



HAL
open science

Modèle lexicographique de croissance du vocabulaire fondé sur un processus aléatoire, mais systématique

Alain Polguère, Dorota Sikora

► **To cite this version:**

Alain Polguère, Dorota Sikora. Modèle lexicographique de croissance du vocabulaire fondé sur un processus aléatoire, mais systématique. Claudine Garcia-Debanc, Caroline Masseron, Christophe Ronveaux. Enseigner le lexique, Presses Universitaires de Namur, pp.35-63, 2013. ⟨hal-00875192⟩

HAL Id: hal-00875192

<https://hal.science/hal-00875192v1>

Submitted on 21 Oct 2013

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.



HAL Authorization

Modèle lexicographique de croissance du vocabulaire fondé sur un processus aléatoire, mais systématique

Alain Polguère* et Dorota Sikora**

* Université de Lorraine, ATILF, UMR 7118, Nancy, F-54000, FRANCE

** Université du Littoral-Côte d'Opale, HLLI, EA 4030, Dunkerque, F-59140, FRANCE

1. Introduction

1.1 Perspective lexicographique sur l'acquisition du vocabulaire

La réflexion sur le vocabulaire que nous proposons est non conventionnelle, dans le contexte de la didactique des langues, car elle s'insère dans une perspective lexicographique de construction d'une ressource lexicale – le *Réseau Lexical du Français* (Lux-Pogodalla & Polguère 2011 ; Gader et coll. 2012). Nous nous focalisons ici sur un élément du travail lexicographique qui, nous le pensons, est particulièrement pertinent pour une réflexion sur le vocabulaire, son apprentissage et son enseignement : il s'agit de la spécification et du processus de croissance de la nomenclature du Réseau Lexical du Français.

Comme l'indique son nom, le Réseau Lexical du Français, désormais *RLF*, modélise le lexique non pas par un texte dictionnaire, mais sous forme d'un réseau extrêmement riche et complexe d'unités lexicales. Le travail lexicographique de construction du RLF vise à créer un réseau lexical répondant aux trois objectifs suivants :

1. il permet de contrôler de façon rationnelle la croissance du RLF dans le cours de sa construction ;
2. son processus de croissance reflète une logique possible de la croissance du vocabulaire d'un individu ;
3. il peut servir de support à une navigation véritablement linguistique dans le système lexical de la langue.

Avant d'entrer dans le vif du sujet, il est nécessaire d'introduire quelques précisions terminologiques portant sur les notions fondamentales dont nous ferons usage dans le reste de l'article (1.2). Nous résumerons ensuite les principales caractéristiques du RLF, en tant que ressource lexicale (1.3).

1.2 Précisions terminologiques

Le cadre linguistique théorique dans lequel nous nous situons est celui de la *Lexicologie Explicative et Combinatoire* (Mel'čuk et coll. 1995), qui est le module lexical de la *théorie Sens-Texte* (Mel'čuk 1997 ; Milićević 2006). Dans ce contexte, nous faisons usage d'une terminologie bien spécifiée, dont un petit sous-ensemble doit être brièvement introduit. Nous nous concentrons bien évidemment sur les termes dénotant les différents types d'entités lexicales manipulées dans le cadre d'une activité lexicographique¹.

1 On pourra se reporter à Polguère (2008) pour une présentation des notions fondamentales (concepts et termes

L'unité d'étude en lexicologie et de description en lexicographie est l'**unité lexicale**, désignée dorénavant de façon plus compacte par le terme de **lexie**. Il existe deux types formels majeurs de lexies :

1. les **lexèmes**, qui se manifestent dans les énoncés en tant que signes linguistiques du type mots-formes – JUTEUX I [*couper une pêche juteuse dans son verre de vin.*], JUTEUX II [*tremper dans une affaire juteuse*], etc. ;
2. les **locutions**, qui se manifestent dans les énoncés en tant que syntagmes non compositionnels – HUILE DE COUDE [*Ce travail va demander un peu plus d'huile de coude.*], VIDER SON SAC [*Elle a vidé son sac devant les journalistes.*], etc.

Comme on le voit ci-dessus, nous écrivons les noms de lexies – lexèmes ou locutions – en petites majuscules, accompagnés si nécessaire de numéros lexicographiques distinctifs. Les lexies qui, comme JUTEUX I et JUTEUX II ci-dessus, se manifestent par les mêmes signifiants linguistiques et possèdent une intersection de sens significative sont regroupées au sein de **vocables polysémiques** – ici, le vocable JUTEUX, qui est égal à l'ensemble de lexies { JUTEUX I, JUTEUX II }. On identifie dans chaque vocable polysémique une **lexie de base**, qui est le « noyau sémantique » autour duquel se développe la structure polysémique du vocable. Dans le cas du vocable JUTEUX, il s'agit de JUTEUX I, dont JUTEUX II est dérivé par métaphore. Bien entendu, la numérotation des acceptions doit refléter le caractère primordial de l'acception de base ainsi que les écarts sémantiques entre toutes les lexies du vocable².

Les dictionnaires et autres ressources lexicales ont normalement comme **entrées** dans leur **nomenclature** les vocables de la langue. Il est cependant essentiel de bien garder à l'esprit que ce sont les lexies elles-mêmes – entités tout à fait circonscrites sur le plan sémantique – qui sont les unités de connaissance lexicale et, donc, les unités d'étude en lexicologie et lexicographie.

La lexie (lexème ou locution) peut-être vue comme une généralisation du signe linguistique et elle se modélise selon le même patron de description que ce dernier ; il s'agit, selon la conception Sens-Texte du signe linguistique, de décrire les trois composants fondamentaux suivants :

- **signifié** ;
- **signifiant** ;
- **combinatoire restreinte**.

Le dernier composant, la combinatoire restreinte, correspond à toutes les propriétés de combinatoire du signe linguistique – donc de la lexie – qui ne peuvent être entièrement déduites ni de son signifié ni de son signifiant (partie du discours, genre pour les noms en français, régime syntaxique, collocations, etc.).

Nous nous arrêtons là dans notre bref survol, ayant couvert la terminologie dont nous aurons à faire usage dans la discussion qui suit. Une fois cette terminologie élémentaire introduite, il convient notamment de ne plus utiliser le terme *mot* qu'avec une extrême prudence, car il ne s'agit justement pas d'un élément de la terminologie linguistique. Il pourra éventuellement s'employer pour désigner des entités dont la nature linguistique exacte n'est pas clairement identifiée. Par exemple, quand on parle des « listes de mots » construites pour l'enseignement du vocabulaire par des chercheurs ou pédagogues, qui n'indiquent pas toujours avec précision de quel type d'entités lexicales il s'agit : vocables, lexies, mots-formes, etc.

associés) de lexicologie.

2 Dans l'approche de la Lexicologie Explicative et Combinatoire, la numérotation en chiffres romains indique des écarts de type métonymique ou métaphorique.

1.3 Caractérisation générale du RLF

Le projet lexicographique sur lequel nous appuyons notre réflexion se situe dans la lignée des travaux menés sur les dictionnaires dits explicatifs *et combinatoires*, notamment le *Dictionnaire explicatif et combinatoire du français contemporain* (Mel'čuk et coll. 1984, 1988, 1992, 1999). Le RLF se distingue cependant radicalement des travaux antérieurs de ce type par sa structure et la méthodologie lexicographique mise en œuvre pour le construire.

Sur le plan de la structure, nous avons mentionné en 1.1 que le RLF n'est pas un dictionnaire, mais un *réseau lexical* dont les nœuds sont les lexies de langue et dont les arcs correspondent aux liens paradigmatiques (c'est-à-dire, sémantiques) et syntagmatiques (c'est-à-dire, de combinatoire restreinte) connectant ces lexies. Une telle structuration présente de nombreux avantages par rapport aux modèles traditionnels textuels que sont les dictionnaires, notamment :

- Un graphe lexical est une structure qui, selon toute vraisemblance, se rapproche beaucoup plus de la structure du *lexique mental* (Aitchison 2003) que la « plate » page de texte. On doit donc pouvoir effectuer sur une telle structure des opérations de navigation, d'inférence analogique, etc., qui s'approchent de celles effectuées de façon quasi instantanée par le locuteur lorsqu'il accède à ses connaissances lexicales.
- Un tel modèle n'est pas lié à un format de présentation spécifique, forçant un choix et un agencement fixes de l'information lexicale. Par contre, s'il est bien conçu, il peut être exploité pour la génération automatique de pages de dictionnaire aux contenus et formats variés. L'approche lexicographique du RLF relève ainsi de ce que l'on appelle la *lexicographie des dictionnaires virtuels* (Atkins 1996 ; Polguère 2012a).

Sur le plan de la méthodologie lexicographique, le projet de construction du RLF se distingue des travaux antérieurs en Lexicographie Explicative et Combinatoire par le recours à un éditeur lexicographique dédié – appelé *Dicet* – qui permet de construire la ressource lexicale non en écrivant des articles lexicographiques, mais en tissant des liens lexicaux : il s'agit donc d'un éditeur de graphes. La présentation de l'éditeur Dicet et des implications qu'a son utilisation sur le travail lexicographique sort du cadre de la présente discussion. On pourra se reporter à Gader et coll. (2012) pour la description de Dicet dans son état actuel et à Polguère (2012b) pour des réflexions sur la nature particulière du nouveau « geste lexicographique » que Dicet permet de mettre en œuvre.

Un des premiers problèmes auxquels il a fallu faire face dans le cadre de la construction du RLF a été celui de l'identification de sa nomenclature cible, sachant qu'il était impératif de fonctionner, sur le long terme, par étapes successives : des lexies les plus centrales du français, en s'étendant graduellement vers la périphérie du lexique de cette langue. La réflexion menée sur la nomenclature projetée pour le RLF et les différentes stratégies qui ont été élaborées dans ce contexte alimentent la discussion qui va suivre. Nous entamons cette dernière par un examen de la notion de vocabulaire de base, considérée dans une perspective didactique.

2. Vocabulaires de base en didactique de langues

2.1 Des listes de mots aux vocabulaires de base

La construction de listes de vocabulaire à usage pédagogique est une pratique qui a pris beaucoup d'importance au siècle dernier, pour ce qui est de l'enseignement du français et de l'anglais³. Ainsi,

3 Il ne nous appartient pas de faire ici un historique de la question. Pour le français, Galisson (1971 : 8–9) remarque que l'idée d'établir des listes de vocabulaires pédagogiques est ancienne puisque au XVIII^e, notamment, l'Abbé de l'Épée a dressé une liste de 5 400 mots nécessaires selon lui pour l'éducation des enfants sourds-muets. On trouvera d'autres informations sur l'historique des listes de vocabulaires pédagogiques pour le français ou l'anglais dans

Odgen (1930) propose un anglais de base, c'est-à-dire une sélection de 850 vocables d'anglais simplifié à usage international. Dans les années cinquante, les auteurs du *Français fondamental* ont entrepris de sélectionner un vocabulaire qui, à la fois, « fournit des moyens d'expression suffisants » et constitue « une première étape en vue de l'acquisition du “ français complet ” » (Gougenheim et coll. 1959 : 11).

Si les méthodes mises en œuvre par aussi bien par Odgen (1930) que par Gougenheim et coll. (1959) ont suscité de nombreuses critiques⁴, l'utilité de listes de vocabulaire dans l'enseignement d'une langue première (L1) et dans celui d'une langue seconde (L2) semble difficilement contestable. L'exemple de la *Oxford Wordlist* (Lo Bianco et coll. 2008)⁵, par exemple, illustre le chemin parcouru. L'outil, mis en place au terme des recherches menées entre 2007 et 2009 auprès des élèves et des étudiants australiens, permet de générer des listes des vocables les plus fréquents selon des critères géographiques, sociologiques, ethniques ou selon le parcours scolaire.

Le développement des études sur corpus et des méthodes statistiques d'analyse de fréquences ont permis d'affiner la notion de vocabulaire de base. Les vocables dont les occurrences ont les taux de fréquence les plus élevés peuvent à juste titre être considérés comme les plus utiles aussi bien pour un locuteur natif que pour un apprenant de L2.

Une démarche purement statistique aurait cependant des effets pervers. En s'appuyant sur des recherches menées sur l'anglais, McCarthy (1999) observe par exemple que *Saturday* a une fréquence d'emploi nettement supérieure à celle de *Wednesday*. Or, il serait étrange de dire que *Saturday* est plus important pour la maîtrise de l'anglais que *Wednesday*, alors que les noms des jours de la semaine forment un paradigme qui n'est pertinent que si on le prend comme un tout. McCarthy (1999) parle ainsi de **vocabulaire de base** en tant que partie du lexique d'une langue qui se caractérise par le potentiel communicatif le plus élevé : il est déterminé sur des bases fréquentielles, certes, mais revu et corrigé sous l'angle d'une pertinence pédagogique.

2.2 Évaluation quantitative d'un vocabulaire de base

La stratégie d'élaboration d'une nomenclature minimale pour une ressource lexicale telle que le RLF repose sur la thèse selon laquelle un vocabulaire de base recouvre le noyau dur du lexique d'une langue. Les études apportant une évaluation quantitative d'un vocabulaire de base dans le contexte de l'apprentissage d'une L2, notamment du français, et de l'acquisition d'une L1 ont en commun l'emploi du terme *mot* : la taille des vocabulaires est mesurée en nombres de « mots ». Il convient de réinterpréter *mot* dans la terminologie de la Lexicologie Explicative et Combinatoire, qui est celle mise en œuvre dans le projet lexicographique du RLF.

Dans le contexte de l'acquisition du français langue maternelle, Picoche (1995 : 369) indique que « l'essentiel des échanges verbaux s'effectue à partir de 3 000 mots du vocabulaire fondamental ». Elle précise par ailleurs, dans Picoche (2007 : 95⁶), que « le vocabulaire dont un individu a besoin pour s'exprimer sans difficulté dans la vie courante n'est que de quelques milliers de mots et que 8 ou 10 000 mots sont déjà une belle richesse »⁷. En discutant les différents aspects de la compétence lexicale, Picoche (1995, 2007) insiste sur le caractère polysémique des « mots ». On ne peut connaître un « mot » qu'une fois qu'on maîtrise ses différents sens, la combinatoire de chacun

Déro (1998 : Chapitre 3) et Chetouani (1997).

4 Pour une discussion critique, notamment du *Français fondamental*, voir Galisson (1971) et, plus récemment, Cortier & Bouchard (2008).

5 <http://www.oxfordwordlist.com/pages/>

6 Il s'agit du numéro de page du fichier électronique de Picoche (2007) ; cette édition électronique d'un ouvrage paru originellement en 1993 n'est pas paginée.

7 Le premier chiffre avancé par Picoche est très proche de celui que proposaient Gougenheim et coll. (1964) : leur français fondamental de premier et de second degré comprend en effet environ 3 500 entrées.

d'entre eux, ses dérivés morphologiques, ses connotations, etc.. Cela indique que le « mot » chez Picoche correspond approximativement au vocable tel qu'il est défini par la Lexicologie Explicative et Combinatoire. Il en ressort donc qu'un vocabulaire fondamental du français, celui qui permet d'assurer les échanges quotidiens, compte environ 3 000 vocables.

Dans le domaine de l'apprentissage d'une L2, le « mot » recouvre généralement une *famille de mots*, notion introduite par Nation (1982) pour servir d'unité de mesure dans l'évaluation des connaissances lexicales de l'apprenant, la croissance de son vocabulaire et la progression de la maîtrise d'une L2 qui en résulte. Une famille de mots comprend toutes les formes fléchies, ainsi que les dérivés forgés à l'aide des affixes les plus fréquents. Soit le cas du verbe DIRIGER ; il conditionne la famille de mots suivante : le verbe REDIRIGER, le verbe pronominal SE DIRIGER, les adjectifs DIRIGEANT_{Adj}, DIRECTEUR_{Adj}, DIRIGÉ, les noms DIRECTION, DIRIGEANT_N, DIRECTEUR_N, DIRIGEANTE, DIRECTRICE. Par ailleurs, on s'accorde à considérer, dans le domaine de l'apprentissage d'une L2, que la connaissance d'un vocable comme DIRIGER englobe la maîtrise des propriétés de chacune de ses acceptions – sens et combinatoire restreinte –, même si ces différents composants ne sont pas nécessairement désignés en ces termes.

La famille de mots est une famille de vocables, pour utiliser notre terminologie, et forts de ce constat, nous avons étudié les données quantitatives relatives aux vocabulaires de base en L2. Il s'avère que les chiffres avancés pour l'anglais (Laufer 1996 ; Adolphs & Schmitt 2003) et pour le français (Milton 2009) sont quasiment identiques. En effet, selon ces études, le niveau seuil de connaissances lexicales se situe, pour un apprenant étranger, à environ 3 000 vocables qui assurent une « couverture textuelle » à hauteur de 95%. Autrement dit, un apprenant avec un bagage lexical de 3 000 vocables est en mesure de comprendre un texte non technique en L2 (Laufer 1996). La compréhension de textes oraux n'est pas moins exigeante. En s'appuyant sur les travaux de Adolphs & Schmitt (2003), Milton (2009) constate que la communication orale requiert également un vocabulaire de 3 000 de vocables.

3. Des « familles de mots » aux dérivés sémantiques proches

Dans cette section, nous proposons une alternative plus rigoureuse et intuitivement plus satisfaisante aux familles de mots : les associations de dérivés sémantiques proches. Ces associations jouent un rôle central dans notre stratégie d'élaboration d'une nomenclature de base pour la construction d'une ressource lexicale comme le RLF, ainsi que pour implanter une stratégie d'enseignement du lexique.

3.1 Retour sur la dérivation

Commençons par effectuer un bref retour sur la dérivation en langue. Après avoir apporté quelques précisions sur les notions de dérivation morphologique, synchronique et diachronique, nous serons en mesure d'explicitier la notion clé de dérivation sémantique, à partir de laquelle sera construite celle de dérivé sémantique proche.

3.1.1 Dérivation morphologique et modèles lexicaux

La dérivation en langue est habituellement comprise comme une *dérivation morphologique*. Cette dernière peut être caractérisée très schématiquement de la façon suivante : on considère qu'il existe un lien de dérivation morphologique entre deux lexies L_1 (source de la dérivation) et L_2 (cible) – lien noté $L_1 \xrightarrow{m} L_2$ – lorsque L_2 peut être décrite comme (i) sémantiquement et (ii) morphologiquement construite à partir de L_1 . Cette caractérisation est volontairement très générale, presque simpliste, afin d'être compatible avec les différentes approches théoriques de la dérivation morpho-

logique, notion qu'il ne nous appartient pas d'examiner en détail ici⁸.

On notera que la dérivation morphologique, en tant que mécanisme langagier, peut être conçue de deux façons.

1. Il peut s'agir d'une **dérivation synchronique**. C'est alors un mécanisme morphologique d'application systématique, que le locuteur peut utiliser afin de construire une lexie « à la demande », comme dans le cas de la dérivation française par préfixation en RE- 'de nouveau' : DÉCHIRER ^m → REDÉCHIRER, DESSINER ^m → REDESSINER, etc.
2. Il peut s'agir d'une **dérivation diachronique**. C'est alors un mécanisme qui permet de créer des néologismes, qui vont éventuellement se lexicaliser dans la diachronie. Ainsi, la dérivation morphologique par suffixation mise en jeu dans BIGLER ^m → BIGLEUX n'existe que dans la diachronie du français. Un locuteur qui dit *Ce type est bigleux* n'a pas appliqué un mécanisme de dérivation par suffixation en -EUX pour construire l'adjectif BIGLEUX. Il a simplement utilisé un adjectif déjà présent dans le lexique de la langue. Si le locuteur applique la règle de dérivation en question selon son bon vouloir, il crée un néologisme, comme dans le cas d'une dérivation LOUCHER ^m → *LOUCHEUX.

Il est important de souligner que, lorsqu'on étudie la dérivation en lexicologie française, on s'intéresse en général à la dérivation morphologique – telle que nous venons de la décrire –, et qu'il s'agit dans l'immense majorité des cas de dérivations diachroniques : en français, les règles de dérivation morphologique sont pour la plupart d'application non systématique et, lorsqu'appliquées, servent à créer des néologismes, c'est-à-dire des lexies qui vont être perçues comme des créations du locuteur⁹.

Il est bien évident qu'un modèle lexicographique doit rendre compte des dérivés diachroniques, lexicalisés dans la langue, mais qu'il ne lui appartient pas de prendre en charge la dérivation synchronique. Cette dernière est prise en charge par la composante morphologique de la grammaire de la langue. Ce n'est que lorsqu'une dérivation morphologique normalement synchronique se lexicalise, c'est-à-dire lorsque la lexie dérivée se trouve associée à des particularités qui ne peuvent être calculées par la règle grammaticale de dérivation, qu'il convient de la prendre en compte dans le modèle lexical. Les particularités en question sont principalement de deux ordres.

1. Individualisation sémantique : le sens de la lexie dérivée n'est pas directement calculable à partir (i) du sens de la source de la dérivation et (ii) du sens du signe morphologique utilisé pour opérer la dérivation. Par exemple, le sens de RELIRE [*Je cherche quelqu'un pour relire mon article.*] est différent de 'lire de nouveau'. Dans un tel cas, on doit considérer que la lexie véritablement dérivée de LIRE – *J'ai lu son article, et je l'ai même relu deux fois* – a donné lieu à une expansion polysémique, qui impose la modélisation d'un vocable RELIRE comportant au moins deux acceptions : RELIRE I 'lire qqch. de nouveau' (morphologiquement transparent) et RELIRE II 'lire un texte pour le corriger ou l'améliorer'.
2. Individualisation sur le plan de la combinatoire : la lexie dérivée possède une combinatoire propre, qui ne peut être complètement expliquée par la combinatoire de la source de la dérivation et par la règle qui l'a produite. Par exemple, la forme participiale *connu* dans *C'est quelqu'un de très connu* possède une combinatoire lexicale propre : elle contrôle notamment des collocations du type *notoirement connu* ou *connu comme le loup blanc*. Il s'agit donc

8 Soulignons, de plus, que nous raisonnons sur une dérivation morphologique qui s'appliquerait aux lexies, alors que, strictement parlant, il s'agit d'un mécanisme mettant en jeux des entités **morphologiques** – les mots-formes et leurs morphèmes constitutifs. Pour un examen de la dérivation morphologique en langue, voir notamment Mel'čuk (1994 : Chapitre IV).

9 Cela n'est pas du tout le cas d'une lexie comme REDÉCHIRER, que l'on ne qualifiera pas de néologisme : sans être présente dans les dictionnaires, elle est de façon virtuelle dans le lexique dans la langue, du fait du caractère systématique de la dérivation qui la construit.

bien d'une lexie adjectivale (et non d'un « emploi adjectival d'un participe »), qu'il convient de décrire dans un article lexicographique consacré à la lexie en question.

Ces faits sont bien connus des lexicographes, qui sont nécessairement confrontés, lors de la construction de la nomenclature de leurs modèles lexicaux, au problème du diagnostic du caractère lexicalisé ou non des dérivés morphologiques. Les décisions doivent être prises au cas par cas, en menant une analyse de chaque dérivé avant son insertion éventuelle dans la nomenclature.

3.1.2 Dérivation sémantique

Dans son ouvrage sur l'enseignement du vocabulaire (Picoche 2007), J. Picoche fait le commentaire suivant à propos de la dérivation :

« Le fait que les dérivés opèrent une sélection parmi les sens du mot de base a pour conséquence que certains sens de ce mot restent sans dérivé. C'est, à côté de beaucoup de circonstances fortuites, une des raisons pour lesquelles il existe de nombreux mots qui, sans être morphologiquement des dérivés, fonctionnent pourtant comme des dérivés : Si le nom d'action correspondant à *retomber* est *retombée*, le nom d'action correspondant à *tomber* est *chute* ; *Max gifle Jean* se transforme en *Max donne une gifle à Jean* mais *Max frappe Jean* en *Max donne un coup (de poing) à Jean* ; *manger* peut être transformé en *faire un repas* ; *dormir* en *faire un somme* ; *se tromper* en *faire une erreur*. Il existe donc des familles de mots non morphologiques, mais purement sémantiques. À côté de la série totalement homogène *fin, finir, final, finalement*, on peut aligner une série morphologiquement hétéroclite mais sémantiquement tout aussi homogène : *terme, achever, dernier, enfin*¹⁰. »

Ce que dit ici J. Picoche est tout à fait limpide et a des implications tellement évidentes sur la façon dont il convient d'envisager l'enseignement et l'apprentissage de la langue que l'on peut s'étonner que la lexicologie et la didactique du français continuent de se focaliser sur une dérivation strictement morphologique. Il est pourtant de bon sens que cette dernière doit être intégrée dans un cadre plus vaste. La Lexicologie Explicative et Combinatoire (Mel'čuk et coll. 1995) concorde entièrement avec la perspective de J. Picoche sur la dérivation – pour des raisons aussi bien théoriques que pédagogiques – et complète la dérivation morphologique par une notion plus large, à caractère véritablement lexicologique : la **dérivation sémantique**, que nous noterons $L_1^s \rightarrow L_2$ pour la distinguer de la dérivation morphologique (notée $L_1^m \rightarrow L_2$)¹¹.

On trouve dans Mel'čuk & Polguère (2007 : 18–19) une présentation élémentaire de la dérivation sémantique. Trois conditions y sont spécifiées pour qu'existe entre deux lexies L_1 et L_2 une relation de dérivation sémantique $L_1^s \rightarrow L_2$, conditions que nous reproduisons ci-dessous avec de légères adaptations :

1. L_2 est en relation sémantique forte avec L_1 ; dans le cas le plus standard, L_2 se définit en termes de L_1 .
2. La relation sémantique entre L_1 et L_2 est récurrente dans la langue.
3. La relation entre L_1 et L_2 peut s'exprimer de façon morphologique dans la langue.

On voit que la dérivation sémantique possède une couverture plus large que la dérivation morphologique puisqu'il n'est pas nécessaire qu'existe une relation morphologique entre la source L_1 et la cible L_2 . Cependant, les liens de dérivation sémantique couvrent un spectre sémantique relativement étroit : il s'agit du sous-ensemble des liens sémantiques qui sont récurrents en langue et tendent à se morphologiser. Ainsi, il ne suffit pas qu'une lexie L_2 se définisse en termes d'une autre lexie L_1 pour qu'elle soit considérée comme son dérivé sémantique. Il faut aussi que le lien sémantique soit suffi-

10 Chapitre 1, section 3 « Les familles ont une histoire », sous-section « Les simples relations ».

11 Soulignons que J. Picoche fait elle-même usage du terme *dérivé sémantique*.

samment « standard » pour que la langue ait développé des outils morphologiques permettant son expression. Illustrons cela avec la paire BOIRE [*Elle a bu sa vodka d'un trait.*] ~ VERRE [*Son verre était à moitié vide.*]. Il est justifié de dire que la lexie VERRE est un dérivé sémantique de la lexie BOIRE, car : (i) ‘verre’ s’analyse approximativement en ‘petit récipient servant à boire’ ; (ii) le sens ‘objet servant à faire L₁’ est très souvent lexicalisé ; (iii) cette lexicalisation peut très naturellement s’effectuer de façon morphologique – par exemple, (S’)ABREUVER ^{s/m}→ ABREUVOIR, HACHER ^{s/m}→ HACHOIR, TONDRE ^{s/m}→ TONDEUSE, etc.

L’écart sémantique entre deux lexies liées par une dérivation sémantique peut tout à fait être (presque) nul, leur différence se ramenant à un changement de partie du discours. Par exemple, la lexie ODEUR [*Je ne supporte pas l’odeur de ce savon.*] est un dérivé sémantique de la lexie SENTIR [*Ce savon sent mauvais.*] car : (i) ‘odeur de X’ s’analyse approximativement en ‘ce que X sent’ ; (ii) l’écart sémantique entre les deux lexies se ramène en fait à un simple changement de partie du discours, la nominalisation d’un sens verbal étant un phénomène très courant ; (iii) la nominalisation s’encode fréquemment de façon morphologique – par exemple, MARCHANDER ^{s/m}→ MARCHANDAGE, POLLUER ^{s/m}→ POLLUTION, etc. La synonymie exacte ou la conversivité – qui relèvent aussi de l’équivalence sémantique – peuvent tout autant être considérées comme des cas de dérivations sémantiques, car il s’agit de relations basiques entre lexies, pour lesquelles les langues offrent des outils morphologiques ; par exemple :

- dérivation synonymique par troncation : APPARTEMENT ^{s/m}→ APPART ;
- dérivation conversive par forme réfléchi : ÉTONNER [*Ma question étonne Éva.*] ^{s/m}→ S’ÉTONNER [*Éva s’étonne de ma question.*].

Remarquons que, dans le cas du réfléchi, si la lexie S’ÉTONNER est morphologiquement dérivée de ÉTONNER (elle est morphologiquement plus complexe), elle ne se distingue sémantiquement de cette dernière que par une permutation de structure actancielle : *X étonne Y* et *Y s’étonne de X* dénotent exactement la même situation. Il est donc légitime de considérer aussi bien une dérivation sémantique ÉTONNER ^s→ S’ÉTONNER qu’une dérivation sémantique inverse S’ÉTONNER ^s→ ÉTONNER.

Nous allons maintenant expliquer comment la notion de dérivation sémantique peut être utilisée pour construire une alternative à la notion de famille de mots (examinée plus haut, en 2.2) sur des bases qui nous semblent plus rigoureuses, théoriquement et pédagogiquement.

3.2 Dérivés sémantiques proches

3.2.1 Gradation dans la proximité sémantique des dérivés

À partir d’une lexie donnée L, on peut identifier une association de lexies {L} _{dsém}, constituée de L et de tous ses dérivés sémantiques. Une association de ce type, qui repose sur la notion bien circonscrite de dérivation sémantique, nous semble plus rigoureusement spécifiée qu’une famille de mots, qui relève en fin de compte d’une notion assez floue et dont la pertinence interlinguistique reste problématique au vu des différences considérables que les langues manifestent sur le plan morphologique.

Une association construite strictement en fonction de la dérivation sémantique est cependant potentiellement trop vaste pour constituer ce que l’on pourrait voir comme étant une « molécule » de connaissance lexicale. Ainsi, BEURRE est un dérivé sémantique de LAIT, mais on ne considérera pas, dans un contexte d’apprentissage du vocabulaire, qu’il appartient à la « famille lexicale immédiate » de LAIT, au même titre que LAITIER par exemple ; et cela, pour des raisons qui ne sont pas uniquement morphologiques. On doit donc aller encore plus loin dans la caractérisation d’une notion se substituant à celle de famille de mots, en restreignant le spectre des liens sémantiques qui est cou-

vert par les associations de dérivations sémantiques. Telle est la finalité de l'élaboration de la notion de dérivation sémantique proche. Pour introduire celle-ci, commençons par souligner une caractéristique importante de la dérivation sémantique : le caractère graduel de la proximité sémantique dont elle rend compte.

La dérivation sémantique recouvre un spectre assez large de relations sémantiques : des relations correspondant à un écart sémantique nul à celles pour lesquelles l'écart est suffisamment important pour que la présence même d'un lien de dérivation sémantique puisse être débattue. Ainsi, à une extrémité de l'échelle de gradation en termes de proximité sémantique, se trouve la proximité « extrême » qu'est l'identité de sens – c'est-à-dire la synonymie exacte : VOITURE^s → BAGNOLE. À l'autre extrémité, se trouvent des proximités « faibles » représentées, par exemple, par une dérivation sémantique adjectivale non vide et non morphologique comme S'ENDORMIR^s → SOMNOLENT.

La théorie Sens-Texte (cf. section 1.2) est, à notre connaissance, la seule approche théorique qui propose un outil formel permettant de décrire exhaustivement les différents types de liens de dérivation sémantique présents dans les langues, ainsi qu'une intégration de la description de ces *liens lexicaux paradigmatiques* avec celle des *liens syntagmatiques* instanciés dans les syntagmes phraséologiques du type base + collocatif appelés *collocations* (Hausmann 1989 ; Mel'čuk 1995), comme :

- *pluie + forte, grosse...* ; angl. *rain + driving, heavy, strong...* ;
- *clou + planter, enfoncer [un clou]* ; angl. *nail + to drive [a nail]* ;
- etc.

L'outil descriptif « unificateur » pour les liens paradigmatiques et syntagmatiques proposé par la théorie Sens-Texte est le système des *fonctions lexicales standard*. Les fonctions lexicales Sens-Texte forment un système extrêmement riche, que nous ne pouvons présenter ici, nous contentant de renvoyer le lecteur à l'abondante littérature traitant de la question¹². Il nous faut cependant faire usage ici des fonctions lexicales, ainsi que des formalisations qui leur sont associées, puisque la notion de dérivé sémantique proche que nous allons maintenant introduire repose entièrement sur la spécification d'un sous ensemble des fonctions lexicales paradigmatiques standard. Nous prendrons donc soin d'explicitier la valeur sémantico-syntaxique de chaque fonction lexicale que nous serons amenés à introduire.

3.2.2 Dérivés sémantiques proches

Le système des fonctions lexicales standard de la théorie Sens-Texte nous a permis de caractériser de façon précise un ensemble de treize liens privilégiés de dérivation sémantique connectant des lexies manifestant une très grande proximité sémantique : ce que nous appellerons désormais la *dérivation sémantique proche*. Nous identifions une relation de dérivation sémantique proche entre deux lexies L_1 et L_2 , relation notée $L_1 \text{ dsp} \rightarrow L_2$, si :

- soit L_2 est un synonyme exact de L_1 ;
- soit L_2 est une « transformée » sémantique minimale de L_1 .

Bien entendu, dire qu'une lexie est une transformée sémantique minimale d'une autre ne relève aucunement d'un critère objectif. Une part importante de subjectivité entre dans la sélection d'un ensemble donné de liens lexicaux qui, seuls, seront considérés comme correspondant à de la dérivation sémantique proche. Une fois cette sélection établie, cependant, l'inclusion ou l'exclusion d'un lien donné lexical $L_1 \rightarrow L_2$ se fera sur une base tout à fait objective : il ne s'agira de dérivation sé-

¹² Voir Mel'čuk (1996), pour une présentation générale, et Polguère (2007), pour une étude du caractère standard vs non standard des fonctions lexicales.

mantique proche que si le lien en question correspond à une des treize relations préalablement identifiées.

La liste que nous proposons a été tout d'abord le produit d'une réflexion menée à partir de l'étude du système des fonctions lexicales paradigmatiques standard et de son utilisation dans des projets lexicographiques tels que le *Dictionnaire explicatif et combinatoire du français contemporain* (Mel'čuk et coll. 1984, 1988, 1992, 1999), la base lexicale *DiCo* (Mel'čuk & Polguère 2006), etc. La pratique lexicographique montre notamment que certains liens standard ont une fréquence d'apparition dans les articles lexicographiques beaucoup plus élevée que d'autres et que, dans tous les cas, il s'agit des liens manifestant un écart sémantique très faible entre la lexie source et la lexie cible. Sans pouvoir entrer ici dans des considérations très théoriques, mentionnons que ce phénomène est tout à fait en accord avec la nature fonctionnelle profonde des langues, qui sont organisées comme des générateurs de paraphrases (Milićević 2007).

Notre liste initiale de dérivations sémantiques proches a été légèrement révisée de façon empirique, suite au travail effectué sur la nomenclature dite *d'amorçage* du RLF, qui est présentée plus bas dans la section 4. Nous donnons maintenant la liste à laquelle nous sommes finalement parvenus. Dans l'énumération qui suit, chaque dérivation sémantique proche est introduite par la fonction lexicale paradigmatique standard correspondante, suivi d'une brève description de sa valeur sémantico-syntaxique, ainsi que d'une ou plusieurs illustrations et, éventuellement, d'un commentaire. Le symbole « L » désigne ici l'**argument** de la fonction lexicale, c'est-à-dire la source de la dérivation sémantique proche.

1. **Syn** : synonyme exact de L ; par ex. VOITURE → AUTOMOBILE_N.
2. **Anti** : antonyme exact de L ; par ex., LÉGAL → ILLÉGAL.
3. **Syn_{sexe}** : quasi-synonyme (plus précisément, synonyme à intersection) de L qui dénote le même type d'individu que L, mais de sexe opposé ; par ex., ACTEUR → ACTRICE, COUSIN → COUSINE. Ce lien de fonction lexicale est parfaitement symétrique ; cependant, la lexie du genre masculin est souvent en même temps un générique, en français. D'un point de vue formel, **Syn_{sexe}** est équivalent, dans le langage formel des fonctions lexicales, à une fonction dite *semi-standard* encodée ainsi : **de sexe opposé Syn** .
4. **Fem** : lexie dénotant un être vivant femelle correspondant à l'être vivant mâle dénoté par L ; par ex., COQ → POULE. La fonction lexicale **Fem** est un « néologisme » dans le système des fonctions lexicales, introduit dans le cadre du projet de construction du RLF. Strictement parlant, elle ne fait pas partie du système des fonctions lexicales standard de la théorie Sens-Texte. Elle est fonctionnellement peu éloignée de la fonction lexicale précédente, mais s'en distingue tout de même par le fait qu'elle ne relève pas de la synonymie véritable. Elle serait dans les faits beaucoup plus proche d'une relation de contrastivité (fonction lexicale standard **Contr**).
5. **Masc** : lexie dénotant un être animé mâle correspondant à l'être animé femelle dénoté par L ; par ex., POULE → COQ. Il s'agit donc de la relation exactement symétrique de la précédente.
6. **V₀** : verbalisation de la lexie non verbale L ; par ex., BAGARRE → SE BAGARRER.
7. **S₀** : substantivisation vide de la lexie non nominale L ; par ex., PARTIR → DÉPART.
8. **Adj₀** : adjectivisation vide de la lexie non adjectivale L ; par ex., ÉCOLE → SCOLAIRE.
9. **Adv₀** : adverbialisation vide de la lexie non adverbiale (normalement, adjectivale) L ; par ex., LENT → LENTEMENT.
10. **S_i** : nom dénotant le i^e (1^{er}, 2^e, 3^e...) actant de L ; par ex., CONDUIRE ('X conduit

Y') → CONDUCTEUR [= **S**₁].

11. **A**_i : adjectif dénotant le fait d'être le i^e (1^{er}, 2^e, 3^e...) actant de L ; par ex., FATIGUE ('fatigue de X causée par Y') → FATIGUÉ [= **A**₁]. On admettra aussi comme dérivation sémantique proche les cas où l'adjectif dérivé ajoute un sens de perfectivité par rapport au sens de la source de la dérivation, comme dans le cas de DÉCOURAGER ('X décourage Y') → DÉCOURAGÉ (qui signifie '[Y] qui a été découragé par X'). Strictement parlant, il s'agit ici de la fonction lexicale standard complexe **A**₂Perf.
12. **Able**_i : adjectif dénotant le fait d'être susceptible d'être le i^e (1^{er}, 2^e, 3^e...) actant de L ; par ex., SE FIER ('X se fie à Y') → FIABLE [= **Able**₂].
13. **Mult** exact : nom collectif qui dénote spécifiquement un ensemble de L ; par ex., FEUILLE → FEUILLAGE. On notera que le collectif doit être « exact » ; il doit donc contenir le sens de L dans sa définition et signifier 'ensemble de L'. Ainsi, BANC [*de poissons, d'huitres...*] n'est pas un dérivé sémantique proche de POISSON, car c'est une lexie trop vague.

La remarque faite à propos de **Mult** doit être généralisée à toutes les autres fonctions lexicales mentionnées ci-dessus : elles sont prises dans leur « sens étroit », tel que spécifié dans les descriptions qui les accompagnent. Par exemple, strictement parlant, VICTIME [*d'un meurtre*] est un **S**₂ valide de MEURTRE [*par l'individu X de l'individu Y*], mais VICTIME ne devrait pas être considéré comme étant un dérivé sémantique proche de MEURTRE, car son sens est beaucoup plus vague que 'individu Y tué par meurtre' (*[un] meurtré).

Grâce à la liste donnée ci-dessus, nous sommes en mesure de substituer aux « familles de mots » les associations lexicales induites par la dérivation sémantique proche. Ces dernières sont des petits ensembles de lexies sémantiquement très proches, en relation de (quasi-) paraphrase, souvent, mais non nécessairement, liées morphologiquement. Nous allons maintenant voir comment les associations de dérivations sémantiques proches nous permettent de gérer la croissance de la nomenclature d'une ressource lexicale telle que le RLF.

4. Nomenclature du RLF

4.1 Construction d'une nomenclature en trois volets

Rappelons que le RLF a pour vocation, non seulement de décrire les lexies du français, mais aussi de rendre compte de la structuration même du lexique en tant que réseau de lexies. Il s'agit de modéliser le positionnement des lexies au sein d'un réseau de relations paradigmatisées et syntagmatisées. Une double piste de réflexion, à la fois didactique et lexicologique, nous a amenés vers l'hypothèse que le noyau dur du système lexical du français que doit cibler le RLF correspond à son vocabulaire de base tel qu'on le conçoit dans un contexte d'enseignement/apprentissage du lexique. La description, dans une ressource lexicale, des acceptions des vocables qui constituent un vocabulaire de base permet en effet de procéder dans un ordre qui, contrairement au cheminement alphabétique traditionnellement appliqué pour la construction des dictionnaires, est déterminé par la structure intrinsèque de la langue : on entre dans le lexique par un noyau lexical minimal servant à amorcer le processus de croissance d'une nomenclature ; on s'attache ensuite aux dérivés sémantiques proches issus de ce noyau lexical ; puis on « s'aventure » dans l'ensemble du lexique selon la logique des liens sémantiques et de combinatoire issus du vocabulaire acquis. Ce *modus operandi* présente l'avantage d'assurer une croissance de la nomenclature qui se fait d'une manière à la fois aléatoire et contrôlée par le système de liens lexicaux privilégiés que nous venons de discuter (section 3).

Il résulte de tout cela que le processus de croissance de la nomenclature du RLF s'articule en trois volets :

1. identification d'une **nomenclature d'amorçage (NA)**, qui réunit les vocables constituant le noyau du vocabulaire de base du français et qui est injectée dans la base lexicale du RLF dès sa création pour « amorcer » sa nomenclature et son processus de croissance¹³ ;
2. génération systématique d'une **nomenclature directement induite (NDI)** de la NA, qui ajoute aux vocables de la NA les dérivés sémantiques proches de leur lexie de base¹⁴ ;
3. génération aléatoire, mais logiquement maîtrisée, d'une **nomenclature indirectement induite (NII)**, qui comprend, entre autres, les vocables dont les acceptions sont des cibles de liens de fonctions lexicales issus des entrées de la NA et de la NDI.

Nous allons tout d'abord traiter du problème de l'identification de la NA (4.2), pour ensuite passer à celui de l'implantation et du contrôle de sa croissance (4.3).

4.2 Identification de la nomenclature d'amorçage (NA)

De nombreuses ressources permettent aujourd'hui d'accéder à un vocabulaire de base du français. En construisant la NA, nous avons systématiquement pris en compte les données des trois listes suivantes :

1. la nomenclature du *Robert Benjamin* (Collectif Robert 2009), un dictionnaire pédagogique destiné au cycle des apprentissages fondamentaux de l'école primaire et comportant 6 500 entrées ;
2. la liste Eduscol, disponible sur le site du Ministère de l'Éducation nationale français¹⁵ ; cette ressource repose sur des études de fréquence menées sur le *Trésor de la Langue française* par Brunet (1981) et comprend 1 463 entrées, qui doivent être travaillées tout au long du cycle des approfondissements (8 à 10 ans) ;
3. un vocabulaire du Ministère de l'Éducation, du Loisir et du Sport (MELS) du Gouvernement du Québec, élaboré à l'Université de Montréal par une équipe de didacticiens (Lefrançois et coll. 2011) à partir des études de fréquence menées sur un corpus de manuels scolaires et de littérature jeunesse ; ce vocabulaire, à l'usage des enseignants intervenant dans les premières classes du cycle secondaire, comporte 4 548 vocables¹⁶.

Une liste de 6 784 vocables a ainsi été arrêtée, avec cependant la nécessité de la soumettre à un examen critique conformément à la démarche proposée par McCarthy (1999). Il s'agissait en effet, d'une part, d'éliminer des éléments qui ne relèvent pas de vocabulaire de base et de l'autre, de rationaliser une liste de vocabulaire issue des ressources s'inscrivant dans des cadres méthodologiques bien différents. Nous avons ainsi été amenés à éliminer des vocables « parasites » tels que ANORAK – issu de la nomenclature du *Robert Benjamin* –, qui est certes fréquent dans le discours des écoliers et de leurs parents, mais dont le potentiel communicatif est peu élevé. Par ailleurs, nous avons dû introduire dans la future NA des vocables tels que SE COMPORTER, DÉCOURAGER ou MANIFESTER, absents des listes consultées, mais dont l'utilité nous apparaissait incontestable.

Pour transformer une liste de vocabulaire en nomenclature d'amorçage, il a été nécessaire d'émettre

13 Lors de l'injection de la NA dans la base RLF, un vocable a été créé automatiquement pour chaque élément de la liste, avec une acception par vocable créé. Immédiatement après l'injection, les articles lexicographiques des acceptions sont vides, si ce n'est pour l'indication d'une partie du discours.

14 Rappel : la notion de lexie de base d'un vocable a été introduite à la section 1.2.

15 Cf. <http://eduscol.education.fr/pid23250-cid50486/vocabulaire.html> (mise à jour le 9 février 2010, consultée le 25 mai 2011).

16 Nous remercions Pascale Lefrançois (Université de Montréal) et Ophélie Tremblay (Université du Québec à Montréal) de nous avoir donné accès à leurs données de recherche. Nous remercions également Caroline Bégin (MELS) et Hélène Cajolet-Laganière (Université de Sherbrooke) d'avoir autorisé la diffusion de l'information contenue dans le rapport de recherche Lefrançois et coll. (2011).

continuellement des hypothèses sur la structuration des sous-réseaux lexicaux. Ainsi, face aux quatre vocables CRÉATION, CRÉER, CRÉATEUR et CRÉATRICE, issus de la compilation des listes de référence, nous avons considéré que la lexie de base du verbe CRÉER permettait d'accéder par des liens de dérivation sémantique proche aux lexies :

- CRÉATION en tant que dérivé nominal de type S_0 ;
- CRÉATEUR et CRÉATRICE en tant que dérivés du type S_1 (noms du premier actant).

Nous avons donc sélectionné le verbe CRÉER comme pivot de l'association lexicale de dérivés sémantiques proches à insérer dans la NA. Cette démarche nous a conduits à une NA comptant 3 734 vocables.

4.3 *Modèle d'expansion de la nomenclature d'amorçage*

Les fonctions lexicales décrites dans la section 3 permettent d'étendre la NA du RLF en ajoutant à la ressource lexicale les vocables dont une acception est la cible d'un lien de dérivation sémantique proche issu de la lexie de base d'un vocable de la NA. Cette dernière est ainsi complétée par une nomenclature directement induite (NDI). Ce mode de croissance a permis d'atteindre une nomenclature NA+NDI d'environ 10 000 vocables.

Notons que la NDI inclut, en plus des vocables introduits du fait de liens de dérivation sémantique proche, tous les vocables possédant une des caractéristiques suivantes.

- Ce sont des verbes pronominaux correspondant aux verbes non pronominaux de la NA (et vice versa).
- Ce sont des homonymes de vocables de la NA.
- Ce sont des locutions formées à partir des lexèmes de la NA¹⁷. Ainsi, la NA comprend les vocables RETOURNER, TAILLER et VESTE ; ceux-ci permettent d'induire les locutions très courantes du français RETOURNER SA VESTE et TAILLER UNE VESTE, qui doivent être intégrées dans la NDI du RLF.

La nomenclature NA+NDI constitue une bonne hypothèse pour le noyau du lexique français. Le processus d'expansion de la nomenclature du RLF ne s'arrête cependant pas là. La croissance systématique déterminée par la dérivation sémantique proche se poursuit par l'ajout d'une nomenclature dite *indirectement induite* (NII). Celle-ci comprend des vocables dont les acceptions sont des cibles (i) de liens paradigmatiques plus riches que la dérivation sémantique proche et (ii) de liens syntagmatiques. Ainsi, le nom CÉLIBATAIRE, présent dans la NA, entraînera l'insertion dans la NII de l'adjectif ENDURCI, qui est un des collocatifs intensificateurs de CÉLIBATAIRE (fonction lexicale syntagmatique **Magn**). Même si cet adjectif n'est bien évidemment pas au cœur du lexique français, son insertion en phase 3 de croissance de la nomenclature est justifiée par le fait qu'il est un collocatif privilégié d'une lexie qui, elle, a été identifiée comme appartenant au lexique de base. La croissance de la nomenclature du RLF se poursuit ainsi graduellement, au fur et à mesure que les lexicographes sont amenés à tisser de nouveaux liens lexicaux issus des vocables déjà présents dans la nomenclature¹⁸.

La Figure 1 ci-dessous visualise une partie du processus d'expansion de la NA à partir de la lexie de base du vocable DÉTRUIRE.

17 La NA du RLF, comme la plupart des vocabulaires de base, est exclusivement constituée de lexèmes (cf. section 1.2). Les locutions étant des lexies à part entière, elles doivent apparaître dans la nomenclature du RLF et il faut donc prévoir un processus permettant leur insertion dans la base lexicale. C'est ce processus que nous caractérisons brièvement ici.

18 Pour une présentation de la méthodologie de description des vocables du RLF, voir Polguère (2011).

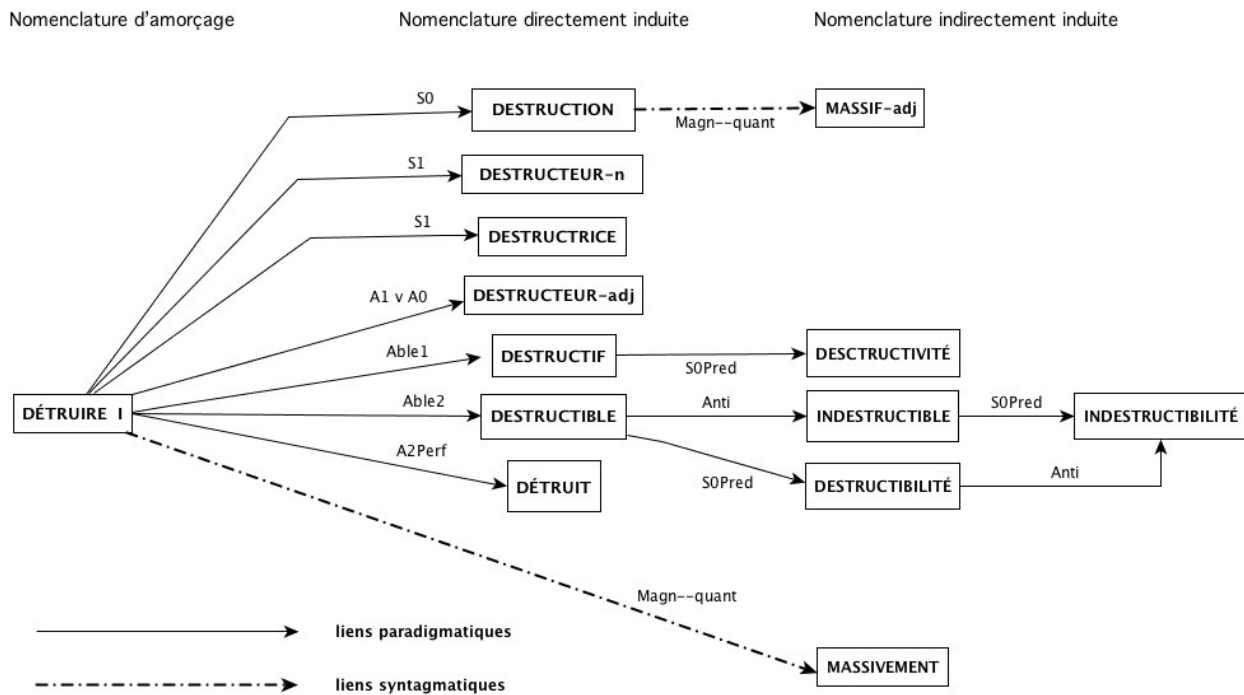


Figure 1 : Schéma de croissance de la nomenclature du RLF

Une fois la NA injectée dans le RLF, le tissage des liens de dérivation sémantique proche a été effectué de façon systématique sur toute cette nomenclature initiale ; puis s'est engagé le processus d'expansion quasi organique de la base, conditionné par le travail lexicographique proprement dit de description des vocables. Le graphe ci-dessous, Figure 2, résume l'évolution de la couverture lexicale du RLF, du début (ou presque) du tissage des liens de fonctions lexicales jusqu'au moment de la rédaction du présent texte. La croissance lente, mais régulière, du taux de polysémie – nombre de vocables divisé par le nombre de lexies (acceptions) – correspond au dégagement progressif de la structure polysémique des vocables de la base. La décroissance ponctuelle que l'on peut constater sur plusieurs mois a coïncidé en grande partie avec la phase de tissage des liens de dérivation sémantique proche à partir de la NA, qui a mené à l'introduction de la NDI et donc a généré un apport soudain de nouveaux vocables dans la base.

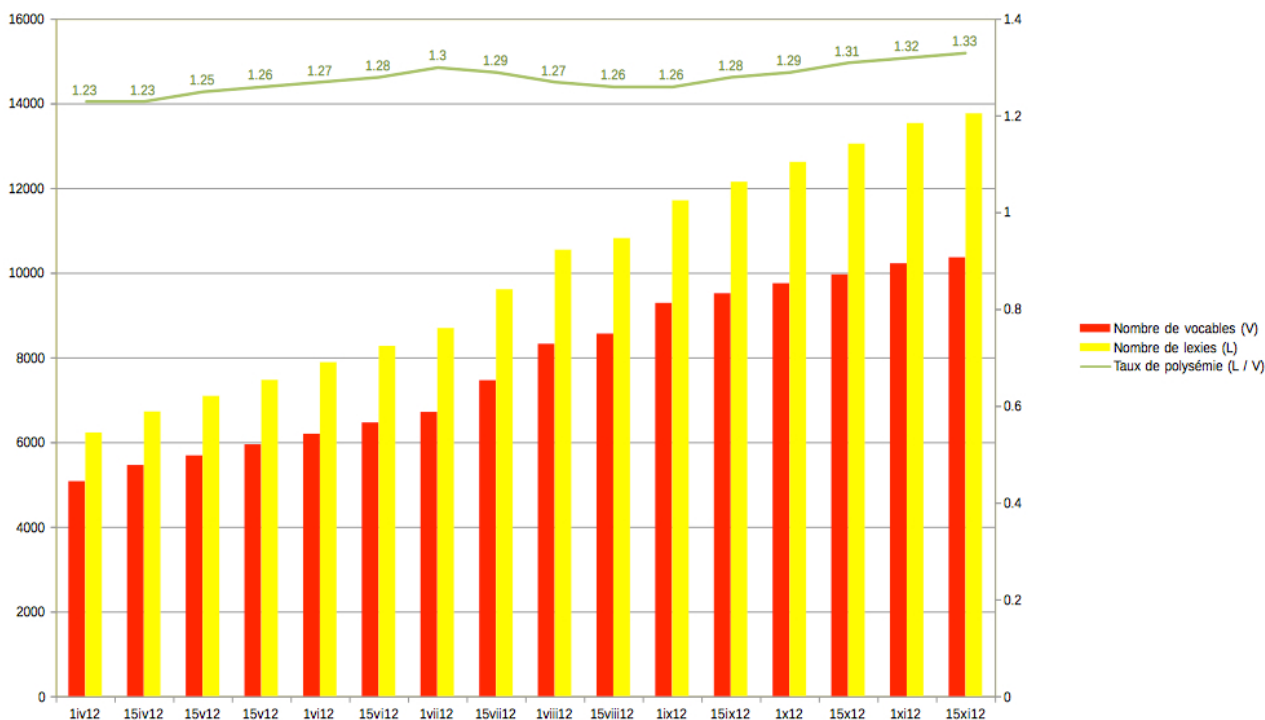


Figure 2 : Statistiques de l'expansion de la couverture lexicale du RLF

5. Conclusions

Le modèle d'identification et d'expansion de la nomenclature du RLF que nous venons d'explicitier nous semble particulièrement intéressant d'un point de vue pédagogique pour au moins deux raisons.

Tout d'abord, il s'inspire directement des réflexions menées par les spécialistes de l'enseignement du vocabulaire. Nous avons cherché à rationaliser et consolider la notion trop floue de « famille de mots », en lui substituant celle d'association lexicale fondée sur la dérivation sémantique proche. Nous avons de plus pris en compte plusieurs listes de vocables établies dans une perspective pédagogique pour synthétiser la NA du RLF. La couverture lexicale initiale du RLF résulte donc, en premier lieu, d'une réflexion pédagogique sur les vocabulaires.

Ensuite, le processus de croissance de la couverture lexicale du RLF pourrait, selon nous, inspirer les didacticiens intéressés à construire une méthode d'exploration et d'acquisition du lexique. En effet, ce processus présente certaines analogies avec la façon dont peut s'effectuer le développement « naturel » du vocabulaire d'un apprenant. La nomenclature du RLF croît de façon à la fois aléatoire – il n'y a pas de programmation stricte de l'ordre d'insertion de nouveaux vocables dans la base – et systématique – la croissance se fait en suivant, à l'intérieur de champs lexicaux bien identifiés (Polguère à paraître), les liens sémantiques et de combinatoire qui tissent le réseau lexical de la langue. Or, il ne fait aucun doute que la croissance du vocabulaire d'un individu est à la fois en grande partie aléatoire et en même temps conditionné par les liens forts qu'entretiennent les lexies de la langue.

Notons pour conclure que le RLF sera entre les mains de ses utilisateurs un outil pédagogique aux applications multiples. Deux aspects du potentiel pédagogique du RLF méritent d'être soulignés.

Premièrement, sa logique de structuration est transposable d'une langue à une autre, un grand nombre de fonctions lexicales ayant un caractère standard. Apprendre à naviguer dans le système lexical de sa langue maternelle, à comprendre, puis à prévoir les relations de sens prépare le locu-

teur à se lancer dans l'étude d'une L2, en lui fournissant les outils conceptuels nécessaires à cette navigation.

Deuxièmement, le RLF est destiné à devenir une ressource favorisant la structuration et la consolidation des connaissances lexicales et par là même le développement du vocabulaire actif. En effet, plus les liens rattachant une lexie à ses congénères sont nombreux, plus son ancrage dans le vocabulaire d'un individu est solide (Meara 2007). C'est cette approche qui avait déjà été utilisée dans le cadre de la rédaction du *Lexique actif du français* (Mel'čuk & Polguère 2007), ouvrage à finalité pédagogique qui décrit un échantillon du lexique français en termes des dérivations sémantiques et collocations. Le RLF est destiné à aller bien au-delà, non seulement en visant une couverture lexicale beaucoup plus vaste, mais aussi en s'affranchissant du support dictionnaire grâce à sa structure de graphe lexical¹⁹.

Bibliographie

Adolphs, S., Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, 24 (4), 425–438.

Aitchison, J. (2003). *Words in the Mind : An Introduction to the Mental Lexicon*. 3^e édition, Oxford UK : Blackwell.

Atkins, B. T. S. (1996). Bilingual Dictionaries : Past, Present and Future. In M. Gellerstam, J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström & C. R. Pappmehl (dir.) : *Euralex'96 Proceedings*, Göteborg : Université de Göteborg, 515–590.

Brunet, É. (1981). *Le Vocabulaire français de 1789 à nos jours d'après les données du « Trésor de la Langue Française »*, Genève/Paris : Slatkine/Champion.

Chetouani, L. (1997). *Vocabulaire général d'enseignement scientifique*. Paris : L'Harmattan.

Collectif Robert (2009). *Le Robert Benjamin*. Paris : Le Robert.

Cortier, C. & Bouchard, R. (dir.) (2008). « Quel oral enseigner, cinquante ans après le Français fondamental ? », *Le Français dans le monde. Recherches et applications*, 43, Cle-International.

Déro, M. (1998). *Inventaire du vocabulaire et développement des connaissances du Cours Préparatoire à la Sixième de Collège*. Thèse de doctorat, Université de Haute Bretagne – Rennes 2.

Gader, Nabil, Lux-Pogodalla, V. & Polguère, A. (2012). Hand-Crafting a Lexical Network With a Knowledge-Based Graph Editor. In *Proceedings of CogALex III*, Mumbai.

Galisson, R. (1971) *Inventaire thématique et syntagmatique du français fondamental*. Paris : Librairies Hachette et Larousse.

Gougenheim, G., Michéa, R., Rivenc, P. & Sauvageot, A. (1959). *L'élaboration du français fondamental, 1^{er} degré*. Paris : Didier.

Hausmann, F. J. (1989). Le dictionnaire de collocations. In F. J. Hausmann, O. Reichmann, H. E. Wiegand & L. Zgusta (dir.) : *Wörterbücher : ein internationales Handbuch zur Lexicographie / Dictionaries: An International Encyclopaedia of Lexicography / Dictionnaires : Encyclopédie internationale de lexicographie*. Handbücher zur Sprach- und Kommunikationswissenschaft, Band 5.1, Berlin/New York NY : Walter de Gruyter, 1010–1019.

Laufer, B. (1996). The lexical threshold of L2 reading : where it is and how it relates to L1 reading ability. In K. Sajavaara & C. Fairweather (dir.), « Approaches to Second Language Acquisition »,

19 Voir à ce propos l'approche de la lexicographie des dictionnaires virtuels, mentionnée dans la section 1.3 et présentée dans Polguère (2012b).

Jyvaskyla Cross Language Studies, 17, 55–62.

Lefrançois, P., Tremblay, O., Lombard, V. (2011). *Constitution de listes de mots pour l'apprentissage de l'orthographe et du lexique au primaire et au début du secondaire*. Rapport de recherche pour le Ministère de l'Éducation, du Loisir et du Sport du Québec (MELS), Université de Montréal, Montréal, 14 février 2011.

Lo Bianco, J., Scull, J. & Ives, D. (2008). *The Words Children Write 2008 : Research Summary of the Oxford Wordlist*. Melbourne : Oxford University Press.

Lux-Pogodalla, V., Polguère, A. (2011). Construction of a French Lexical Network. Methodological Issues. *Proceedings of the First International Workshop on Lexical Resources, WoLeR 2011. An ESSLLI 2011 Workshop*, Ljubljana, 54–61.

McCarthy, M. (1999) What constitutes a basic vocabulary for spoken communication ? *English Language and Literature*, 1, 233–249.

Meara, P. (2007). Towards a new approach to modelling vocabulary acquisition. In N. schmitt & M. McCarthy, Michael (dir.) : *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge UK : Cambridge University Press, 109-121.

Mel'čuk, I. (1994). *Cours de morphologie générale (théorique et descriptive) Vol. 2. Deuxième partie : Significations morphologiques*. Montréal & Paris : Les Presses de l'Université de Montréal & CNRS Éditions.

Mel'čuk, I. (1995). Phrasemes in Language and Phraseology in Linguistics. In M. Everaert. E.-J. van der Linden, A. Schenk & R. Schreuder (dir.) : *Idioms : Structural and Psychological Perspectives*, Hillsdale N.J./Hove UK : Laurence Erlbaum Associates, 167–232.

Mel'čuk, I. (1996). Lexical Functions : A Tool for the Description of Lexical Relations in the Lexicon. In L. Wanner (dir.) : *Lexical Functions in Lexicography and Natural Language Processing*, Language Companion Series 31, Amsterdam/Philadelphie : John Benjamins, 37–102.

Mel'čuk, I. (1997) *Vers une linguistique Sens-Texte*. Leçon inaugurale, Paris : Collège de France.

Mel'čuk, I. et coll. (1984, 1988, 1992, 1999). *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques I-IV*, Montréal : Les Presses de l'Université de Montréal.

Mel'čuk, I., Clas, A., Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve : Duculot.

Mel'čuk, I., Polguère, A. (2006). Dérivations sémantiques et collocations dans le DiCo/LAF. *Langue française*, 50, 66–83.

Mel'čuk, I., Polguère, A. (2007). *Lexique actif du français. L'apprentissage du vocabulaire fondé sur 20 000 dérivations sémantiques et collocations du français*. Collection « Champs linguistiques », Bruxelles : De Boeck & Larcier.

Milićević, J. (2006) A Short Guide to the Meaning-Text Linguistic Theory. *Journal of Koralex*, 8, 187–233.

Milićević, J. (2007). *La paraphrase. Modélisation de la paraphrase langagière*. Berne : Peter Lang.

Milton, J. (2009) *Measuring Second Language Acquisition*. Bristol UK/Buffalo NY/Toronto : Multilingual Matters.

Nation, I. S. P. (1982). Beginning to learn foreign language vocabulary : a review of research. *RELC Journal*, 13, 61–68.

- Odgen, C. K. (1930). *Basic English : A General Introduction with Rules and Grammar*. London : Paul Treber & Co.
- Picoche, J. (1995) L'Enseignement du vocabulaire en français langue maternelle au niveau de l'enseignement secondaire. In N. Andrieux-Reix & G. Hasenhor (dir.), *Études de lexicologie et dialectologie*, Paris : Conseil international de la langue française, 365–377.
- Picoche, J. (2007). *Enseigner le vocabulaire, la théorie et la pratique*. Champigny-sur-Marne : Éditions Allouche, publication électronique sur CD-ROM [version révisée de l'ouvrage paru en 1993 sous le titre *Didactique du vocabulaire français*, collection « fac.linguistique », Paris : Nathan].
- Polguère, A. (2007). Lexical function standardness. In L. Wanner (dir.) : *Selected Lexical and Grammatical Issues in the Meaning-Text Theory. In Honour of Igor Mel'čuk*, Language Companion Series 84, Amsterdam/Philadelphie : John Benjamins, 43–95.
- Polguère, A. (2008). *Lexicologie et sémantique lexicale. Notions fondamentales*. Collection « Paramètres », Montréal : Les Presses de l'Université de Montréal.
- Polguère, A. (2011). *Mémo RLF. Étapes dans la rédaction d'une entrée de vocable*. Projet RELIEF/RLF, Nancy : ATILF CNRS (dernière mise à jour : 20 février 2012).
- Polguère, A. (2012a). Lexicographie des dictionnaires virtuels. In Y. Apresjan, I. Boguslavsky, M.-C. L'Homme, L. Iomdin, J. Milićević , A. Polguère & L. Wanner, L. (dir.) : *Meanings, Texts, and Other Exciting Things. A Festschrift to Commemorate the 80th Anniversary of Professor Igor Alexandrovič Mel'čuk*, Studia Philologica, Moscou : Jazyki slavjanskoj kultury Publishers, 509–523.
- Polguère, A. (2012). Like a Lexicographer Weaving Her Lexical Network. In *Proceedings of CogA-Lex III*, Mumbai.
- Polguère, A. (à paraître). Les petits soucis ne poussent plus dans le champ lexical des sentiments. In F. Baider & G. Cislaru (dir.) : *Cartographie des émotions*, Villeneuve-d'Ascq : Presses Universitaires du Septentrion.