **est** le dictionnaire-livre, alors que le dictionnaire a toujours été un pis-aller, une structure formellement lourde, sans souplesse et à bien des égards en contradiction avec la nature relationnelle du lexique.

G. Miller, dans un article précisément intitulé « Dictionaries in the mind » [38], résume assez bien l'incompatibilité entre la structure des dictionnaires et celle du lexique mental de la façon suivante :

The way a lexicographer goes about writing a dictionary is not at all the way a psychologist would go about constructing a psychological theory. And it is certainly not the way a child goes about learning a vocabulary. [38 : 172]

Cette assertion est justifiée si l'on se réfère uniquement à une lexicographie des dictionnaires-livres, la seule qui a véritablement existé jusqu'à présent. Or, cette situation peut et doit changer. La lexicographie, si l'on ne veut pas qu'elle disparaisse et que disparaisse avec elle le vaste savoir accumulé sur la nature des faits lexicaux au cours des siècles d'activité lexicographique, doit évoluer d'une discipline visant la construction de dictionnaires-livres vers une discipline visant des modèles véritablement multidimensionnels, dont les articles de dictionnaire ne seront qu'un des produits dérivés possibles [39, 40].

Pour conclure sur la question du modèle des dictionnaires-livres, soulignons les deux faiblesses principales qui le rendent impropre à la représentation du lexique mental.

Premièrement, il s'agit d'un modèle rigide, contraint par sa structure textuelle linéaire, qui impose une vision statique du lexique mental. Ce dernier, cependant, doit être représenté comme un système dynamique, car le lexique est en transformation constante. De plus, un modèle formel du lexique mental doit aussi être un outil d'expérimentation approprié pour rendre compte de la dynamique présente dans les processus cognitifs tels que l'apprentissage et la perte de maîtrise des règles lexicales.

Deuxièmement, le dictionnaire est une représentation qui est un piètre support pour le calcul, même dans le cas de dictionnaires aussi formalisés que le sont, par exemple, les *Dictionnaires explicatifs et combinatoires* [41]. De plus, le modèle dictionnaire présente la faiblesse de n'être pas un **diagramme**, au sens de C. S. Peirce [42], et bride donc toute possibilité de raisonnement diagrammatique [43] sur la structure relationnelle du lexique. Cette dernière est cachée dans ce que l'on appelle la *médiostructure* des dictionnaires : l'en-

semble des connexions lexicales impliquées par la mention de lexies au sein des articles lexicographiques d'autres lexies [44].

Ces faiblesses formelles distinguent le modèle dictionnaire textuel des deux familles de modèles de type graphe dont il va maintenant être question (en 3.2 et 3.3) et qui constituent de véritables tentatives de représentation multidimensionnelle du lexique mental.

3.2 Le lexique en tant que taxinomie

La métaphore du lexique mental non dictionnaire la plus connue est indiscutablement celle implantée dans la ressource lexicale WordNet¹², initiée dans les années quatre-vingt par G. Miller et inspirée de ses recherches antérieures en psycholinguistique [45]. WordNet, du fait de son histoire, possède un double visage et il est présenté par ses créateurs eux-mêmes aussi bien comme un modèle formel du lexique mental et un support pour la *psycholexicology* [45 : 236–238] que comme une base de données destinée, notamment, au traitement automatique des langues : « WordNet is an online lexical database designed for use under program control » [46 : 39].

Il est inutile de faire une énième présentation de WordNet et nous nous contentons d'énumérer ses principales caractéristiques formelles, en nous focalisant sur le WordNet de Princeton (lexique de l'anglais américain) et en nous intéressant tout particulièrement à la structure globale de la ressource.

- a) L'unité de **description** de WordNet est le *sense*¹³, qui correspond approximativement à notre notion de lexie. Cela signifie que WordNet rend bien compte de la polysémie, en distinguant les acceptions des vocables ; par exemple, les deux *senses* : *stream*#1 'cours d'eau' vs *stream*#2 'courant [de pensée, etc.]'.
- b) Néanmoins, l'unité de **structuration** est le *synset* : ensemble de lexies et expressions entretenant une relation de synonymie. Cela signifie que WordNet connecte de façon première les *synsets*, même s'il s'agit d'une structure hétérogène comportant aussi des connexions directes entre *senses*.
- c) Deux relations lexicales tissent l'ossature de WordNet : 1) la relation de synonymie entre *senses* définit leur appartenance aux *synsets* ; 2) la relation d'hyper-/hyponymie hiérarchise les *synsets*, des plus vagues sémantiquement aux plus spécifiques.
- d) Cette organisation modélise le lexique de façon hiérarchique afin de rendre compte d'une organisation des *senses* en classes : **il s'agit donc d'un modèle taxinomique du lexique mental.**
- e) Cette organisation est avant tout adaptée au lexique nominal, qui est le seul pour lequel la structuration hiérarchique fondée sur l'hyper-/hyponymie semble complètement opératoire. Les sous-lexiques des autres parties du discours majeures – verbes, adjectifs et adverbes – sont structurés de façon sensiblement différente, pour ce qui est des connexions entre *synsets* notamment.
- f) WordNet compartimentalise donc le lexique en parties du discours, certains liens transversaux entre parties du discours – dérivations morphologiques, par exemple [47, 48] – étant cependant intégrés.

La Figure 4 ci-dessous illustre la structure de WordNet avec un extrait de la hiérarchie nominale centrée autour du *synset* {*stream*#1, *watercourse*#2} 'cours d'eau'.

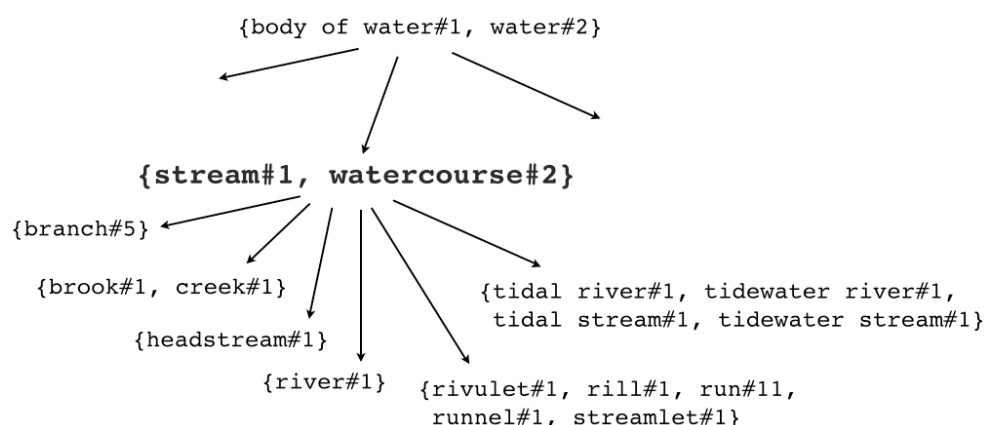


Fig. 4. Extrait de la hiérarchie nominale de WordNet

La structure de WordNet est séduisante à bien des égards, et nous relèverons trois de ses caractéristiques qui la rendent tout particulièrement pertinente dans le cadre d'une modélisation du lexique mental. Tout d'abord, cette structure est visualisable et permet une interprétation géométrique de l'organisation du lexique. WordNet est un graphe et, en visualisant ce graphe, on obtient une forme d'explication de l'organisation du lexique, contrairement à la métaphore textuelle du dictionnaire. Cette structure est aussi séduisante parce que, du fait de sa nature taxinomique, elle est ressentie comme compatible avec la modélisation des processus cognitifs relevant de la classification et de l'héritage de propriétés. Par exemple, le synset **{stream#1, watercourse#2}** de la Figure 4 permet de classer 26 *senses* de WordNet (les deux *senses* du synset lui-même, plus les *senses* des synsets hyponymes) qui partagent des propriétés sémantiques fondamentales puisqu'ils dénotent tous des masses d'eau en mouvement ; si l'on remonte d'un cran, au synset hyperonyme **{body of water#1, water#2}**, ce sont 362 *senses* qui se retrouvent classés en tant que dénotant des masses d'eau ; etc. Finalement, et c'est peut-être le point le plus important, la structure de WordNet semble légitime parce qu'elle se fonde sur deux relations lexicales qui sont effectivement fondamentales et perçues consciemment ou inconsciemment comme telles par tout un chacun : la synonymie et l'inclusion de sens.

Malgré tous ces atouts, nous considérons que le graphe de WordNet n'est pas une métaphore appropriée de la structure du lexique mental logique (aussi bien que physiologique) pour les trois raisons suivantes.

Tout d'abord, les modèles du lexique mental ne doivent pas, selon nous, être organisés comme des taxinomies, car il n'existe pas de principe premier de classification des lexies. De multiples classifications, fonctionnant de façon orthogonale, peuvent être appliquées aux lexiques des langues, selon la perspective que l'on désire adopter : classification par classes conceptuelles (*grosso modo*, les classes de synsets de WordNet), par champs sémantiques (ensemble de lexies sémantiquement liées même si leurs dénotations peuvent être très distinctes), par champs lexicaux (classes de vocables dont les lexies de base appartiennent à un champ sémantique donné)¹⁴, par parties du discours, par domaines conceptuels (terminologies), etc. Toutes ces classifications ont leur légitimité et leur utilité potentielle, mais aucune ne parvient à rendre compte de la logique globale de la structuration du lexique. La classification et la hiérarchisation des classes semblent mettre de l'ordre dans le désordre lexical, mais une telle approche est vite rattrapée par la réalité d'un lexique qui refuse de se faire mettre en boîtes et de se laisser ranger du grenier aux oubliettes des prisons

taxinomiques. Le lexique n'est pas une armée de mots, classifiés et hiérarchisés [50]. Toute entreprise lexicographique taxinomique est soit limitée à un aspect bien spécifique de l'information lexicale – comme dans les thésaurus –, soit destinée à voir se dénaturer graduellement l'élégance de sa structure initiale pour parvenir à rendre compte de la richesse et de la complexité des connexions lexicales.

Ensuite, nous n'adhérons pas au postulat de la compartimentalisation du lexique mental en parties du discours. Si la proximité sémantique doit être reflétée par une proximité topologique dans le modèle formel – et c'est un des postulats de WordNet que nous reprenons à notre compte –, le principe de connexion lexicale doit être fondé sur la paraphrase et non sur la stricte synonymie, parce que cette dernière n'admet qu'une équivalence intra partie du discours. *X eats Y*, *X's eating of Y*, *[X] eater of Y*, *[Y] food for X*, *[Y] edible by X*, etc., s'ils ne sont pas synonymes, sont construits sur le même matériau sémantique (ou presque) et sont en relation de **paraphrasage**, tel que conçu dans le cadre de la théorie Sens-Texte [51, 52]. Les lexies EAT, EATING, EATER, FOOD, EDIBLE, etc., doivent donc être topologiquement très proches dans un modèle formel diagrammatique du lexique mental, ce qui exclut absolument une compartimentalisation de ce dernier en parties du discours.

Finalement, les modèles taxinomiques ne pêchent pas simplement par le fait qu'ils donnent une part disproportionnée à la relation d'hyper-/hyponymie dans la structuration du lexique, au détriment des autres relations paradigmatiques. Ils ignorent ou marginalisent les relations syntagmatiques, alors que ces dernières jouent un rôle tout aussi important dans l'organisation de la connaissance lexicale. Chaque lexie est connectée aux lexies qui lui sont sémantiquement proches parce que celles-ci participent de l'expression de contenus identiques ou analogues. Elle est cependant aussi connectée aux lexies avec lesquelles elle possède des affinités combinatoires, notamment au sein d'expressions phraséologiques. Les deux types de connexions – paradigmatiques et syntagmatiques – relèvent d'ailleurs dans bien des cas de besoins communicatifs similaires. Par exemple, la connexion paradigmatique liant une lexie nominale comme ENQUÊTE à sa contrepartie verbale ENQUÊTER est fonctionnellement analogue à la connexion syntagmatique qui la lie au verbe support MENER ; toutes deux permettent l'expression verbale du sens 'enquête' :

(3) *X enquête sur Y* $\equiv$ *X mène une enquête sur Y*.

Un modèle plausible du lexique mental doit donc fusionner les liens lexicaux paradigmatiques et syntagmatiques selon un même principe de structuration.

Nous nous sommes ici focalisé sur WordNet, car il s'agit du produit d'une véritable étude lexicologique et lexicographique, remarquable par sa couverture et par son caractère novateur. Nos remarques peuvent cependant s'appliquer à tout modèle inspiré de présupposés taxinomiques analogues sur l'organisation de la connaissance lexicale. Nous allons maintenant caractériser le second type de modèle lexical de type graphe – le réseau lexical petit-monde (3.3) –, avant de passer à la présentation de son instanciation particulière dans les Systèmes Lexicaux, qui conclura notre étude.

d'arcs en regard du nombre de sommets¹⁵, les graphes petits-mondes sont tels que n'importe quel sommet y est relié à n'importe quel autre par un chemin relativement court (un petit nombre d'arcs). Leur topologie se caractérise bien sûr également par la présence de clusters, qui donnent à ce type de graphe sa structure particulière constituée d'une multitude de grappes de sommets – voir point précédent sur les espaces sémantiques dans les graphes lexicaux.

Illustrons le phénomène de petit-monde avec la lexie VOCABULAIRE. Celle-ci entretient des relations sémantiques étroites avec un grand nombre de lexies françaises (LEXIQUE, TERMINOLOGIE, MOT, LANGUE, etc.), qui définissent autour d'elle un espace sémantique, dont un extrait représentatif est présenté dans la Figure 6 ci-dessous.

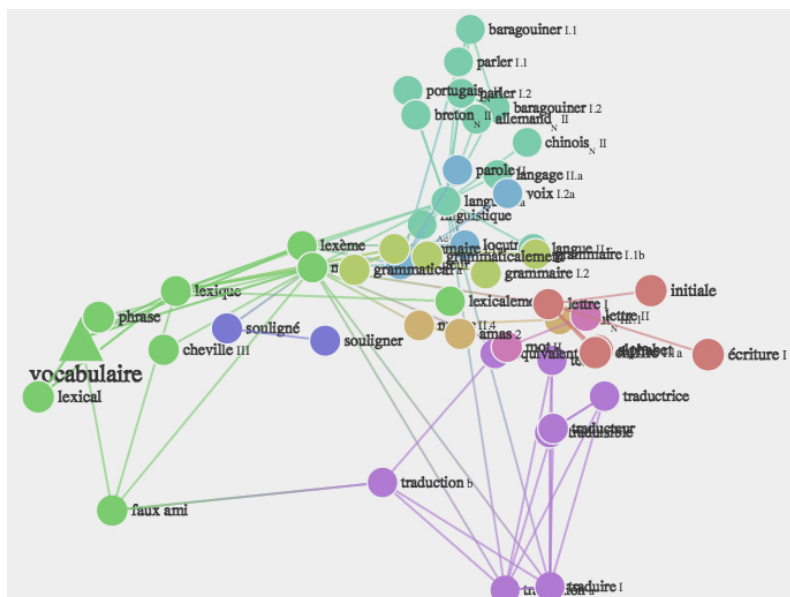


Fig. 6. Espace sémantique de VOCABULAIRE dans le Réseau Lexical du Français

Cependant, VOCABULAIRE, en plus d'entretenir ces relations de proximité sémantique, est aussi connectée presque directement à une lexie qui lui est sémantiquement étrangère comme DÉCORATION 1.2 [*murs ornés d'une décoration en mosaïque*], par le simple fait que, comme cette dernière, elle appartient au club relativement disparate des noms qui ont une affinité combinatoire avec l'adjectif métaphorique RICHE 11.2 (*vocabulaire riche ~ décoration riche*). Le monde lexical est bien petit et les lexies les plus éloignées sémantiquement, comme le sont VOCABULAIRE et DÉCORATION 1.2, peuvent facilement s'y rencontrer. Cet état de fait est visualisé dans la Figure 7 ci-dessous, où l'éloignement géométrique entre sommets représente un éloignement sémantique.



Fig. 7. Rencontre dans le petit-monde lexical

Le phénomène petit-monde des réseaux lexicaux n'est pas assuré seulement par l'existence de lexies intrinsèquement collocatives comme l'intensificateur **RICHE II.2**, qui établissent des ponts entre des aires sémantiquement éloignées du lexique. Il est aussi la résultante du fait plus général que toute langue possède un petit ensemble de lexies extrêmement vagues qui, par nature et contrairement à la moyenne, sont impliquées dans un très grand nombre de relations paradigmatiques ou syntagmatiques ; il s'agit, par exemple, des termes génériques, des primitifs et des molécules sémantiques [55], des collocatifs « jokers » comme *avoir, faire...* en tant que verbes supports ou *très, beaucoup...* en tant qu'intensificateurs, etc. Ces lexies à très haute connectivité, que nous appelons **lexies carrefours**, tissent des raccourcis dans les graphes lexicaux, accentuant ainsi le phénomène petit-monde. En général, ces lexies appartiennent à des vocables hautement polysémiques – des **vocables carrefours** –, dont la description est une priorité méthodologique dans le cadre de la construction lexicographique de réseaux lexicaux du fait de leur importance stratégique dans la topologie du graphe.

La présentation qui vient d'être faite de la notion de graphe petit-monde est assez simpliste, pour ne pas dire vulgarisante, mais nous espérons qu'elle suffit à faire percevoir la nature formelle bien particulière (i) de la modélisation du lexique mental sous forme de **réseau lexical** et (ii) de la géométrisation qui en découle. Notons que nous employons dorénavant le terme *réseau lexical* exclusivement pour dénoter les graphes lexicaux non taxinomiques de type petits-mondes.

Les contraintes topologiques que respectent les réseaux lexicaux sont bien plus faibles que celles des graphes taxinomiques examinés en (3.2). Ces contraintes existent cependant et rendent les réseaux lexicaux calculables, au sens où des inférences peuvent y être effectuées. On peut ainsi exploiter les contraintes topologiques qui s'y appliquent afin de modéliser des processus dynamiques, comme l'acquisition du vocabulaire [56, 57].

Il est insuffisant de postuler que le lexique mental logique possède une structure de graphe petit-monde. Il faut aussi proposer une hypothèse théorique sur la nature des entités qui le constituent (sommets et arcs du graphe) et explorer la validité de cette hypothèse par un travail lexicologique et lexicographique intensif sur les réseaux lexicaux. C'est tout le propos de la recherche sur les Systèmes Lexicaux, que nous allons maintenant présenter.

4 Systèmes Lexicaux

Nous procéderons en deux étapes pour présenter le modèle des Systèmes Lexicaux. Tout d'abord (4.1), nous allons décrire les principaux types formels de réseaux lexicaux, en expliquant notamment pourquoi il nous semble essentiel de viser en lexicologie les modèles fondés sur les graphes de lexies, plutôt que les graphes de vocables (aux sommets ambigus). Nous présenterons ensuite le modèle des Systèmes Lexicaux (4.2), en expliquant comment la notion s'est construite et tentant de démontrer l'intérêt des Systèmes Lexicaux en tant que modèles géométrisables du lexique mental.

4.1 Variété des réseaux lexicaux

Les réseaux lexicaux – au sens où nous l'entendons ici de graphes non taxinomiques (3.3) – sont de plus en plus utilisés en tant que modèles formels dans les études sur la structure des lexiques [12, 58]. Ils se situent dans le courant global de construction et d'exploitation des réseaux dits *sémantiques* [59] et un nombre croissant d'outils théoriques, formels et computationnels est disponible pour appuyer les travaux dans ce domaine. On trouve cependant

une grande diversité de réseaux lexicaux, qui se distinguent notamment par la nature des sommets et des arcs qui les constituent.

Pour ce qui est des sommets – c’est-à-dire, des entités connectées –, on peut considérer qu’il existe deux grandes familles de réseaux : les réseaux de vocables et les réseaux de lexies. Les premiers sont généralement des structures automatiquement construites, par compilation des données de ressources préexistantes, comme des dictionnaires ou, même, des corpus de textes traités statistiquement. Il s’agit de réseaux **de vocables**, car leurs sommets ne correspondent pas à des unités lexicales (des sens bien identifiés), mais à des « mots » généralement ambigus.

Un exemple très connu de ce type de réseau, pour le français, est le graphe du Dictionnaire Électronique des Synonymes ou DÉS¹⁶, issu de la base DicoSyn [60], elle-même construite par compilation des données de sept dictionnaires de synonymes. Ce graphe – parce que ses sommets sont ambigus – va nécessairement comporter de très nombreux chemins (parcours d’arcs) relevant plus de l’association d’idées que de la véritable navigation sémantique, bien qu’il soit fondé sur l’équivalence sémantique (la synonymie). On y trouve ainsi des chemins du type *sieste*—*somme*—*addition*, qui représentent des sortes de compactions de l’information lexicale véritable. Dans le cas du chemin donné ici en exemple, la situation dans le lexique français est la suivante :

- une acception du vocable SOMME¹ (cf. *faire un petit somme*) est un quasi-synonyme d’une acception (peut-être unique) du vocable SIESTE ;
- une acception d’un vocable SOMME² (cf. *faire une somme arithmétique*) – homonyme de SOMME¹ – est un quasi-synonyme d’une acception du vocable ADDITION (cf. *faire une addition arithmétique*).

Nous sommes donc en réalité en présence de deux systèmes de connexions disjoints. Les connexions véritables entre lexies étant fusionnées dans les graphes de vocables, seule l’analyse topologique de ces derniers permet de contourner le problème de l’ambiguïté, en identifiant les espaces sémantiques associés à chaque sommet lexical et en traitant un même sommet comme relevant alternativement de l’un ou l’autre de ces espaces sémantiques. Nous nous attarderons un instant sur deux analyses de ce type – analyse par identification de cliques¹⁷ [61] et analyse fondée sur la proxémie [62], qui donnent de très bons résultats.

On identifie seize cliques à partir du sous-graphe des synonymes du sommet *somme* dans le DicoSyn, avec les configurations suivantes :

- deux cliques contiennent le sommet *addition* – {*addition, compte, somme, total*} et {*addition, prix, somme*} ;
- une clique contient le sommet *sieste* – {*roupillon, sieste, somme, sommeil*} ;
- les deux sommets *addition* et *sieste* n’appartiennent simultanément à aucune clique.

Cette propriété topologique du graphe peut être interprétée comme une forme de modélisation de la déconnexion sémantique entre les sommets *sieste* et *addition* et, conséquemment, comme la présence d’une polysémie à l’intérieur du nœud *somme*.

L’analyse par proxémie – Figure 8 ci-dessous¹⁸ – permet quant à elle d’identifier à partir du même graphe synonymique de *somme* quatre espaces sémantiques (clusters) distincts, dont l’un – en bas à droite de la figure – relève du sens ‘addition’ et l’autre – en haut à droite – relève du sens ‘fait de dormir’. L’analyse proxémique est puissante puisque, en plus de calculer les espaces sémantiques, elle identifie la plus ou moins grande proximité sémantique entre ces espaces, représentée visuellement dans la Figure 8 par la distance géométrique.

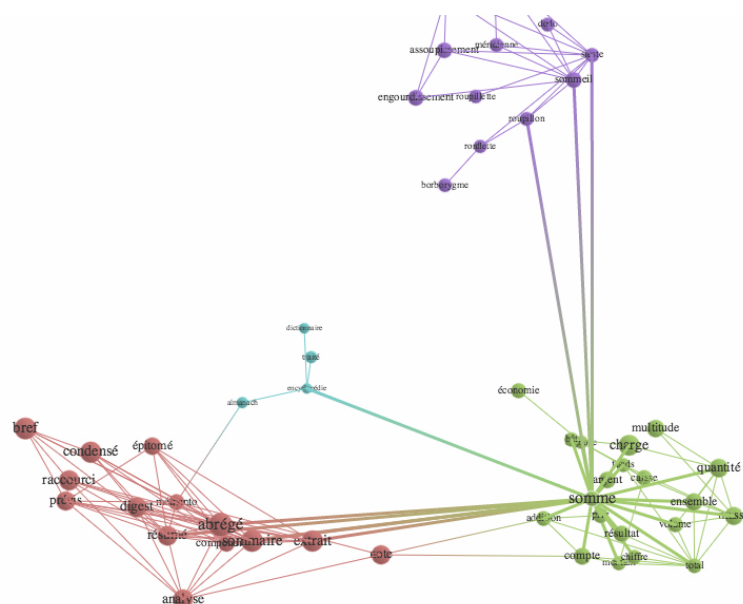


Fig. 8. Analyse proxémique du sous-graphe de synonymie de *somme* dans le DicoSyn

L'analyse topologique des graphes de vocables permet donc de reconstituer une partie de l'information sur la polysémie, implicite dans ces graphes. On ne peut cependant accepter une telle modélisation, d'un point de vue lexicologique, que si l'on se fonde sur une théorie de la polysémie en tant que phénomène continu et non discret [61] et si l'on se cantonne à une description très partielle du lexique, qui ne se donne pas pour but d'identifier toutes les propriétés sémantiques, formelles, relationnelles et de combinatoire des lexies. Le fait que *somme* 'addition' est un nom féminin et *somme* 'fait de dormir' est un nom masculin, le fait que le premier est lié à *soustraction* et le second à *sieste*, le fait que le premier est combinable avec *calculer* (*calculer une somme*) et le second avec *piquer* (*piquer une somme*), etc., tous ces faits démontrent que les propriétés linguistiques fondamentales sont bien associées aux sens individuels, sens qu'il faut donc réifier au sein du modèle lexical, comme il est fait dans tout dictionnaire voulant prétendre à une quelconque utilité.

Un graphe de vocables est très difficilement exploitable dans un contexte d'enseignement des langues, là où il faut justement développer chez l'apprenant la connaissance des lexies par l'assimilation de leurs caractéristiques individuelles qui permettent leur emploi correct dans la phrase et permettent leur positionnement adéquat dans le lexique mental. On peut aussi penser que, même en traitement automatique des langues, de bien meilleurs résultats pourront être obtenus si l'on dispose de graphes non ambigus, où chaque sommet est une entité complexe encodant des caractéristiques propres à un sens donné [63]. Finalement, le graphe de vocables ne sera pas un bon modèle du lexique mental logique, car les informations sur les sens individuels du type de celles mentionnées ci-dessus à propos des différentes lexies SOMME font partie de la connaissance lexicale.

Les graphes de lexies donnent une représentation beaucoup plus exacte de la sémantique qui structure le lexique. Ils sont cependant impossibles à construire de façon entièrement automatique dans l'état actuel des connaissances en traitement automatique des langues, ce qui explique la prédominance des graphes de vocables parmi les ressources lexicales informatisées actuellement disponibles. La construction des graphes de lexies force aussi, comme dans le cas des dictionnaires, à se confronter à la polysémie, qui peut être contour-

née, ou même niée en tant que phénomène discret, si l'on se limite au recours à des graphes de vocables. Il s'agit cependant là de problèmes pratiques, qu'il convient de résoudre plutôt que de les contourner au détriment de l'adéquation théorique.

Examinons maintenant brièvement les principales différences trouvées dans les réseaux lexicaux du point de vue de leurs arcs. Ces derniers peuvent être : (i) orientés ou non ; (ii) étiquetés ou non. Les réseaux lexicaux dont les arcs sont étiquetés se distinguent entre eux par la richesse du jeu d'étiquettes, c'est-à-dire par le nombre et la nature des liens lexicaux pris en compte. Voici deux exemples de réseaux se distinguant par la nature de leurs arcs.

1. Les arcs de la Figure 8 ci-dessus dénotent tous la relation de (quasi-)synonymie qui, de plus, est conçue dans ce cas-ci comme une relation parfaitement symétrique ($L_1 \text{ syn } L_2 \equiv L_2 \text{ syn } L_1$). En conséquence, puisque la relation est unique et systématiquement bidirectionnelle, les arcs ne sont formellement ni orientés ni étiquetés.
2. La Figure 7 (voir plus haut, section 3.3), par contre, est la visualisation d'un extrait du Réseau Lexical du Français, dont la structure est celle d'un Système Lexical (section 4.2, plus bas). En conséquence, même si cette information n'apparaît pas dans la visualisation, chaque arc est étiqueté par une relation lexicale orientée. La Figure 9 ci-dessous visualise la sémantique de l'arc connectant les sommets *vocabulaire* et *riche*_{Adj} II.2. Il s'agit du lien de fonction lexicale **Magn**₂^{quant}, qui relie dans le lexique français la lexie VOCABULAIRE à un de ses collocatifs intensificateurs [64]¹⁹.

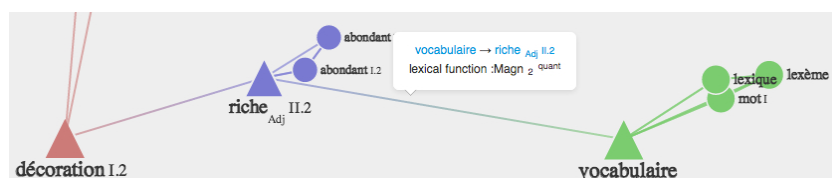


Fig. 9. Visualisation du lien de fonction lexicale d'intensification dans le Réseau Lexical du Français

L'hétérogénéité des arcs est évidente, puisque le lien qui vient d'être visualisé est clairement distinct du lien unissant, par exemple, le sommet *riche*_{Adj} II.2 à *abondant* I.2. Il s'agit ici de quasi-synonymie à intersection – fonction lexicale **Syn**_∩. Dans le cadre de la lexicologie des Systèmes Lexicaux, la synonymie est une relation orientée (non systématiquement symétrique). En conséquence, les deux sommets en question sont en fait connectés par deux arcs de quasi-synonymie réciproques, comme l'illustre la Figure 10.

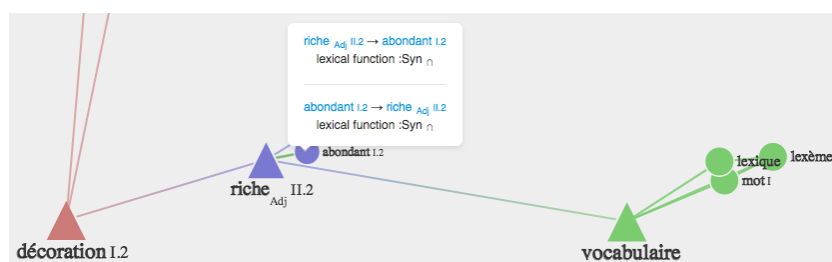


Fig. 10. Visualisation de deux arcs de quasi-synonymie réciproques

Pour résumer, les Systèmes Lexicaux peuvent être très généralement caractérisés, parmi les différents types de réseaux lexicaux, comme étant (i) des graphes de lexies, (ii) dont les arcs sont orientés et étiquetés. Nous allons maintenant nous concentrer sur la description de ce type de structure et expliquer pourquoi il nous semble intéressant de l'exploiter comme modèle géométrisable du lexique mental logique.

4.2 Émergence des Systèmes Lexicaux (SL)

Plusieurs publications sont déjà consacrées à la description de la structure des Systèmes Lexicaux – dorénavant, SL – ainsi qu'aux méthodes lexicographiques élaborées pour ce type de modèles [13, 40, 65, 66]. Nous présentons donc ici les SL selon une autre perspective, en contextualisant la construction de la notion dans le cadre des recherches en Lexicologie Explicative et Combinatoire. Nous espérons que cet éclairage « historique » permettra de mieux faire comprendre la logique de notre démarche et, par là même, d'apporter des éléments de justification de l'approche adoptée pour géométriser les modèles lexicaux.

4.2.1 Origines : expérimentation avec la ressource lexicale DiCo

Le projet de fonder la recherche en Lexicologie Explicative et Combinatoire sur une structure de réseau lexical a graduellement pris corps dans le contexte du travail effectué à l'Université de Montréal sur la base lexicale DiCo [67, 68].

La finalité première du DiCo était d'aller au-delà de la structure textuelle des *Dictionnaires Explicatifs et Combinatoires* ou DEC [41], pour construire une ressource sur le français qui soit véritablement exploitable aussi bien par l'humain que par la machine. Nous nous étions en cela inspiré de la philosophie du projet WordNet (section 3.2, ci-dessus), qui était pour nous le modèle à suivre en matière de développement de ressources lexicales multifonctionnelles. Bien entendu, l'approche lexicographique proprement dite du DiCo était très différente de celle de WordNet. Il s'agissait de la même lexicographie que celle appliquée dans la rédaction des DEC, mais avec des contraintes de formalisation qui ont notamment permis d'exploiter les données de la base lexicale sous deux autres formes :

1. l'ouvrage de support à l'apprentissage du vocabulaire *Lexique Actif du Français* [69] – par reformatage manuel des données exportées du DiCo ;
2. la base de données DiCoSQL, consultable en ligne par recherche « formulaire » sur le site de l'Observatoire de linguistique Sens-Texte (OLST) au moyen de l'interface DiCouèbe²⁰ ou, dans sa version grand public DiCoPop²¹, par compilation automatique des données du DiCo [70,71].

Le travail sur le DiCo a permis de mettre en évidence deux faits fondamentaux. Tout d'abord, il est apparu que la Lexicographie Explicative et Combinatoire, de nature éminemment relationnelle, imposait de travailler sur des structures non textuelles si l'on voulait pouvoir dépasser le stage de l'échantillon lexicographique de quelques centaines de vocables. Les contraintes formelles imposées par le tissage des liens de fonctions lexicales de lexies à lexies, clairement identifiées, présupposent que les lexicographes puissent littéralement naviguer dans le réseau de la langue [72]. Ensuite, nous avons pris conscience du fait que le système des fonctions lexicales standards de la théorie Sens-Texte [14 : 125–152, 64, 73]²² offrait véritablement – au-delà d'un outil de modélisation des liens lexicaux, de la lexicalisation et des relations paraphrase – une théorie de la structure lexicale des langues.

Rappelons que le système des fonctions lexicales standards Sens-Texte est fondé sur un ensemble relativement petit de fonctions lexicales standards dites *simples* (un peu plus d'une soixantaine) qui peuvent être combinées et qui correspondent aux relations lexicales

universellement présentes dans le lexique des langues. On distingue deux familles de fonctions lexicales, correspondant aux deux grands types de liens lexicaux :

1. Les fonctions lexicales **paradigmatiques** connectent les lexies à leurs **dérivés sémantiques** : synonymes, antonymes, translatifs (verbaux, nominaux...), noms d'actants et de circonstants, etc.
2. Les fonctions lexicales **syntagmatiques** – par exemple, la fonction lexicale **Magn**, dont une variante (**Magn₂^{temp}**) a été introduite à la section 4.1 – connectent les lexies à leurs **collocatifs** potentiels : intensificateurs, verbes supports, verbes de réalisation, etc.

Puisque le système des fonctions lexicales (i) synthétise ce qui est universel dans les connexions entre lexies – que ce soit du point de vue de leurs relations sémantiques que de celui de leurs relations de combinatoire –, (ii) est au cœur des règles de paraphrasage et de lexicalisation, et (iii) a démontré son caractère opératoire de profilage sémantique et combinatoire des lexies dans le cadre de projets lexicographiques, nous avons fait l'hypothèse qu'il représentait le principe général de structuration des lexiques. En d'autres termes, nous avons considéré que l'ossature des réseaux lexicaux des langues était déterminée **en tout premier lieu** par les liens de fonctions lexicales standards et que, en tissant méthodiquement ceux-ci en tant qu'arcs de graphes connectant des sommets lexicaux, on parviendrait à construire des modèles relationnels plausibles des lexiques.

Nous avons développé cette hypothèse en proposant une première caractérisation formelle de tels graphes, que nous avons nommés *Systèmes Lexicaux*, du fait de la nature systémique de leur organisation. Une version « preuve de concept » d'un SL ayant été générée par compilation des données du DiCoSQL [13]²³, il ne restait plus qu'à passer à l'étape suivante et à construire des SL grandeur nature selon une méthodologie lexicographique.

4.2.2 Vers le graphe petit-monde du Réseau Lexical du Français (RL-fr)

Les premières recherches linguistiques sur l'exploitation de la structure de graphe des réseaux lexicaux (non taxinomiques) ont paru alors que s'effectuait le travail sur le DiCo. Il nous semble, par ailleurs, que les chercheurs français ont joué ici un rôle de pionniers [60, 61]. Les travaux de B. Gaume, tout particulièrement, sur l'analyse des petits-mondes lexicaux et les principes d'analyse par proxémie [62] ont été pour nous une révélation. En effet, ces travaux nous démontraient, tout d'abord, que le type de modélisation que nous envisagions pour le lexique avait été théorisé d'un point de vue mathématique. Ensuite, ils suggéraient que les propriétés formelles des SL les rendraient calculables et géométrisables, au sens où elles permettraient de faire émerger et de visualiser de l'information nouvelle à propos de la connaissance lexicale à partir de l'analyse topologique des SL. La consultation des travaux sur les graphes petits-mondes démontrait à quel point les SL que nous postulions étaient en harmonie formelle avec nombre de systèmes émergeant naturellement de phénomènes relationnels – connexions entre individus dans les réseaux sociaux, interactions biologiques, etc. –, que B. Gaume désigne sous le terme de **grands graphes de terrain** [62]. Pour les lexicologues et lexicographes, il semble normal de considérer que le lexique est un grand graphe de terrain et il est particulièrement satisfaisant de disposer d'un tel terme pour désigner l'entité à laquelle ils se confrontent.

Lorsqu'a débuté le projet RELIEF au laboratoire ATILF [65], projet qui s'est déroulé sur quatre ans (2010–2014) et avait pour finalité de construire une première version d'un Réseau Lexical du Français multifonctionnel, la notion de SL était élaborée dans ses grandes lignes. Il restait cependant à effectuer les avancées suivantes :

- mettre au point un éditeur lexicographique dédié à la lexicographie des SL – éditeur qui sera conçu en collaboration avec le partenaire industriel du projet RELIEF (MVS²⁴) et programmé par ce dernier [75] ;
- inventer une lexicographie des SL fondée sur la navigation dans le graphe lexical et la génération automatique d’articles lexicographiques (= textes lexicographiques) à partir des données du graphe lexical [40] – les articles lexicographiques correspondent approximativement au contenu des sommets du graphe lexical, qui ne sont pas des entités atomiques et qui possèdent au contraire une structure informationnelle extrêmement riche ;
- exploiter et perfectionner la hiérarchie d’étiquettes sémantiques du français [76], un outil descriptif développé dans le cadre du travail sur le DiCo et le LAF qui permet de projeter une taxinomie sémantique sur les SL et, donc, d’y modéliser les inférences fondées sur l’héritage de propriétés ;
- vérifier qu’une telle approche est véritablement implémentable, fonctionnelle et conduit à la construction de SL valides, possédant notamment une structure de graphe petit-monde qui permettra leur exploitation informatique.

Ces différents buts ont été atteints avec la construction d’un premier Réseau Lexical du Français, RL-fr, de taille significative²⁵. De plus, des expérimentations sur d’autres langues que le français ont été effectuées : échantillons de RL- développés, dans le cadre de recherches doctorales, pour l’espagnol (RL-es) et le coréen (RL-kr) ; compilation des données du WordNet de Princeton sous forme d’un SL (RL-en), dont la comparaison avec le SL du RL-fr est présentée dans [77] ; amorçage de la construction d’un modèle du russe (RL-ru)²⁶. Le travail sur ces SL se poursuit sur le long terme, à la fois pour développer ces SL en tant que ressources lexicales et pour les utiliser comme supports pour la recherche en lexicologie et autres disciplines connexes.

4.2.3 Caractéristiques formelles du SL du RL-fr

Le travail de construction du RL-fr a permis de faire évoluer la notion même de SL. Si la plupart des postulats initiaux sur la structure de ces réseaux lexicaux ont été repris, certaines caractéristiques ont été précisées ou reconsidérées.

Signalons tout d’abord que le RL-fr est plus qu’un SL, puisqu’il incorpore aussi des connaissances métalinguistiques, qui servent à spécifier et structurer le modèle lexical proprement dit. Ainsi, on trouve encapsulés dans le RL-fr des modèles pour : les traits grammaticaux attribuables aux lexies, la morphologie des lexies, la hiérarchie taxinomique d’étiquettes sémantiques, le système des fonctions lexicales, etc. Ces modèles, que nous ne pouvons détailler ici, sont extérieurs au SL proprement dit, sur lequel nous focalisons notre attention.

La principale évolution formelle qu’ont connue les SL dans le cadre du projet lexicographique du RL-fr a été l’introduction de liens lexicaux extérieurs au système des fonctions lexicales. Même si ce dernier reste l’élément structurant fondamental des SL, trois autres types de liens sont graduellement venus le concurrencer, liens que nous présentons ci-dessous, des plus sémantiques aux plus formels.

1) Liens d’inclusion définitionnelle. Nous venons d’entamer le travail sur l’intégration dans le RL-fr de définitions lexicales formalisées selon les principes de la Lexicologie Explicative et Combinatoire. L’élaboration de telles définitions conduit au tissage de liens d’inclusion définitionnelle dans les SL [78]. Par exemple, la lexie FRAISE [*manger un bol de fraises du jardin*] doit pointer par un lien d’inclusion définitionnelle vers la lexie adjectivale ROUGE, puisque sa définition doit spécifier que les fraises sont typiquement de couleur

rouge. Grâce au tissage d'un arc correspondant dans le RL-fr, il est possible de parcourir le graphe lexical pour atteindre le sommet FRAISE à partir du sommet ROUGE, en remontant le lien d'inclusion définitionnelle. Il s'agit d'un parcours sémantique disponible dans le lexique du français – via un lien sémantique fort existant entre les deux lexies en cause²⁷ – que ne modélise pas le système des fonctions lexicales standards.

2) Liens de dérivation polysémique. Le RL-fr contient la modélisation explicite de la structure polysémique des vocables grâce au tissage de liens de dérivation entre copolysèmes (= lexies d'un même vocable). Certains de ces liens – comme le lien d'extension sémantique qui unit [la] LUNE I.1 '[le] satellite de la Terre' à son copolysème dérivé LUNE I.2 [de X] 'satellite de la planète X' – sont de nature sémantique et relèvent d'une très forte intersection de sens. D'autres – comme le lien de dérivation métaphorique qui unit la même lexie LUNE I.1 à son autre copolysème dérivé LUNE II [de X] 'fesses de l'individu X' – vont connecter des lexies qui appartiennent à des espaces sémantiques tout à fait disjoints. Ils permettent donc d'effectuer des bonds sémantiques à l'intérieur du graphe du RL-fr. Ce dernier est, à notre connaissance, la seule ressource lexicale d'envergure qui incorpore une modélisation relationnelle systématique de la structure polysémique des vocables²⁸.

3) Liens d'inclusion formelle. Le RL-fr contient la description systématique de la structure lexico-syntaxique des locutions [79], ce qui implique le tissage de liens d'inclusion formelle entre les sommets du graphe correspondant à des locutions – par exemple, 'LUNE DE MIEL'²⁹ – et ceux correspondant aux lexèmes dont ces locutions sont constituées. Il s'agit ici de liens qui sont souvent purement formels. (Une lune de miel n'a rien à voir avec la Lune et le miel.) Ils connectent donc, comme les liens de copolysémie métaphoriques mentionnés ci-dessus, des espaces sémantiques disjoints dans le SL. De tels liens renforcent le phénomène de petit-monde dont il a été question à la section 3.3.

Notons que les liens d'inclusion formelle phraséologique qui viennent d'être présentés participent, avec les liens de fonctions lexicales syntagmatiques, à la formalisation de la phraséologie lexicale, qui occupe une place essentielle dans l'approche de la Lexicologie Explicative et Combinatoire. Comme il a été montré dans [80], cette modélisation de la phraséologie équivaut à l'enchâssement d'une partie significative de la syntaxe des langues dans les SL et participe, de façon plus générale, à la prise en compte de l'interaction entre lexicologie et grammaire dans ce type de modèle²⁹.

Une dernière caractéristique formelle des SL qu'il convient de souligner est le fait qu'il s'agit de modèles relativistes : toute information sur une lexie présente dans un SL porte un degré de confiance qui encode une estimation de sa validité. Par défaut, l'information entrée manuellement par les lexicographes porte un degré de confiance maximal. Ceux-ci peuvent cependant ajuster le degré en question, pour indiquer qu'il s'agit d'une information qui doit être relativisée, pour diverses raisons : des désaccords existent entre lexicographes (ou avec le contenu des corpus consultés) ; on est en présence d'une information instable, impliquée dans un processus d'évolution diachronique ; etc. Si l'information n'est pas introduite manuellement, mais par semi-automatisation de la tâche lexicographique fondée, notamment, sur une implantation du raisonnement analogique [81], un degré de confiance relativement faible indiquera le caractère hypothétique de l'information en question. Cette nature relativiste des SL en fait des structures particulièrement propices à la modélisation d'inférences effectuelles sur le lexique mental – cf. (2), section 2.1 –, dont on peut faire l'hypothèse qu'elles reposent sur des informations potentiellement instables et dont la validité est de nature gradable. Le lexique mental est un système complexe dynamique, qui ne peut être modélisé que par une structure qui permet l'encodage de cette dynamité. Nous sommes ainsi convaincu que la nature relativiste des SL en fait des modèles adaptés à un domaine d'application privilégié de la recherche lexicologique : l'enseignement et apprentissage de la langue³⁰.

5 Conclusion

Nous concluons en revenant au problème initialement posé : celui de la géométrie du lexique. Rappelons pour cela les principaux points qui ont été abordés.

Après avoir expliqué l'importance du choix d'une bonne représentation formelle de la structure du lexique, nous avons été amené à distinguer deux notions de lexique mental : (i) le lexique mental logique – objet d'étude de la lexicologie – et (ii) le lexique mental physiologique – objet d'étude, notamment, de la neurolinguistique. Nous avons postulé que ces deux lexiques possédaient des structures distinctes, bien que formellement compatibles³¹.

Nous avons ensuite tenté de démontrer que les meilleures représentations formelles de des lexiques logiques étaient celles qui explicitaient la structure relationnelle de ces derniers, de façon non taxinomique. Nous avons désigné de tels modèles par le terme de *réseau lexical*. Les réseaux lexicaux possèdent une topologie particulière, dite *petit-monde*, dont nous pensons qu'elle rend adéquatement compte de l'organisation en « réseau social » des lexiques. En plus de permettre le calcul sémantique, la topologie des réseaux lexicaux en fait des modèles directement géométrisables et visualisables, qui peuvent donc servir de supports au raisonnement diagrammatique au sens peircien (cf. section 3.1).

Finalement, nous avons présenté les Systèmes Lexicaux (SL) : un type particulier de réseau lexical géométrisable, à la base duquel est notamment construit de façon lexicographique le Réseau Lexical du Français.

Nous espérons que cet exposé aura su tout d'abord convaincre de l'intérêt de disposer de modèles relationnels géométrisables des lexiques. Notre but était également de présenter les SL comme des modèles adéquats non seulement en lexicologie, mais aussi pour établir la connexion avec les recherches menées sur le lexique mental physiologique. Il nous semble notamment qu'une compatibilité naturelle existe entre l'approche des SL et les recherches les plus récentes en neuroscience sur la localisation des connaissances sémantiques lexicales dans le cerveau (*map of the semantic system*) présentées dans [83]³². Remarquons que la méthodologie employée dans [83] s'appuie sur l'exploitation des profils statistiques des mots en corpus (*word embedding spaces*) pour interpréter les mesures de l'activité neuronale en situation de décodage linguistique. De ce fait, on peut considérer que ces travaux établissent une connexion entre les observables conditionnant traditionnellement le modèle du lexique logique (réalisations linguistiques) et ceux conditionnant le modèle du lexique physiologique (comportements physiologiques en situation d'activité langagière) : cf. la double flèche horizontale en haut de la Figure 2 (section 2.2)³³. Il n'est pas exclu que les SL eux-mêmes puissent être exploités dans ce type d'expérimentation, comme modèles de référence pour le lexique logique : cf. la double flèche horizontale de la Figure 2 entre les modèles des lexiques logique et physiologique.

Notons finalement que nous avons fait l'impasse sur la difficile question de la cohabitation des SL. Les lexiques cohabitent naturellement : cohabitation interlangue de lexiques mentaux, dans les cas de multilinguisme, ou cohabitation intralangue au sein d'un même lexique mental, si l'on considère les sous-systèmes lexicaux que constituent, par exemple, les terminologies [84]. Ces interactions restent à étudier formellement en lexicologie des réseaux lexicaux.

Nous sommes infiniment reconnaissant à Bruno Gaume pour ses commentaires sur une première version de cet article et pour les discussions qui ont suivi. Ces échanges nous ont permis de rendre notre exposé plus rigoureux. Bien entendu, nous assumons l'entière responsabilité de toutes les erreurs et inexactitudes qui subsistent. Le travail de construction des Systèmes Lexicaux serait impossible sans le génial éditeur lexicographique programmé par le non moins génial Nabil Gader (MVS). Nous remercions Igor Mel'čuk pour... tout. Merci aussi aux étudiants et lexicographes avec lesquels nous

avons travaillé dans le cadre des projets DiCo, à l'Université de Montréal, et RELIEF, au laboratoire ATILF CRNS de Nancy. Les visualisations de graphes lexicaux des Figures 5 à 10 ont été générées grâce aux outils de Kodex.Lab³⁴.

Références

1. M. A. K. Halliday, C. M. I. M. Matthiessen, *An introduction to functional grammar* (Hodder Arnold, London, 2004)
2. R. R. Hoffman, « Metaphor in Science », dans R. P. Honeck & R. R. Hoffman (dir.) *Cognition and Figurative Language*, 393–423 (Lawrence Erlbaum, Hillsdale NJ, 1980)
3. G. Lochak, *La géométrisation de la physique* (Flammarion, Paris, 1994)
4. A. Polguère, « Collocations et fonctions lexicales : pour un modèle d'apprentissage », *Revue Française de Linguistique Appliquée*, **E(1)**, 117–133. (2003)
5. *Le Petit Robert de la langue française* (Dictionnaires Le Robert, Paris, 2014)
6. J. R. Ross, *Constraints on Variables in Syntax* (Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge MA, 1967)
7. L. Tesnière, *Éléments de syntaxe structurale* (Klincksieck, Paris, 1959)
8. Y. Lecerf, « Programme des conflits, modèle des conflits », *Traduction automatique*, **1(4) / 1(5)**, 11–18 / 17–36 (1960)
9. P. Ihm, Y. Lecerf, *Éléments pour une grammaire générale des langues projectives*. (Rapport technique, EUR 210.f, Communauté Européenne de l'Énergie Atomique – Euratom, Bruxelles, 1963)
10. S. Marcus, « Sur la notion de projectivité », *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **11**, 181–192 (1965)
11. D. Béchet, O. Lacroix, « CDGFr, un corpus en dépendances non-projectives pour le français », *Actes de la 22^e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2015)*, 522–528 (Association pour le Traitement Automatique des Langues, Caen, 2015)
12. R. Solé, B. Corominas-Murtra, S. Valverde, L. Steels, « Language Networks : Their Structure, Function, and Evolution », *Complexity*, **15(6)**, 20–26 (2010)
13. A. Polguère, « Lexical systems : graph models of natural language lexicons », *Language Resources and Evaluation*, **43(1)**, 41–55 (2009)
14. I. Mel'čuk, A. Clas, A. Polguère, *Introduction à la lexicologie explicative et combinatoire* (Duculot, Paris/Louvain-la-Neuve, 1995)
15. I. Mel'čuk, *Language: From Meaning to Text* (LRC Publishing House & Academic Studies Press, Moscow & Boston, 2016)
16. J. Queiroz, F. Merrell, « Semiosis and pragmatism : Toward a dynamic concept of meaning », *Sign Systems Studies*, **34(1)**, 37–65 (2006)
17. A. Wierzbicka, « Mental Language and Semantic Primitives », *Communication and Cognition*, **X(3–4)**, 155–179 (1977)
18. M. Zock, « Words in Books, Computers and the Human Mind », *Journal of Cognitive Science*, **16(4)**, 355–378 (2015)
19. A. Polguère, « Lexical Contextualism : The Abélard Syndrome », dans N. Gala, R. Rapp & G. Bel-Enguix (dir.) *Language Production, Cognition, and the Lexicon*, 53–73 (Springer, Cham Heidelberg New York Dordrecht Londres, 2015)
20. J. Aitchison, *Words in the Mind : An Introduction to the Mental Lexicon* (4^e édition, Wiley-Blackwell, Oxford, 2012)
21. P. Marquer, *L'organisation du lexique mental. Des « contraires » aux expressions idiomatiques*. (L'Harmattan, Paris, 2005)

22. A. Polguère, « Terminologie grammaticale... et lexicale. » *Repères*, **49**, 115–130 (2014)
23. D. Sperber, D. Wilson, « The mapping between the mental and the public lexicon » dans P. Carruthers & J. Boucher (dir.) *Language and thought. Interdisciplinary themes*, 184–200 (Cambridge University Press, Cambridge, 1998 – Première parution dans *UCL Working Papers in Linguistics*, **9**, University College London, Londres, 1997)
24. R. C. Oldfield, « Things, words and the brain », *Quarterly Journal of Experimental Psychology*, **18(4)**, 340–353 (1966)
25. W. Szubko-Sitarek, *Multilingual Lexical Recognition in the Mental Lexicon of Third Language Users* (Second Language Learning and Teaching Series, Springer-Verlag, Berlin Heidelberg, 2015)
26. A. M. Treisman, « Contextual cues in selective listening », *Quarterly Journal of Experimental Psychology*, **12(4)**, 242–248 (1960)
27. G. A. Miller, « Semantic Relations among Words », dans M. Halle, J. Bresnan & G. A. Miller (dir.) *Linguistic Theory and Psychological Reality*, 60–118 (The MIT Press, Cambridge MA & London, 1981)
28. A. Lenci, « Distributional semantics in linguistic and cognitive research. », *Rivista di Linguistica*, **20(1)**, 1–31 (2008)
29. *Antidote 9* (Druide informatique inc., Montréal, 2016)
30. S. Charest, É. Brunelle, J. Fontaine, « Au-delà de la paire de mots : extraction de cooccurrences syntaxiques multilexémiques », *Actes de la 17^e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2010)* (Association pour le Traitement Automatique des Langues, Montréal, 2010)
31. M. Baroni, S. Bernardini, A. Ferraresi, E. Zanchetta, « The WaCky wide web : a collection of very large linguistically processed web-crawled corpora », *Language Resources and Evaluation*, **43(3)**, 209–226 (2009)
32. T. Blakely, K. J. Miller, R. P.N. Rao, M. D. Holmes, J. G. Ojemann, J. G. « Localization and classification of phonemes using high spatial resolution electrocorticography (ECoG) grids », *Proceedings of the 30th Annual International IEEE EMBS Conference*, 4964–4967 (Vancouver, 2008)
33. C. Marzi, V. Pirrelli, « A Neuro-Computational Approach to Understanding the Mental Lexicon », *Journal of Cognitive Science*, **16(4)**, 493–534 (2015)
34. G. T. Altmann, *The Ascent of Babel : An Exploration of Language, Mind, and Understanding* (Oxford University Press, Oxford, 1997)
35. S. Dehaene, *Les neurones de la lecture* (Odile Jacob, Paris, 2007)
36. J. Segui, « Évolution du concept de lexique mental », *Revue de neuropsychologie*, **7(1)**, 21–26 (2015)
37. C. Müller-Spitzer, « Textual structures in electronic dictionaries compared with printed dictionaries : A short general survey », dans R. H. Gouws, U. Heid, W. Schweickard & H. E. Wiegand (dir.) *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume : Recent Developments with Focus on Electronic and Computational Lexicography*, 367–381 (De Gruyter Mouton, Berlin & Boston, 2013)
38. G. A. Miller, « Dictionaries in the mind », *Language and Cognitive Processes*, **1(3)**, 171–185 (1986)
39. B. T. S. Atkins « Bilingual Dictionaries: Past, Present and Future », *Euralex'96 Proceedings*, 515–590 (Gothenburg University, Gothenburg, 1996)
40. A. Polguère « Lexicographie des dictionnaires virtuels », dans Y. Apresjan, I. Boguslavsky, M.-C. L'Homme, L. Iomdin, J. Miličević, A. Polguère & L. Wanner (dir.) *Meanings, Texts, and Other Exciting Things. A Festschrift to Commemorate the 80th Anniversary of Professor Igor Alexandrovič Mel'čuk*, 509–523 (Studia Philologica, Jazyki slavjanskoj kultury Publishers, Moscou, 2012)

41. I. Mel'čuk et coll., *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques. Volumes I-IV* (Les Presses de l'Université de Montréal, Montréal, 1984, 1988, 1992, 1999)
42. P. Farias, J. Queiroz, « On diagrams for Peirce's 10, 28, and 66 classes of signs », *Semiotica*, **147**(1/4), 165–184 (2003)
43. A. Shimojima, « The Graphic-Linguistic Distinction », *Artificial Intelligence Review*, **15**, 5–27 (2001)
44. R. H. Gouws, D. J. Prinsloo, « Cross-Referencing as a Lexicographic Device », *Lexikos*, **8**, 17–36 (1998)
45. G. A. Miller, R. Beckwith, C. D. Fellbaum, D. Gross, K. J. Miller, « Introduction to WordNet : An On-line Lexical Database », *International Journal of Lexicography*, **4**(3), 235–244 (1990)
46. G. A. Miller, « WordNet: A Lexical Database for English », *Communications of the ACM*, **38**(11), 39–41 (1995)
47. C. D. Fellbaum, G. A. Miller, « Morphosemantic Links in WordNet », *Traitement Automatique des Langues (T.a.l.)*, **44**(2), 69–80 (2003)
48. C. D. Fellbaum, A. Osherson, P. E. Clark, « Putting Semantics into WordNet's "Morphosemantic" Links », dans Z. Vetulani & H. Uszkoreit (dir.) *Responding to Information Society Challenges : New Advances in Human Language Technologies*, 350–358 (Springer Lecture Notes in Informatics 5603, Springer, Dordrecht, 2009)
49. A. Polguère, « Les petits soucis ne poussent plus dans le champ lexical des sentiments », dans F. Baider & G. Cislaru (dir.) *Cartographie des émotions. Propositions linguistiques et sociolinguistiques*, 21–41 (Presses Sorbonne Nouvelle, Paris, 2013)
50. V. Barbu Mititelu, C. Forăscu, C. D. Fellbaum, P. Vossen, « Tuning Hierarchies in Princeton WordNet », dans A. Lohk, C. D. Fellbaum & L. Vöhandu (dir.) *Proceedings of the Eighth Global WordNet Conference*, 177–183 (Bucarest, 2016)
51. I. Mel'čuk, « Paraphrase et lexique : la théorie Sens-Texte et le Dictionnaire Explicatif et Combinatoire », dans I. Mel'čuk, N. Arbatchewsky-Jumarie, L. Iordanskaja & S. Mantha (dir.) *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques III*, 9–58 (Les Presses de l'Université de Montréal, Montréal, 1992)
52. J. Milićević, *La paraphrase. Modélisation de la paraphrase langagière* (Sciences pour la communication, Peter Lang, Bern, 2007)
53. S. Milgram, « The small world problem », *Psychology Today*, **1**(1), 61–67 (1967)
54. D. J. Watts, S. H. Strogatz, « Collective dynamics of 'small-world' networks », *Nature*, **393**, 440–442 (1998)
55. A. Wierzbicka, « The Theory of the Mental Lexicon », dans S. Kempgen, P. Kosta, T. Berger & K. Gutschmidt (dir.) *Die slavischen Sprachen/The Slavic Languages : An International Handbook of their Structure, their History and their Investigation*, 848–863 (Mouton de Gruyter, Berlin & New York, 2009)
56. B. Gaume, K. Duvignau, « Pour une ergonomie cognitive des dictionnaires électroniques », *Document numérique*, **8**(3), 157–181 (2004)
57. B. Gaume, K. Duvignau, E. Navarro, Y. Desalle, H. Cheung, S.-K. Hsieh, P. Magistry, L. Prévot, « Skillex : a graph-based lexical score for measuring the semantic efficiency of used verbs by human subjects describing actions », *Traitement Automatique des Langues (T.a.l.)*, **55**(3) (sous presse)
58. A. E. Motter, A. P.S. de Moura, Y.-C. Lai, P. Dasgupta, « Topology of the conceptual network of language », *Physical Review E*, **65**, 065102-1–4 (2002)
59. M. E. Bales, S. B. Johnson, « Graph theoretic modeling of large-scale semantic », *Journal of Biomedical Informatics*, **39**, 451–464 (2006)

60. S. Ploux, « Modélisation et traitement informatique de la synonymie », *Linguisticae Investigationes*, **21(1)**, 1–27 (1997)
61. S. Ploux, B. Victorri, « Construction d’espaces sémantiques à l’aide de dictionnaires de synonymes », *Traitement Automatique des Langues (T.a.l.)*, **39(1)**, 161–182 (1998)
62. B. Gaume, « Balades aléatoires dans les Petits Mondes Lexicaux », *I3 Information Interaction Intelligence*, **4(2)**, 39–96 (2004)
63. P. Norvig, G. Lakoff, « Taking: A Study of Lexical Network Theory », *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society (BLS 13)*, 195–206 (Berkeley Linguistics Society, Berkeley, 1987)
64. I. Mel’čuk, « Lexical Functions : A Tool for the Description of Lexical Relations in the Lexicon », dans L. Wanner (dir.) *Lexical Functions in Lexicography and Natural Language Processing*, 37–102 (Language Companion Series 31, John Benjamins, Amsterdam/Philadelphia, 1996)
65. V. Lux-Pogodalla, A. Polguère, « Construction of a French Lexical Network : Methodological Issues », *Proceedings of the First International Workshop on Lexical Resources, WoLeR 2011. An ESSLLI 2011 Workshop*, 54–61 (Ljubljana, 2011)
66. A. Polguère, « From Writing Dictionaries to Weaving Lexical Networks », *International Journal of Lexicography*, **27(4)**, 396–418 (2014)
67. A. Polguère, « Une base de données lexicales du français et ses applications possibles en didactique », *Revue de linguistique et de didactique des langues (Lidil)*, **21**, 75–97 (2000)
68. I. Mel’čuk, A. Polguère, « Dérivations sémantiques et collocations dans le DiCo/LAF », *Langue française*, **150**, 66–83 (2006)
69. I. Mel’čuk, A. Polguère, *Lexique actif du français. L’apprentissage du vocabulaire fondé sur 20 000 dérivations sémantiques et collocations du français* (Champs linguistiques, De Boeck & Larcier, Bruxelles, 2007)
70. J. Steinlin, S. Kahane, A. Polguère, « Compiling a “classical” explanatory combinatorial lexicographic description into a relational database », *Proceedings of the Second International Conference on the Meaning Text Theory (MTT’2005)*, 477–485 (Moscou, 2005)
71. A.-L. Jousse, A. Polguère, « Le DiCo et sa version DiCouèbe. Document descriptif et manuel d’utilisation » (OLST, Université de Montréal, 2005)
72. A. Polguère, « Like a Lexicographer Weaving Her Lexical Network », *Proceedings of the Third Workshop on Cognitive Aspects of the Lexicon (CogALex III)*, 1–3 (The COLING 2012 Organizing Committee, Mumbai, 2012)
73. A. Polguère, « Lexical function standardness », dans L. Wanner (dir.) *Selected Lexical and Grammatical Issues in the Meaning-Text Theory. In Honour of Mel’čuk*, 43–95 (Language Companion Series 84, John Benjamins, Amsterdam/Philadelphia, 2007)
74. A. Polguère, « Structural Properties of Lexical Systems : Monolingual and Multilingual Perspectives », *Proceedings of the Workshop on Multilingual Language Resources and Interoperability (COLING/ACL 2006)*, 50–59 (Association for Computational Linguistics, Sydney, 2006)
75. N. Gader, V. Lux-Pogodalla, A. Polguère, « Hand-Crafting a Lexical Network With a Knowledge-Based Graph Editor », *Proceedings of the Third Workshop on Cognitive Aspects of the Lexicon (CogALex III)*, 109–125 (The COLING 2012 Organizing Committee, Mumbai, 2012)
76. A. Polguère, « Classification sémantique des lexies fondée sur le paraphrasage », *Cahiers de lexicologie*, **98**, 197–211 (2011)
77. N. Gader, S. Ollinger, A. Polguère, « One Lexicon, Two Structures : So What Gives ? », dans H. Orav, C. D. Fellbaum & P. Vossen (dir.) *Proceedings of the Seventh*

- Global Wordnet Conference (GWC2014)*, 163–171 (Global WordNet Association, Tartu, 2014)
78. I. Mel'čuk, A. Polguère, « La définition lexicographique selon la lexicologie explicative et combinatoire », *Cahiers de lexicologie* (à paraître)
79. Pausé, M.-S., « Modélisation de la structure lexico-syntaxique des locutions au sein d'un réseau lexical », dans M. Kauffer & Y. Keromnes (dir.) *Actes du colloque international « Approches théoriques et empiriques en phraséologie »* (Eurogermanistik Series, Stauffenburg Verlag, Tübingen, à paraître)
80. A. Polguère, « Lexicon Embedded Syntax », *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, 2–9 (Uppsala, 2015)
81. S. Ollinger, *Le raisonnement analogique en lexicographie, son informatisation et son application au Réseau Lexical du Français* (Thèse de doctorat, Université de Lorraine, Nancy, 2014)
82. D. Larsen-Freeman, L. Cameron, *Complex Systems and Applied Linguistics* (Oxford University Press, Oxford, 2008)
83. A. G. Huth, W. A. de Heer, T. L. Griffiths, F. E. Theunissen, J. L. Gallant, « Natural speech reveals the semantic maps that tile human cerebral cortex », *Nature*, **532**, 453–458 (2016)
84. F. Ingrosso, A. Polguère, « How Terms Meet in Small-World Lexical Networks : The Case of Chemistry Terminology », dans T. Poibeau & P. Faber (dir.) *Proceedings of the 11th International Conference on Terminology and Artificial Intelligence (TIA 2015)*, 167–171 (Granada, 2015)

¹ Dans l'approche de la linguistique systémique, le lexique et la grammaire sont deux modules totalement interreliés, dont l'union est désignée par le terme *lexicogrammar*, le terme *grammar* étant généralement utilisé comme un raccourci pour *lexicogrammar* [1 : p. 7].

² Rappelons qu'une application algébrique linéaire est une application « qui peut être écrite en termes du premier degré des variables » [5] et qui tire son nom du fait qu'elle se représente dans l'espace euclidien par une droite.

³ Les notions tout particulièrement importantes pour l'exposé sont signalées en **police sans empattement**. De plus, nous utilisons les conventions d'écriture habituelles de l'approche Sens-Texte pour signaler typographiquement les différents types d'entités linguistiques (signifiés linguistiques, mots-formes, lexèmes, locutions, vocables, etc.).

⁴ Par *entité plus abstraite*, nous voulons dire que le lexème relève d'un degré d'abstraction encore plus élevé que le mot-forme, relativement à la manifestation orale ou écrite du signifiant lexical. Rappelons que le mot-forme, en tant que signe linguistique au sens saussurien, est déjà une entité informationnelle très abstraite, postulée par la théorie linguistique.

⁵ Dans la terminologie sémiotique de C. S. Peirce, la sémiose est le processus sémiotique : celui dans lequel le signe est interprété et, donc, fonctionne effectivement en tant que vecteur de sens [16].

⁶ Voir la définition donnée par D. Sperber et D. Wilson : « A concept, as we understand the term, is an enduring elementary mental structure, which is capable of playing different discriminatory or inferential roles on different occasions in an individual's mental life » [23 : 189]. Nous élargissons cette notion de concept aux entités informationnelles potentiellement éphémères, impliquées ponctuellement dans des inférences.

⁷ Pour ne citer qu'un exemple : « The term mental lexicon was introduced by Oldfield in 1966 (Oldfield 1966; in Singleton 1999) » [25 : p. 36].

⁸ On cite aussi souvent les travaux antérieurs de A. Treisman, présentés notamment dans [26] comme étant les premiers à traiter de la notion de lexique mental. Là encore, cependant, il s'agit avant tout de prémisses d'un concept qui sera terminologisé et notionalisé plus tard, en tant que lexique mental.

⁹ Nous ne pensons par trahir le contenu de la citation de G. Miller en considérant qu'au lieu de la seule syntaxe, c'est toute la grammaire de la langue qui est ici visée.

¹⁰ Voir [30] pour une présentation de la méthode d'extraction des collocations du « Dictionnaire des cooccurrences » d'Antidote.

¹¹ Notre affirmation que la lexicologie n'est pas une science expérimentale ne doit pas être interprétée comme idéologiquement motivée. Nous ne faisons que constater, à gros traits, un état de fait que la simple consultation des publications dans les domaines concernés permet de mettre en évidence. La frontière entre approches descriptive et expérimentale n'est cependant pas étanche, notamment si l'on adopte une perspective véritablement multidisciplinaire sur le lexique, et nous ne défendons aucunement une stricte compartimentalisation des disciplines.

¹² <http://wordnet.princeton.edu>

¹³ Dans ce qui suit, nous employons de façon systématique la terminologie du projet WordNet dans sa version anglaise, écrite en italique, afin de préserver la clarté notionnelle de l'exposé.

¹⁴ Sur la distinction entre champ sémantique et champs lexical et sur l'intérêt respectif de ces classes en lexicologie/-graphie, voir [49].

¹⁵ Dans la terminologie de la théorie des graphes, on dira que le *degré moyen* (nombre moyen d'arcs entrants/sortants) des sommets est faible.

¹⁶ Le DÉS est consultable en ligne, sur le site du CRISCO (Université Caen Normandie), à l'adresse <http://www.crisco.unicaen.fr/des/>.

¹⁷ Une clique d'un graphe G, en théorie des graphes, est un sous-graphe complet de G, c'est-à-dire un une partie de G dont tous les sommets sont connectés directement les uns aux autres.

¹⁸ L'analyse proxémique de DicoSyn et sa visualisation sont disponibles à partir du site du Centre National de Ressources Textuelles et Lexicales (CNRTL) : <http://www.cnrtl.fr/proxemie/>.

¹⁹ Nous reviendrons sur la notion de fonction lexicale dans la section 4.2. **Magn₂^{quant}** est une fonction lexicale, dite *syntagmatique*, qui représente un collocatif intensificateur exprimant le fait que le second actant de la lexie intensifiée dénote une entité qui se trouve être en grand nombre. Dans le cas présent, le second actant de VOCABULAIRE correspond aux éléments du vocabulaire (le premier actant étant la personne possédant/utilisant le vocabulaire en question). Un vocabulaire riche est un vocabulaire contenant beaucoup d'éléments lexicaux.

²⁰ <http://olst.ling.umontreal.ca/dicouebe/>

²¹ <http://olst.ling.umontreal.ca/dicopop/>

²² Pour une présentation concise des fonctions lexicales, on pourra aussi consulter la page Wikipedia « lexical function » : https://en.wikipedia.org/wiki/Lexical_function.

²³ Une première version de ce travail a paru en 2006 dans [74].

²⁴ <http://www.mvs.fr>

²⁵ Au moment où nous écrivons, le RL-fr comporte 27 079 sommets (lexies et entités lexicales assimilées) connectés par 69 872 arcs de relations lexicales. Les données linguistiques du RL-fr sont sous licence libre, distribuées pour l'instant de la main à la main en attendant qu'un protocole de distribution public soit mis en place. De plus, une interface de visualisation et de navigation graphique sera prochainement disponible. Elle existe pour l'instant sous forme de prototype et a été utilisée pour générer les visualisations des extraits du graphe du RL-fr données dans le présent article.

²⁶ Travail coordonné par Svetlana Krylosova (INALCO, Paris).

²⁷ L'existence de ce lien en français est à mettre en relation avec la présence d'expressions telles que *rouge comme une fraise*, [*couleur*] *rouge fraise*, etc.

²⁸ Plus de huit mille arcs de copolysémie ont à ce jour été tissés au sein du graphe du RL-fr.

²⁹ Cette observation est à rapprocher de ce qui a été dit à la section 1.1 à propos de la lexico-grammaire de la linguistique systémique.

³⁰ Pour une approche didactique de la connaissance linguistique, en tant que système complexe, voir [82].

³¹ On pourrait ainsi faire l'hypothèse que les signifiés de mots-formes (les formes lexicales) jouent un rôle structurant relativement important dans le cas du lexique physiologique – comme une des clés d'accès à la mémoire lexicale –, alors qu'ils sont en arrière-plan dans la structure du lexique logique, sémantiquement organisé.

³² Merci à S. Kahane d'avoir attiré notre attention sur ces travaux.

³³ La nécessité de la présence de cette double flèche entre observables dans notre schéma nous a été signalée par B. Gaume.

³⁴ <http://kodexlab.com>