

## **Modélisation des liens lexicaux au moyen des fonctions lexicales**

Alain Polguère

OLST — Département de linguistique et de traduction  
Université de Montréal  
C. P. 6128, Succ. Centre-Ville  
Montréal (Québec) H3C 3J7  
Canada  
alain.polguere@umontreal.ca

### **Résumé — Abstract**

Ce tutoriel est une introduction à la modélisation lexicographique des liens lexicaux au moyen des fonctions lexicales de la théorie Sens-Texte. Il s'agit donc d'examiner un sous-ensemble des tâches effectuées en lexicographie formelle basée sur la lexicologie explicative et combinatoire. Plutôt que de viser l'introduction de toutes les fonctions lexicales identifiées par la théorie Sens-Texte, je vais m'attacher à introduire la notion de fonction lexicale de façon méthodique, en présentant d'abord les notions linguistiques plus générales sur lesquelles elle s'appuie (lexie, prédicat, actant, dérivation sémantique, collocation, etc.). Ce document vise essentiellement à récapituler les définitions des notions linguistiques qui vont être vues dans le tutoriel de façon pratique, par le biais d'exercices à caractère lexicographique.

This tutorial is an introduction to the modeling of lexical relations by means of Meaning-Text lexical functions. In other words, it presents a specific aspect of formal lexicography as practiced within the framework of explanatory combinatorial lexicology. Only a subset of Meaning-Text lexical functions will be introduced. My goal is to methodically presents the notion of lexical function itself, based on the introduction of more general linguistic notions it presupposes (lexie, actant, semantic derivation, collocation, etc.). The present document summarizes definitions of notions that will be applied in lexicographic exercises during the tutorial.

### **Mots clés — Keywords**

Fonction lexicale, lexicologie explicative et combinatoire, théorie Sens-Texte, lexicographie formelle.

Lexical function, explanatory combinatorial lexicology, Meaning-Text theory, formal lexicography.

## **1 Introduction**

L'identification et la modélisation des liens lexicaux sont des tâches centrales dans un grand nombre d'applications de la linguistique, par exemple : lexicographie, construction de lexiques formels pour le TALN et enseignement des connaissances lexicales (langue maternelle ou lan-

gue seconde). On peut appeler *lien lexical* deux types de connexions entre les unités lexicales de la langue :

1. Liens sémantiques très généraux de nature paradigmaticque appelés **DÉRIVATIONS SÉMANTIQUES**, par exemple
  - synonymie et quasi-synonymie — *manger* ~ *ingurgiter*, *comprimé* ~ *pilule* ;
  - antonymie et quasi-antonymie — *majorité* ~ *minorité*, *richesse* ~ *pauvreté* ;
  - nominalisation — *lourd* ~ *lourdeur*, *tomber* ~ *chute* ;
  - noms actanciels — *manger* ~ *mangeur* ~ *nourriture*, *livre* ~ *auteur* ~ *lecteur* ~ *contenu*.
2. affinités de combinatoire lexicale au sein d'expressions semi-idiomatiques appelées **COLLOCATIONS**, par exemple
  - intensificateurs types — *gros/profond soupir* ;
  - « atténuateurs » types — *faible/léger soupir* ;
  - verbes supports — *émettre/laisser échapper/pousser un soupir* ;
  - verbes causatifs — *arracher un soupir à qqn*.

Les **FONCTIONS LEXICALES**, l'objet de ce tutoriel, sont un outil notionnel extrêmement puissant pour comprendre, modéliser, encoder, apprendre et enseigner ces phénomènes. La notion de **FONCTION LEXICALE**, élaborée dans le cadre de la **THÉORIE SENS-TEXTE**, est maintenant un concept bien connu en lexicologie, lexicographie et TALN. Elle demeure cependant peu utilisée en regard de son intérêt pratique et théorique. Ce tutoriel vise à donner un meilleur accès à cette notion que ce qui peut être présentement obtenu par la seule étude des différentes publications sur le sujet. Il ne s'adresse pas aux linguistes déjà familiarisés avec la théorie Sens-Texte mais cherche au contraire à rendre la notion accessible à un public aussi large possible. De plus, il ne s'agit pas tant de présenter un formalisme que de mettre en évidence la nature profonde des fonctions lexicales et donc, de permettre leur utilisation dans des contextes aussi variés que la lexicographie, le TALN ou l'enseignement des langues.

La base théorique sur laquelle repose ce tutoriel, qui se veut surtout orienté vers la pratique lexicographique, est la **LEXICOLOGIE EXPLICATIVE ET COMBINATOIRE**, la branche lexicale de la théorie Sens-Texte. Je n'introduirai que les notions théoriques dont nous avons besoin ici. Les deux articles (Mel'čuk, 1997) et (Polguère, 1998) offrent une introduction générale à la théorie Sens-Texte, sous une forme relativement compacte. On trouvera une introduction détaillée à la lexicologie explicative et combinatoire dans (Mel'čuk *et al.*, 1995). Certains passages du présent document ont été extraits et adaptés de (Polguère, 2000a) et (Polguère, à paraître).

Le tutoriel est structuré en trois parties : définition de certaines notions linguistiques de base présumées par les fonctions lexicales, présentation de la notion de fonction lexicale proprement dite et travail pratique sur un sous-ensemble des fonctions lexicales dites « standard » identifiées par la théorie Sens-Texte.



L'approche de l'étude du lexique adoptée ici est purement synchronique : on ne va pas examiner comment modéliser les liens lexicaux du point de vue de l'évolution de la langue ; la langue est étudiée à un moment donné de son évolution. La langue de référence pour ce tutoriel est le français contemporain, tel que parlé en France. Mais je

m'appuierai aussi, bien entendu, sur des exemples empruntés à d'autres langues (notamment l'anglais) et à des variantes dialectales (français de France *vs* du Québec, anglais britannique *vs* américain).

## 2 Quelques notions de sémantique et de syntaxe nécessaires à la bonne maîtrise des fonctions lexicales

Il est nécessaire de définir, au besoin approximativement, plusieurs notions linguistiques de base avant de s'attaquer aux fonctions lexicales proprement dites. Les fonctions lexicales ne peuvent être comprises en profondeur et, donc, utilisées de façon concrète avec cohérence, sans que l'on ait défini au préalable la terminologie de base nécessaire à l'étude du lexique. La situation est exactement la même en syntaxe, par exemple, à la différence près que les notions de base de syntaxe formelle sont en général mieux connues que les notions bases utilisées en lexicologie et lexicographie formelles. Notez que je vais profiter de l'introduction des notions de base pour présenter quelques conventions d'écriture qui seront utilisées dans le tutoriel.

### 2.1 Lexie, vocable

On peut considérer que chaque langue est constituée de deux ensembles de connaissances distincts. Il y a tout d'abord le *LEXIQUE*, qui est un ensemble très vaste, à la structure relativement anarchique et qui peut être vu comme le répertoire des sens de base de la langue avec leur moyen d'expression associé. Il y a ensuite la *GRAMMAIRE*, qui est un ensemble de taille beaucoup plus modeste, comportant des règles générales permettant de combiner entre eux les éléments du lexique afin de produire des énoncés bien formés.

Ce tutoriel s'intéresse aux liens unissant les éléments du lexique entre eux et, éventuellement, aux liens unissant les éléments du lexique à des expressions « toutes faites » de la langue. Pour désigner les éléments du lexique, nous n'utiliserons pas le terme *mot*, qui est trop ambigu et ne relève pas d'une terminologie linguistique bien définie.

---

On appellera *UNITÉ LEXICALE*, ou *LEXIE*, un élément de base du lexique. Chaque lexie de la langue est associée à un sens particulier, unique, définissable et à un ensemble de formes correspondant aux variantes flexionnelles du signifiant de la lexie.

---

Je vais toujours écrire en petites majuscules les noms de lexies. Les sens seront donnés entre guillemets simples (dits *guillemets sémantiques*) et les signifiants en italique. Par exemple, on peut écrire que la lexie CHAT a pour sens 'chat' = 'animal domestique de petite taille...' et est associée aux signifiants *chat* (singulier de CHAT) et *chats* (pluriel).

Les mêmes signifiants sont très souvent utilisés dans la langue en association avec plus d'une seule lexie. Ils servent à exprimer des sens lexicaux liés de façon métaphorique, métonymique, etc. :

---

Un *VOCABLE* est un ensemble de lexies associées aux mêmes signifiants et dont les sens présentent une intersection non triviale.

---

Dans les dictionnaires, dans les travaux de recherche en lexicologie et lexicographie, etc., il est de coutume de numéroter les lexies d'un même vocable. Par exemple, le *Nouveau Petit Robert* décrit le vocable français PORC comme étant constitué de quatre lexies : PORC1 'animal domestique ...', PORC2 'individu sale ...', PORC3 'viande de porc1 ...' et PORC4 'peau du porc1 ...'.

Les lexies d'un vocable sont souvent appelées *ACCEPTIONS* de ce vocable. La *POLYSÉMIE* est la propriété d'un vocable donné telle qu'il contient plus d'une lexie. Ainsi, le vocable PORC est polysémique, comme nous venons de le voir, alors que le vocable PHACOCÈRE (pour rester dans le même domaine) est monosémique.

Il peut arriver que deux lexies distinctes soient associées aux mêmes signifiants alors qu'elles n'entretiennent aucune relation de sens ; il s'agit d'un cas d'*HOMONYMIE*. Les dictionnaires distinguent fréquemment les lexies homonymes par des numéros mis en exposant, et c'est aussi la notation adoptée en lexicologie explicative et combinatoire. Par exemple :

ADRESSE<sup>1</sup> [*Il fait preuve de beaucoup d'adresse.*]

~

ADRESSE<sup>2</sup> [*Est-ce que tu connais son adresse en France ?*]

L'approche adoptée pour le tutoriel étant synchronique (voir ci-dessus, page 38), nous devons nous baser sur la présence ou l'absence d'un lien sémantique en français (en anglais, etc.) contemporain pour décider si nous sommes en présence d'homonymes ou de lexies d'un même vocable. S'il existe un lien étymologique mais que ce lien n'est plus concrétisé par une relation de sens couramment perçue par les locuteurs, nous devons l'ignorer.



Il existe des techniques permettant d'identifier les acceptions d'un vocable dans un corpus. Nous allons les examiner en faisant quelques exercices pratiques d'analyse de corpus.

## 2.2 Prédicat sémantique, actant sémantique et syntaxique

On peut considérer que les sens lexicaux se divisent en deux grandes classes : prédicats sémantiques vs objets sémantiques.

---

Les *PRÉDICATS SÉMANTIQUES* sont des sens de lexies qui dénotent des faits ou des entités impliquant au moins un « participant » appelé *ACTANT SÉMANTIQUE*. Les actants sont habituellement désignés par des variables du type X, Y, Z, etc. — 'X mange Y', 'X donne Y à Z', '[X est] petit', 'amour de X pour Y', ...

---

On associe généralement les lexies prédictives à des faits (actions, événements, états, etc.). Il est cependant très important de garder à l'esprit qu'elles peuvent aussi dénoter des entités. Ainsi, la lexie NEZ [*Il a un tout petit nez.*], comme toutes les lexies dénotant des parties du corps, est un exemple de lexie prédictive dénotant une entité (et non un fait). Un nez est nécessairement le nez de quelqu'un et la mention de la personne à qui appartient le nez — *le nez de Cléopâtre* — permet d'exprimer l'actant sémantique de ce prédicat (un participant nécessaire de la situation « avoir un nez »).

---

Les **OBJETS SÉMANTIQUES** sont des sens de lexies qui dénotent des entités n'impliquant intrinsèquement aucun participant — 'tomate', 'sable', 'Émile Zola', ...

---

Les prédicats sémantiques types sont des verbes. En fait, un verbe est nécessairement un prédicat sémantique. Cependant, les adjectifs et les adverbes sont eux aussi des prédicats. Des lexies comme GROGNON ou MÉCHAMMENT signifient nécessairement '[quelqu'un est] grognon' et '[quelque chose est fait] méchamment'.

On voit donc que le fait d'être un prédicat est une propriété sémantique de ces trois parties du discours. Cependant, les noms peuvent aussi être des prédicats ; ils le sont même très fréquemment. Ainsi, lorsqu'on emploie la lexie AMOUR, on sous-entend, voir (1a) ci-dessous, ou on exprime explicitement, voir (1b), deux actants de ce prédicat : celui qui éprouve de l'amour et la personne pour laquelle ce sentiment est éprouvé :

- (1) a. *C'est un amour platonique.*  
b. *L'amour de Léonce pour Justine est platonique.*

Même si ni Léonce ni Justine ne sont mentionnés dans (1a), on sait que deux « participants » sont nécessairement impliqués dans la situation dont il est question. Cette information nous est communiquée par le sens même du lexème AMOUR, dont une caractéristique est d'être un prédicat à deux actants. On peut ainsi contraster 'amour' avec un autre prédicat nominal, à quatre actants celui-là :

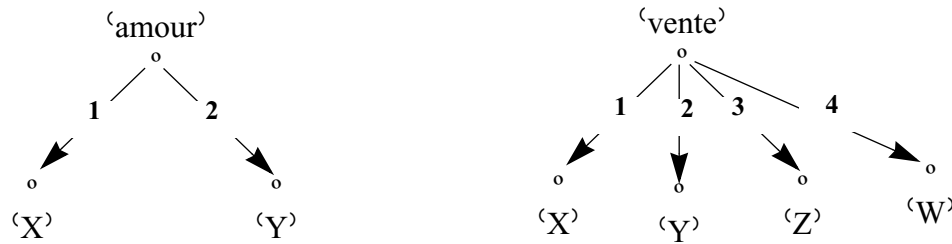
- (2) a. *La vente a eu lieu dans la matinée.*  
b. *La vente par l'agence[= X] de la maison[= Y] au client impatient[= Z] pour une somme astronomique[= W] a eu lieu dans la matinée.*



Il est impossible de décrire correctement une lexie sans considérer sa nature de prédicat sémantique. Cela concerne tous les aspects de la caractérisation de la lexie : sens, comportement dans la phrase, etc. Notamment, il est impossible de bien modéliser les relations lexicales sans avoir effectué un minimum d'analyse sémantique des lexies concernées. L'outil des fonctions lexicales nous force à effectuer cette analyse sémantique lexicale, tout en nous aidant, de façon parallèle, à mener à bien cette analyse.

Cet aspect crucial du fonctionnement et de l'utilisation des fonctions lexicales va peu à peu s'éclaircir, au fur et à mesure que nous progresserons dans le tutoriel.

Notons finalement que, dans la théorie Sens-Texte, les liens qui unissent un prédicat à ses actants sont simplement distingués par une numérotation, en chiffres arabes. On n'utilise pas de marqueurs de rôle sémantique (du type « agent », « patient », etc.) dans la mesure où la relation entre le prédicat et ses actants est modélisée d'une autre façon, beaucoup plus complète : dans la définition lexicographique (voir Section 2.5 ci-dessous), qui explicite les liens simplement désignés dans le reste de l'article de dictionnaire au moyen de chiffres. Les structures sémantiques associées aux lexies des exemples (1a-b) et (2a-b) ci-dessus sont donc :



### 2.3 Diathèse d'une lexie

---

La théorie Sens-Texte appelle *ACTANT SYNTAXIQUES* d'une lexie les éléments de la phrase dont l'emploi est régi par la lexie. La *DIATHÈSE* d'une lexie, quant à elle, est la correspondance entre les actants sémantiques et les actants syntaxiques de cette lexie.

---

Ceci est une caractérisation extrêmement simplifiée des notions d'actant syntaxique et de diathèse. Je ne vais pas entrer dans le détail de la présentation de ces notions. Elles relèvent d'un cours de syntaxe et l'approche syntaxique Sens-Texte, qui s'appuie sur la *SYNTAXE DE DÉPENDANCE* telle que théorisée dans (Tesnière, 1959), a été introduite l'année dernière dans un tutoriel de cette même conférence (Kahane, 2001). Je vais me contenter de donner ici quelques exemples.

La diathèse d'une lexie peut se représenter au moyen de tableaux, appelés *TABLEAUX DE RÉGIME*, montrant explicitement la correspondance entre les actants sémantiques, les actants syntaxiques dits « profonds » — numérotés au moyen de chiffres romains — et leur réalisation dans la phrase à un niveau syntaxique de surface. La numérotation des actants syntaxiques reflète la hiérarchie relative des structures régies. Ainsi, l'actant syntaxique d'un verbe qui correspond au sujet est numéroté **I**, le premier complément d'objet (direct ou indirect) est numéroté **II**, etc.

Voici les tableaux de régime des lexies *AIMER* et *VENTE* :

X = <b>I</b>	Y = <b>II</b>
1. N	1. N

Figure 1 : Régime de *AIMER*

X = <b>I</b>	Y = <b>II</b>	Z = <b>III</b>	W = <b>IV</b>
1. <i>de</i> N	1. <i>de</i> N	1. <i>à</i> N	1. <i>pour</i> N
2. <i>par</i> N	2. $A_{\text{poss}}$		2. <i>d'un montant de</i> N
3. $A_{\text{poss}}$			

Figure 2 : Régime de *VENTE*

Il existe bien entendu plusieurs lexies AIMER et plusieurs lexies VENTE ; en d'autres termes, les vocables AIMER et VENTE sont polysémiques. Néanmoins, je ne distinguerai pas les lexies au moyen d'une numérotation lorsque cela ne sera pas pertinent pour la discussion.

La correspondance entre actants sémantiques et actants syntaxiques n'est pas toujours aussi directe que dans les deux exemples qui viennent d'être donnés. Notamment, une même lexie peut avoir plus d'un régime, comme dans le cas suivant :

Régime 1			Régime 2	
X = I	Y = II		X+Y = I	
1. <i>de</i> N	1. <i>envers</i> N		1. <i>de</i> N <i>et</i> N	
2. A <sub>poss</sub>	2. <i>pour</i> N		2. <i>entre</i> N <i>et</i> N	
			3. A <sub>poss</sub>	

Figure 3 : Régimes de AMOUR

## 2.4 Paraphrase et sens

Deux expressions linguistiques ayant (approximativement) le même sens sont appelées des *PARAPHRASES*. Il n'y a pratiquement pas d'autre façon naturelle de procéder pour décrire le sens que de faire appel à des paraphrases. C'est d'ailleurs ainsi que fonctionnent les définitions de la plupart des dictionnaires. Cela nous amène à définir le sens linguistique de la façon suivante :

Le *SENS* d'une expression linguistique est la seule propriété qu'elle partage avec toutes ses paraphrases.

Cette définition pourrait paraître circulaire dans la mesure où elle revient *grosso modo* à dire que le sens d'une expression linguistique est la seule propriété que partage cette expression avec toutes les autres expressions ayant le même sens. Cependant, cette circularité n'est qu'apparente : avoir le même sens (ou être une paraphrase) est, comme être grammaticalement correct, un concept primitif (c'est-à-dire, indéfinissable) de la langue. En tant que locuteurs du français, nous pouvons immédiatement dire si les trois phrases françaises (3a-c) ci-dessous sont des paraphrases — si elles ont le même sens —, et cela sans avoir besoin de suivre un cours de linguistique :

- (3) a. *Je pense donc je suis.*  
 b. *Le fait que je pense démontre que j'existe.*  
 c. *Ma pensée est la preuve de mon existence.*

Une des caractéristiques de la langue, qui l'oppose à bien des systèmes sémiotiques « artificiels » comme la logique formelle, les langages de programmation, etc., est d'offrir à la personne qui l'emploie un très grand nombre d'options plus ou moins équivalentes pour exprimer un contenu donné. La relation de paraphrase est en quelque sorte une donnée première du

sens, quelque chose que nous n'allons pas définir mais que nous prendrons comme un concept primitif permettant de définir la notion de sens elle-même.

Le lien de paraphrase est cependant de nature très complexe, même si sa perception par le locuteur se fait de façon instantanée. Il faut notamment distinguer la paraphrase exacte, qui est finalement assez difficile à obtenir, et différentes variétés de paraphrases approximatives. Le lien de paraphrase renvoie en fait à un continuum ; c'est ce qu'illustrent les exemples suivants, où se manifeste un écart sémantique croissant par rapport à la phrase de référence (4a) :

- (4) a. *Cette pierre est très lourde.*  
 b. *Cette pierre pèse très lourd.*  
 c. *Le poids de cette pierre est élevé.*  
 d. *Cette pierre est difficile à transporter.*

La phrase (4b) peut être considérée comme une paraphrase exacte de (4a), le remplacement de ÊTRE par PESER n'introduisant aucune nuance de sens. Par comparaison, (4c) s'écarte légèrement du sens initial, notamment parce que la structure de la seconde phrase, où POIDS est le sujet grammatical et ÉLEVÉ attribut, diminue l'importance de PIERRE, qui était sujet dans la première phrase : (4a) dit quelque chose à propos d'une pierre alors que (4c) dit quelque chose à propos du poids de cette pierre. Bien entendu, il est clair que ce sont les mêmes informations qui sont communiquées, et donc que l'on a bien affaire à deux paraphrases. Mais une nuance existe tout de même au niveau de la façon dont l'information communiquée est « emballée » dans la phrase. On dira que (4a) et (4c) n'ont pas la même *STRUCTURE COMMUNICATIVE*. La notion de structure communicative est étudiée en détail dans le cadre de la théorie Sens-Texte dans (Mel'čuk, 2001). Finalement, on remarque que la phrase (4d) n'est pas déjà plus une paraphrase linguistique véritable de (4a) ; on pourrait la considérer comme étant une paraphrase « conceptuelle ».

Si j'ai introduit ici la notion de paraphrase c'est d'une part parce qu'elle est au cœur de la modélisation du sens et, d'autre part, parce que les fonctions lexicales sont, justement, des outils descriptifs permettant de modéliser de façon élégante les choix paraphrastiques qui s'offrent au locuteur lorsqu'il veut exprimer un message donné.

## 2.5 Définition lexicale par genre prochain et différences spécifiques

Les sens lexicaux sont décrits en lexicologie explicative et combinatoire au moyen de définitions lexicographiques présentant les trois caractéristiques suivantes :

1. elles proposent une paraphrase du sens de la lexie définie ;
2. elles sont constituées de sens lexicaux plus simples que celui de la lexie définie ;
3. elles ont une composante sémantique centrale appelée *GENRE PROCHAIN* — qui est en quelque sorte une paraphrase approximative minimale du sens de la lexie définie — et des composantes sémantiques périphériques appelées *DIFFÉRENCES SPÉCIFIQUES* — qui correspondent aux composantes sémantiques spécifiques au sens de la lexie définie.



La dernière caractéristique mentionnée ci-dessus donne son nom au type de définition en question : la **DÉFINITION PAR GENRE PROCHAIN ET DIFFÉRENCES SPÉCIFIQUES**, que l'on peut aussi appeler **DÉFINITION ANALYTIQUE**. Cette terminologie est très ancienne puisqu'elle provient du Livre V de l'*Organon* d'Aristote : *Les Topiques*.

La définition (5a) ci-dessous est une définition analytique telle qu'on pourrait la trouver dans un dictionnaire comme le *Nouveau Petit Robert*. Par contraste, (5b) est, plutôt qu'une définition, une simple énumération de synonymes approximatifs :

- (5) a. LABEUR travail<sub>[genre prochain]</sub> long et pénible<sub>[diff. spécifiques]</sub>  
b. LABEUR travail, corvée, besogne.

La plupart des dictionnaires de langue proposent des définitions analytiques, mais ils ne le font pas toujours de façon systématique. Dans la pratique, il faut être capable d'ébaucher des définitions lexicales analytiques pour bien modéliser les relations de fonction lexicale, qui, comme nous le verrons bientôt, sont toutes ancrées dans le contenu sémantique de la lexie.



Nous allons faire rapidement une ébauche de définition lexicale analytique en tentant de procéder par étapes logiques : 1) identification de la nature prédicative de la lexie, 2) identification du genre prochain, 3) caractérisation sémantique des actants, 4) identification des différences spécifiques et 5) validation de la définition.

## 2.6 Idiomaticité

En théorie, les énoncés obéissent au principe dit de **COMPOSITIONALITÉ SÉMANTIQUE** voulant que le sens d'un énoncé soit la résultante de la composition du sens des éléments qui le constituent. Ce principe rend compte du fait que le sens d'une phrase comme

- (6) *Les invités sont arrivés.*

est directement calculable à partir du sens de chacune des lexies qui le composent. Pour pouvoir calculer le sens de la phrase (6), il suffit de connaître :

- le sens des lexies LE, INVITÉ, ARRIVER ;
- les sens grammaticaux de pluriel et de passé ;
- les règles grammaticales du français.

Les énoncés sont cependant truffés de cas qui contredisent ce principe général et qui manifestent d'une façon ou d'une autre une certaine **IDIOMATICITÉ**. L'idiomaticité, qui est une transgression du principe de compositionnalité sémantique, est un phénomène graduel, comme l'illustre la série d'exemples ci-dessous, où les expressions qui nous intéressent sont en gras :

- (7) *Sylvain **a levé le rocher** avec un treuil.*

Expression respectant entièrement le principe de compositionnalité sémantique. C'est une **EXPRESSION LIBRE**.

- (8) *Le chien a levé la patte en passant près du réverbère.*

Expression quasi figée signifiant ‘uriner en levant la patte’. Elle est modélisée en tant que lexie dans la théorie Sens-Texte sous le nom de *QUASI-PHRASÈME*. Dire qu’une telle expression est une lexie implique qu’elle doit posséder sa propre entrée dans le dictionnaire de la langue.

- (9) *Sylvain a levé la séance.*

Expression linguistique semi-idiomatique non lexicalisée, appelée collocation ou semi-phrasème. Les collocations sont présentées plus en détail dans la Section 3.1.2 ci-dessous.

- (10) *Sylvain travaille trop. Il devrait lever le pied un peu.*

C’est une *EXPRESSION FIGÉE*, c’est-à-dire entièrement non compositionnelle, signifiant ‘ralentir, en faire moins’. Les expressions figées sont des lexies de la langue et doivent être décrites comme telles. Elles sont appelées *LOCUTION* dans la terminologie traditionnelle (locution verbale, nominale, etc.) et *PHRASÈME COMPLET* dans le cadre de la théorie Sens-Texte. Noter qu’une lexie qui n’est pas phrasémique, c’est-à-dire qui est formellement constituée d’un seul élément lexical (CHAT, ADRESSE<sup>1</sup>, ADRESSE<sup>2</sup>, AIMER, etc.) est appelée *LEXÈME*. Lorsque l’on parle de lexie, on a souvent tendance à ne penser, par défaut, qu’aux lexèmes. Cependant, les phrasèmes sont en nombre considérable dans toute langue naturelle. Leur description, dans les dictionnaires de langue courants, est malheureusement souvent enchâssée dans l’entrée d’un des éléments lexicaux les constituant.



Les noms de lexies qui sont des quasi-phrasèmes ou des phrasèmes complets (locutions) seront écrits en petites majuscules (comme pour toutes les lexies), encadrés par des coins relevés : « LEVER LA PATTE » et « LEVER LE PIED ». Noter qu’il serait incorrect d’écrire \*« LEVER LA SÉANCE » puisque une telle lexie n’existe pas : seule existe l’expression semi-idiomatique — la collocation — *lever la séance*. On trouvera dans (Mel’čuk, 1995) une présentation détaillée des différents types de phrasèmes en langue.

Nous en avons terminé avec la présentation des notions linguistiques dont la maîtrise est pré-supposée par la notion de fonction lexicale. Nous pouvons maintenant entrer dans le vif du sujet.

### 3 La notion de fonction lexicale

Pour présenter les *FONCTIONS LEXICALES* (dorénavant *FL*), je vais commencer par introduire plus précisément les deux types de phénomènes qu’elles permettent de modéliser : les dérivations sémantiques et les collocations (dont la présentation vient d’être ébauchée). Nous passerons ensuite à la définition précise de la notion de FL, puis examinerons le langage formel d’encodage des liens de FL.

#### 3.1 Les liens lexicaux : dérivation sémantique vs lien collocationnel

##### 3.1.1 Les dérivations sémantiques

Les dérivations sémantiques sont des *LIENS PARADIGMATIQUES* entre lexies :

---

La lexie  $L_2$  est un **DÉRIVÉ SÉMANTIQUE** de la lexie  $L_1$  si  $L_1$  et  $L_2$  sont liées par une relation sémantique récurrente en langue.

---

On peut identifier trois grandes classes de dérivations sémantiques :

1. La relation sémantique entre  $L_1$  et  $L_2$  peut être l'égalité exacte ou approximative, ce qui est un cas un peu extrême de dérivation sémantique. Il s'agit ici notamment des synonymes ou, plus fréquemment, des quasi-synonymes : COMPLIMENT ~ ÉLOGE. Entrent aussi dans cette classe les conversifs — DONNER ~ RECEVOIR —, les nominalisations — REGRETTER ~ REGRET —, verbalisations, adjectivisations et adverbialisations, ainsi que les termes génériques — SENTIMENT pour COLÈRE.
2.  $L_1$  et  $L_2$  peuvent avoir des sens opposés : ce sont des (quasi-)antonymes — GRAND ~ PETIT.
3. Finalement,  $L_2$  peut dénoter un élément de la situation dénotée par  $L_1$  : un participant — VENDEUR pour VENDRE —, un circonstant — ARME [du crime] pour CRIME —, ou une caractéristique d'un participant/circonstant — SOUPÇONNEUX [= 'qui tend à soupçonner'] pour SOUPÇONNER.

On doit employer ici le terme spécifique de *dérivation sémantique* pour éviter la confusion avec l'interprétation courante du terme *dérivation*, qui implique généralement une relation morphologique entre lexies. La dérivation sémantique couvre un éventail beaucoup plus large de relations paradigmatiques entre lexies puisque  $L_1$  et  $L_2$  ne doivent pas nécessairement être liées morphologiquement pour qu'il existe entre elles un lien de dérivation sémantique (voir, par exemple, TOMBER et sa nominalisation « supplétive » CHUTE).

### 3.1.2 Les liens collocationnels

La notion de collocation a été étudiée dans un des tutoriels de TALN 2001 (Daille, 2001) et (Williams, 2001). Il importe de noter que je ne considère ici que la notion de collocation telle que définie dans le cadre de la théorie Sens-Texte. En effet, on trouve dans la littérature le terme *collocation* utilisé pour dénoter au moins deux notions assez distinctes.

1. Tout d'abord, la collocation est fréquemment définie sur une base statistique : une combinaison de mots dont la fréquence d'occurrence dans les corpus est significativement élevée. Selon cette définition, qui tire son origine d'une interprétation strictement quantitative de la notion proposée par (Firth, 1957), une expression telle que *manger un gâteau* **pourrait** être considérée comme étant un meilleur candidat au statut de collocation que *manger un coup*, si les mesures statistiques démontrent que, proportionnellement, *manger* tend à apparaître plus fréquemment dans les corpus de référence avec *gâteau* qu'avec *coup*. Or, dans le cadre de la lexicologie explicative et combinatoire, les mesures statistiques peuvent aider à repérer les collocations mais elles ne sont pas définitives. Sans qu'il soit nécessaire de faire de calculs statistiques, on dira que *manger un gâteau* est une

expression libre alors que *manger un coup* est une collocation mettant en jeu la FL  
**Oper<sub>2</sub>** :

**Oper<sub>2</sub>(coup)** = recevoir, fam encaisser, fam manger [ART ~].

2. Dans le cadre de la lexicologie explicative et combinatoire, la définition de la notion de collocation ne fait donc pas référence à la fréquence d'apparition d'une expression mais 1) à la façon dont elle est « fabriquée » par le locuteur et 2) à sa nature semi-idiomatique :

---

L'expression *AB* (ou *BA*), formée des lexies A et B, est une **COLLOCATION** si la lexie A est sélectionnée librement — strictement d'après son sens <sup>(A)</sup> —, alors que la lexie B est sélectionnée pour exprimer un sens <sup>(C)</sup> en fonction de A. Dans un tel cas, on dira qu'il existe un lien collocationnel allant de A à B.

---

Voici quelques exemples de collocations en français :

- (11) a. *grosse*<sub>[= B]</sub> *tempête*<sub>[= A]</sub>  
 b. *dormir*<sub>[= A]</sub> *profondément*<sub>[= B]</sub>  
 c. *pleuvoir*<sub>[= A]</sub> *des cordes*<sub>[= B]</sub>

Une collocation est contrôlée par l'élément qui retient son sens dans la collocation et qui est choisi librement : il s'agit de la **BASE** de la collocation. Dans les exemples ci-dessus, les bases des collocations sont les éléments étiquetés A. Les éléments étiquetés B sont appelés **COLLOCATIFS**. Pour mémoire, notons que les trois collocations ci-dessus s'analysent au moyen de la FL **Magn**.

On dit que la base contrôle la collocation car, du point de vue du locuteur, c'est le collocatif qui est choisi en fonction de la base, et non l'inverse. Si j'entre dans une pièce mal éclairée et que je veux signaler ce fait, je peux dire, par exemple, (12a) ou (12b) :

- (12) a. *Il fait noir ici.*  
 b. *Il fait sombre ici.*

Cependant, si je veux revenir sur ce que je viens de dire, pour le modifier et insister sur le fait qu'on n'y voit vraiment rien, mon choix initial va contraindre la façon dont je peux maintenant m'exprimer :

- (13) a. *Il fait même noir **comme dans un four**.*  
 b. *Il fait même **vraiment très** sombre.*  
 c. *\*Il fait même sombre **comme dans un four**.*

Il n'y a aucune raison **logique** pour que (13a) soit correcte et non (13c), si ce n'est le fait que la locution 「FAIRE NOIR」 admet *comme dans un four* comme collocatif, ce que n'admet pas la locution 「FAIRE SOMBRE」.

Les collocations, en tant que phénomène linguistique, possèdent les trois caractéristiques suivantes :

1. Elles sont universellement présentes dans toutes les langues.
2. Elles sont omniprésentes dans les textes, qu'ils soient oraux ou écrits.
3. Elles semblent plus ou moins arbitraires, ne peuvent pas se traduire mot à mot d'une langue à l'autre et sont donc très difficiles à acquérir.

Toutes ces caractéristiques montrent qu'il est primordial de disposer d'un mécanisme qui nous permette non seulement de décrire les collocations de façon rigoureuse, mais aussi de faire des prédictions à leur propos. C'est ici que les FL vont intervenir.

Notons finalement que les collocations sont généralement assez mal traitées dans les dictionnaires de langue. Elles ne sont pas décrites en tant que telles ; seules quelques unes, parmi les collocations les plus « courantes », apparaissent mentionnées dans les articles de dictionnaire des lexies qui en sont la base. Il existe bien entendu des dictionnaires spécialement dédiés à la présentation des collocations, par exemple : (Beauchesne, 2001) pour le français et (Benson *et al.*, 1997) pour l'anglais. Il existe aussi des dictionnaires dits « de locutions » ou « d'expressions idiomatiques », qui contiennent en général les descriptions de nombreuses collocations, mélangées aux descriptions de véritables locutions ; voir, par exemple, (Rey, Chantreau, 1997), pour un ouvrage de référence, et (Kirk-Greene, 1997), pour un outil pédagogique destiné aux anglophones apprenant le français.

### 3.2 Généralisation des liens lexicaux : définition de la notion de FL

Les FL servent à modéliser les liens de dérivation sémantique et les liens collocationnels. On peut procéder en deux étapes pour définir la notion de FL : d'abord définir ce qu'est une FL au sens large, puis définir la FL type, appelée **FL STANDARD**.



La présentation ci-dessous, sous forme de définitions commentées, ne fait qu'énumérer les faits « théoriques ». Il nous faut voir ensemble des données concrètes pour comprendre l'intérêt véritable de ce qui est présenté de façon relativement sèche ici. On trouvera une présentation détaillée de la notion de FL — définition, exemples, implications pratiques, etc. — dans (Mel'čuk, 1996). Une description succincte est faite dans (Mel'čuk, 1997) et (Polguère, 1998).

#### 3.2.1 La notion de FL au sens large

La notion de FL est relativement complexe et sa définition l'est tout autant. Nous allons étudier une définition structurée en composantes qui sont clairement identifiées par un nom (*Fonctionnement*, *Statut linguistique*, etc.) et dont la pertinence sera examinée au fur et à mesure de leur introduction (points énumérés après les «  $\Rightarrow$  » ci-dessous).

---

Une fonction lexicale  $f$  est une correspondance (sur le modèle des fonctions mathématiques) ayant les propriétés suivantes :

Fonctionnement Au sein d'une langue donnée  $\mathcal{L}$ , elle associe à toute unité lexicale  $L$  à laquelle elle peut logiquement s'appliquer — appelée *MOT-CLÉ* de  $f$  — un ensemble relativement petit (et éventuellement vide) d'expressions lexicales de  $\mathcal{L}$  appelé *VALEUR* de l'application de  $f$  à  $L$  :  $f(L) = \{L_{\text{valeur}}\}$ .

- ⇒ Connexions à l'intérieur d'une langue donnée, et non d'une langue à l'autre.
- ⇒ Valeur non vide anticipée pour un grand nombre de lexies pour une FL « normale ».
- ⇒  $f$  sélectionne un sous-ensemble du lexique auquel elle peut s'appliquer.

Statut linguistique L'application  $f(L)$  se caractérise en fonction de son contenu, qui est soit une relation sémantique liant  $f(L)$  à  $L$  soit un rôle sémantico-syntaxique que  $f(L)$  peut jouer relativement à  $L$  au niveau syntaxique profond.

- ⇒ *FL PARADIGMATIQUE* pour un contenu qui est une dérivation sémantique.
- ⇒ *FL SYNTAGMATIQUE* pour un contenu qui est un lien collocationnel.
- ⇒ Le contenu sémantique d'une FL s'applique toujours à une composante donnée de la définition du mot-clé, y compris pour les FL syntagmatiques.

Condition générale Pour toute paire  $f(L_1)$  et  $f(L_2)$ , tous les éléments de  $f(L_1)$  et tous les éléments de  $f(L_2)$  sont liés de façon identique à  $L_1$  et  $L_2$  respectivement, ce lien correspondant au contenu de  $f(L)$  tel que décrit dans le Statut linguistique.

- ⇒ Les FL permettent d'établir des généralisations et de modéliser la paraphrase, voir (Mel'čuk, 1988).

Condition de validité Dans une langue donnée  $\mathcal{L}$ , il existe des cas où deux applications  $f(L_1)$  et  $f(L_2)$  retournent deux ensembles non vides distincts et une telle différence ne peut pas être anticipée.

- ⇒ Seuls les liens présentant un caractère irrégulier correspondent à une FL véritable.
- 

Cette définition ne fait en réalité que préparer le terrain pour la définition de la notion de FL « véritable », c'est à dire la FL standard. En fait, c'est uniquement lorsque l'on a à notre disposition un ensemble, même limité, de FL standard que l'on peut bien saisir toute la valeur de la notion de FL au sens large.

### 3.2.2 *FL standard vs FL non-standard*

Les FL se subdivisent, en fonction de leur degré de généralité, en FL standard vs FL non-standard.

---

Une FL **f** est une *FL STANDARD* si elle possède les propriétés suivantes :

Statut linguistique L'application **f**(L) est un cas particulier de « lexème généralisé » qui n'apparaît qu'au niveau de représentation syntaxique profond des énoncés. **f** est une *FL STANDARD SIMPLE* si c'est un élément de base du langage descriptif formel des FL; autrement, il s'agit d'une *FL COMPLEXE*.

⇒ Les FL sont des entités abstraites, des sortes de métalexies ; par exemple, **Magn**(*dormir*) vaut pour toutes les expressions qui permettent d'exprimer l'intensification auprès de la lexie DORMIR.

⇒ La distinction entre niveau syntaxique profond et de surface est nécessaire pour bien comprendre le fonctionnement des FL.

Couverture Il existe un grand nombre de lexies L pour lesquelles l'application **f**(L) peut être logiquement considérée. De plus, il existe un grand nombre d'applications logiquement possibles **f**(L) qui retournent une valeur non vide.

⇒ Sélection d'un nombre important de mots-clés potentiels.

⇒ Au moins une centaine d'applications **f**(L) dans chaque langue retournent une valeur non vide.

⇒ Cette condition est satisfaite du fait du contenu sémantique très vague de toute FL standard. Elle est de plus liée directement au caractère **universel** des FL standard.

Diversité Il existe une assez grande diversité parmi toutes les valeurs retournées pour **f**(L) dans chaque langue donnée **L**.

⇒ Pour chaque langue donnée **L**, on doit trouver au moins une douzaine de valeur distinctes de **f**(L).

Autrement, **f** est considérée comme étant une *FONCTION LEXICALE NON-STANDARD* du fait d'un des échecs suivants

Échec partiel La condition de diversité n'est pas remplie.

Échec complet La condition de couverture n'est pas remplie, ce qui implique que la condition de diversité ne l'est pas non plus. Si la condition de couverture n'est pas remplie au point que **f** ne peut s'appliquer qu'à une seule lexie (ou à deux ou trois lexies sémantiquement très proches), **f** est appelée une *PSEUDO-FL*.



Tel que mentionné plus haut, la notion de FL est essentiellement orientée vers celle de FL standard. Une FL non-standard peut être vue comme une FL qui « n'a pas réussi » à être une FL standard ; notamment, elle n'a pas le caractère universel d'une FL standard.

### 3.3 Le langage formel descriptif des FL

Je ne fais ici qu'énumérer les principales caractéristiques du langage formel des FL, qui sera vu de façon pratique au cours du tutoriel. On pourra notamment examiner les encodages présents dans l'extrait d'article lexicographique pour CHAT, donné plus bas (Section 4.2). Pour une dis-

cussion des problèmes posés par ce langage et pour des alternatives, voir (Kahane, Polguère, 2001).

### 3.3.1 Les noms de FL standard simples

Les noms de FL standard simples sont normalement des formes abrégées de mots latins (**Magn** pour *magnus*, **Bon** pour *bonus*, etc.), ceci afin de refléter le caractère universel de ces FL.

### 3.3.2 Les noms de FL non-standard

Les noms de FL non-standard sont des paraphrases en langue naturelle du lien encodé. La langue utilisée étant bien entendu normalement celle dont on décrit le lexique. Par exemple, on trouve dans l'article de CHAT (Section 4.2) les jolies dérivations sémantiques suivantes :

**Excrément du C.** : *crotte* [de ~]; **fam pipi, fam pisse** [de ~]  
**Gravier spécial sur lequel le C. fait ses besoins** : *litière*

Une FL peut être semi-standard. Dans ce cas, la formule d'encodage est hybride, incorporant une partie formalisée et une partie qui est une paraphrase en langue naturelle de la composante non-standard du lien de FL. Par exemple :

**essayer de**  
**Oper<sub>2</sub>(compliment)** = *rechercher* [ART ~] | C. surtout au pl



Comme on peut le voir dans les exemples ci-dessus, il est souvent nécessaire de donner dans la description de la valeur de **f(L)** non seulement les lexies liées elles-mêmes, mais aussi les structures syntaxiques qu'elles contrôlent. Il s'agit du **RÉGIME** de chaque élément de la valeur. Celui-ci est indiqué entre crochets.

### 3.3.3 Les FL complexes

Chaque FL doit être décrite en fonction de son contenu, bien entendu, mais aussi en fonction de sa combinatoire avec d'autres FL. Les FL standard simples se combinent en effet naturellement pour former ce que l'on appelle des **FL COMPLEXES**. Par exemple, dans l'article de ABANDON1.2b (état d'un lieu) :

**Oper<sub>1</sub>** : *être* [à l'~]/[dans ART ~ | avec modificateur]  
**PermOper<sub>1</sub>** : *laisser* [N<sub>Y</sub> à l'~]  
**IncepOper<sub>1</sub>** : *tomber* [dans ART ~]  
**PermIncepOper<sub>1</sub>** : *laisser aller* [N<sub>Y</sub> à l'~]

### 3.3.4 Les configurations de FL

Il peut arriver qu'un lien de FL soit basé non sur la combinaison de plusieurs FL formant une FL complexe, comme ci-dessus, mais sur l'application simultanée et de façon parallèle de plusieurs FL au même mot-clé. Par exemple :



**Magn<sup>quant</sup>+Oper<sub>12</sub>(compliment)** = *se répandre [en ~s auprès de N<sub>Y</sub>], prodiguer [ART ~s à N<sub>Y</sub>]*

indique que les collocations en question signifient **à la fois** « faire » des compliments (la FL **Oper<sub>12</sub>**) et que ces compliments sont nombreux (la FL **Magn<sup>quant</sup>**).

### 3.3.5 Les valeurs fusionnées de FL

La frontière entre FL paradigmatique et FL syntagmatique n'est pas étanche. Une FL syntagmatique peut servir à connecter deux lexies qui sont en fait liées par un lien de dérivation sémantique. Le lien collocationnel est à ce moment comme intégré au sens de lexie retournée par la FL. On parle dans un tel cas de **VALEUR FUSIONNÉE** de la FL. Les valeurs fusionnées apparaissent à droite du symbole // dans les encodages de liens de FL. Par exemple :

**Magn(victoire)** = *grande | antépos; claire, complète, éclatante, écrasante, retentissante // triomphe.*

Dans la liste de valeurs ci-dessus, le symbole // nous indique que TRIOMPHE est un cas de fusion du **Magn** : *triomphe = victoire + Magn(victoire)*. En d'autres termes, TRIOMPHE, en tant que valeur de **Magn**, ne correspond pas ici à un cas de collocation, mais plutôt à un cas de dérivation sémantique ; on aurait pu aussi bien encoder la relation entre VICTOIRE et TRIOMPHE par la formule suivante :

**Syn<sub>⊃</sub>(victoire) = triomphe,**

où l'indice  $\supset$  indique que le synonyme est sémantiquement plus riche.

Il faut noter qu'une valeur est considérée comme fusionnée même si elle correspond tout de même à un collocatif véritable dans le cas où elle se trouve en fait gouverner syntaxiquement la base de la collocation alors que la FL impliquée fonctionne normalement de façon inverse :

**Magn<sup>quant</sup>(injure)** = // *chapelet [d'~s] < torrent [d'~s]*

**Magn(L)** est normalement un modificateur de L et pourtant la valeur ci-dessus est constituée de noms régissant syntaxiquement la base, d'où l'utilisation du symbole de fusion.

## 4 Un sous-ensemble du système des fonctions lexicales

Cette section constitue la partie centrale du tutoriel, là où est enseigné un sous-ensemble représentatif du système des fonctions lexicales et où sont mises en pratique les définitions ci-dessus. L'essentiel du travail va se faire sur la base de travaux pratiques visant la modélisation de liens lexicaux à partir de données textuelles.

### 4.1 Analyse de données

L'analyse de données peut prendre au moins deux formes, que l'on examinera tour à tour durant le tutoriel :

1. On peut analyser les données linguistiques trouvées dans des corpus électroniques dans le cadre de la rédaction de descriptions lexicographiques. Dans ce cas, on utilise surtout des concordances.



Exercice d'ébauche d'un article de dictionnaire du type DiCo (Polguère, 2000b) avec recours à un corpus électronique.

2. Les FL permettent d'analyser certains procédés linguistiques utilisés dans des textes. Il peut s'agir de textes littéraires ou, même, de textes d'élèves (dont on cherche à améliorer la maîtrise de la langue).



On peut par exemple chercher à identifier tous les liens de FL mis en jeu dans le passage suivant :

*On s'est mis en rang et notre chef nous a comptés. « Onze ! il a dit. Il en manque un ! » C'était Paulin, qui était assis dans l'eau et qui ne voulait pas en sortir.*

*— Je veux rester dans l'eau ! il criait. Si je sors, je vais avoir froid ! Je veux rester !*

*Le chef, qui avait l'air de s'énerver, l'a ramené en le tirant par le bras et Paulin criait qu'il voulait rentrer chez son papa, chez sa maman, et dans l'eau.*

Sempé J.-J., Goscinny R. (1997), *Les vacances du petit Nicolas*, Folio Junior, Édition spéciale, n° 457, Paris, Gallimard Jeunesse.

## 4.2 Importance des fonctions lexicales non standard

Même si les FL non-standard sont en quelque sorte des FL « anormales » du point de vue la notion théorique de FL elle-même, elles ont une importance pratique (descriptive) considérable. Il est pratiquement impossible de trouver une lexie contrôlant un ensemble assez riche de FL qui soient toutes standard. Il y aura toujours parmi les FL associées à une lexie au moins quelques FL non-standard. Certaines lexies, notamment celles dénotant des entités concrètes, ont même tendance à contrôler un très grand nombre de liens non standard ou semi-standard. Voici par exemple la zone de FL de l'article CHAT du DiCo — pour une description des formalismes propres au DiCo, voir (Polguère, 2000b) :

```
{QSyn} matou, minet, minou, "vieilli" mistigri, "arg" greffier
{Gener} animal domestique; animal de compagnie; félin-n
{Connotation} agilité, souplesse; délicatesse, prudence; paresse
{Femelle} chatte
{Petit du C.} chaton
{A0} de [~] //félin-adj ["une démarche de chat <féline>"]
/*Nom pour X*/
{S1} maître, propriétaire [de ART ~]
/*De grande taille*/
{Magn} gros | postpos < énorme
/*Qui n'a plus de X <Qui vit à l'état sauvage>*/
{sans X AntiVer} abandonné, errant, "rare" haret-adj, sauvage
| postpos //"rare" haret-n
{Recherché par X} perdu | postpos
{D'une espèce
particulière} de gouttière, européen | postpos; birman, persan
| postpos; abyssin-adj/n, angora-adj/n, siamois-adj/n
| postpos si adj //chartreux
```

{Dont le pelage est rayé} tigré | postpos  
{Dont le pelage est de trois couleurs} tricolore | postpos  
{Dont le pelage est en mauvais état} galeux, pelé | postpos  
{Qui a été rendu stérile [=f1]} opéré | postpos; castré,  
châtré | postpos  
{[X] faire en sorte que f1} faire castrer/opérer/stériliser [ART ~]  
{[C. femelle] en rut} \_en chaleur\_, \_en chasse\_  
{Qui chasse les animaux N} tueur [de N-pl] ["un chat tueur de rats"]  
/\*Faire des caresses à un C.\*/  
{Real\^-caresse} câliner, caresser, gratter [ART ~]  
/\*C. fait des mouvements\*/  
{Fact0-mouvement} \_faire le gros dos\_; s'étirer; se pelotonner,  
se mettre/rouler \_en boule\_, se rouler \_en pelote\_;  
se raccourcir [pour sauter]  
/\*C. se nettoie le pelage\*/  
{Fact0-nettoyage} se laver, se lécher, se lustrer, faire sa toilette  
/\*C. gratte un meuble <une surface> avec ses griffes\*/  
{Fact0-gratter} faire ses griffes, se faire les griffes  
/\*C. se repose\*/  
{NonFact0} faire la sieste, se prélasser  
/\*C. chasse l'animal N\*/  
{Fact0-chasser} chasser [N], faire la chasse[à N]  
/\*Animal que chassent habituellement les C.\*/  
{S2Fact0-chasser} oiseau, souris, rat  
/\*C. attaque N\*/  
{AntiBonInvolv} griffer, mordre [N]  
/\*C. prend une attitude menaçante\*/  
{PreparAntiBon2Involv} sortir ses griffes  
/\*C. cesse d'avoir une attitude menaçante\*/  
{LiquPreparAntiBon2Involv} rentrer ses griffes; \_faire patte de  
velours\_  
/\*C. boit\*/  
{Fact0-boire} laper  
/\*Ce qu'aiment boire les C.\*/  
{S2--usualFact0-boire} lait#b  
/\*C. émet un cri\*/  
{Son-cri} miauler; feuler | pour exprimer la colère ou la menace  
/\*Cri du C.\*/  
{S0Son-cri} miaulement, « Miaou ! »; feulement  
/\*C. émet un son de gorge exprimant le plaisir [=f2]\*/  
{Son-plaisir} "enfantin" faire ronron, ronronner  
/\*Nom pour f2\*/  
{S0Son-plaisir} ronron, ronronnement  
/\*C. émet un son de gorge exprimant la menace\*/  
{Son-menace} cracher  
{Élément du corps d'un C.} moustaches; queue; coussinet, griffe,  
patte; poil, fourrure, pelage, robe  
{Caractéristique de la langue des C.} râpeuse  
{Excrément du C.} crotte [de ~]; "fam" pipi, "fam" pisse [de ~]  
{Gravier spécial  
sur lequel le C. fait ses besoins} litière  
{Contenant dans lequel le C. fait ses besoins} bac (à litière)  
{Petite ouverture pratiquée en bas  
d'une porte qui permet à un C. d'entrer et sortir} chatière  
{Ce que l'on dit pour appeler un C.} « Minet ! », « Minou ! »,  
« Petit, petit ! »  
{Nom propre typique des C.} Gris-Gris, Mistigri

{Nom de personnage de C.  
dans la littérature} \_le Chat botté\_, Raminagrobis

### 4.3 Le problème de la variation individuelle et dialectale

L'encodage des liens de FL nous oblige à faire face au problème de la variation individuelle et dialectale. Un francophone de France aura par exemple beaucoup de mal à enseigner de façon pratique les FL au Québec, s'il n'est pas déjà familiarisé avec le français québécois. Beaucoup de ses exemples tomberont à plat et il devra les adapter en fonction de son public. L'inverse est bien entendu tout aussi vrai. Un dictionnaire de collocations comme (Benson *et al.*, 1997), par exemple, qui vise un public européen aussi bien qu'américain, va très souvent avoir recours aux marques d'usage BE (pour *British English*) et AE (pour *American English*) dans ses articles.

Si la variation dialectale pose problème pour la modélisation du lexique, on notera cependant que les FL peuvent justement être un outil performant pour rendre compte des écarts dialectaux. Elles permettent notamment de bien distinguer les variations qui relèvent d'une différence au niveau du stock lexical *vs* d'une différence au niveau de la combinatoire des lexies. Par exemple, la fameuse expression québécoise *tomber en amour* ne devrait pas se décrire par le biais de l'introduction d'une nouvelle lexie TOMBER propre au français québécois (comme on le voit faire parfois), mais bien en tant que collocation de AMOUR :

**IncepOper**<sub>1</sub>(*amour*) = Québ *tomber* [*en ~*].

Pour un exemple d'utilisation des FL dans la mesure des écarts dialectaux, voir (Lim, Polguère, 1992).

### 4.4 Récapitulatif des fonctions lexicales introduites

Je ne récapitule ici que les principales FL standard qui auront été examinées durant le tutoriel. On trouvera dans (Me'čuk, 1996) la liste complète des FL standard simples. Celle-ci est aussi disponible dans les volumes du *Dictionnaire explicatif et combinatoire du français contemporain*, notamment le Volume IV (Mel'čuk *et al.*, 1999).

#### 4.4.1 Principales FL paradigmatiques examinées

1. **Syn** (du latin *synonymum*) est la FL qui associe à une lexie ses synonymes exacts ou approximatifs :

**Syn**(*voiture*) = *automobile*, **fam** *auto*, **Québ fam** *char*

**Syn**(*individu*) = **fam** *gars*, **fam** *type*, **fam** *mec*

**Syn**<sub>⊂</sub>(*avion*) = *appareil3*

APPAREIL3 est un synonyme moins riche (un hyperonyme) de AVION : son sens est inclus dans celui de AVION, comme l'indique le symbole d'inclusion mis en indice.

**Syn**<sub>⊃</sub>(*appareil3*) = *avion*

À l'opposé, AVION est un synonyme plus riche (un hyponyme) de APPAREIL3 : son sens inclut celui de APPAREIL3, comme l'indique le symbole d'inclusion inverse.

**Syn**<sub>∩</sub>(*jouer*) = *s'amuser*

S'AMUSER est un synonyme « à intersection » de JOUER : son sens possède une intersection significative avec celui de JOUER, comme l'indique le symbole d'intersection. Ceci est démontré par le fait que, bien que les lexies JOUER et S'AMUSER peuvent être substi-

tuées l'une à l'autre dans de nombreux contextes, on peut tout à fait jouer sans s'amuser et, à l'inverse, s'amuser sans jouer. Il y a donc ici intersection et non inclusion de sens.

2. **Anti** (du latin *antonymum*) est la FL qui associe à une lexie ses antonymes :

**Anti**(*petit*) = *grand*  
**Anti**(*haut*<sub>Adj</sub>) = *bas*<sub>Adj</sub>

3. **Conv<sub>ijk</sub>** (du latin *conversivum*) est la FL qui associe à une lexie ses conversifs. La suite **ijk** identifie la permutation des actants :

**Conv<sub>321</sub>**(*donner*) = *recevoir*  
 [Actant<sub>1</sub> *donne* Actant<sub>2</sub> à Actant<sub>3</sub> ≡ Actant<sub>3</sub> *reçoit* Actant<sub>2</sub> de Actant<sub>1</sub>]

4. **S<sub>0</sub>** (*S* pour *substantif*) est la FL qui associe à une lexie verbale, adjectivale ou adverbiale sa contrepartie nominale :

**S<sub>0</sub>**(*courir*) = *course*  
**S<sub>0</sub>**(*dormir*) = *sommeil*

5. **V<sub>0</sub>** est le pendant verbal de **S<sub>0</sub>**, qui associe donc à une lexie nominale, adjectivale ou adverbiale sa contrepartie verbale :

**V<sub>0</sub>**(*sommeil*) = *dormir*

6. **S<sub>1</sub>, S<sub>2</sub>, ...** lient une lexie prédicative au nom standard de son premier, deuxième, ... actant :

**S<sub>1</sub>**(*courir*) = *coureur*  
**S<sub>1</sub>**(*dormir*) = *dormeur*  
**S<sub>1</sub>**(*voler*) = *voleur, coupable*  
**S<sub>2</sub>**(*voler*) = *butin*  
**S<sub>3</sub>**(*voler*) = *victime*

#### 4.4.2 Principales FL syntagmatiques examinées

1. **Magn** (du latin *magnus* 'grand') associe à une lexie l'ensemble des lexies ou expressions linguistiques qui expriment auprès d'elle l'intensification, c'est-à-dire le sens général 'intense', 'très', 'beaucoup', etc. :

**Magn**(*chagrin*) = *grand, gros* < *énorme, immense*  
**Magn**(*amour*) = *grand* < *immense*  
**Magn**(*courir*) = *vite* < <sup>⌈</sup>*à fond de train*<sup>⌋</sup>, <sup>⌈</sup>*à perdre haleine*<sup>⌋</sup>  
**AntiMagn**(*chagrin*) = *petit*

Dans les exemples ci-dessus, le symbole « inférieur à » indique une gradation entre les différents éléments de la valeur retournée. Ainsi, un énorme chagrin est plus « intense » qu'un gros chagrin.

2. **Bon** (du latin *bonus* 'bon') associe à une lexie l'ensemble des lexies ou expressions linguistiques qui expriment auprès d'elle le sens général 'bon', 'bien', etc. — c'est-à-dire, l'évaluation positive/l'approbation du locuteur :

<b>Bon</b> (colère)	= saine, sainte
<b>Bon</b> (compliment)	= bien tourné

Les trois FL suivantes permettent de décrire les constructions dites à *VERBE SUPPORT*.

3. **Oper<sub>i</sub>** (du latin *operari* 'faire') associé à une lexie L un verbe sémantiquement vide qui prend L comme premier complément et le i<sup>ème</sup> actant de L comme sujet grammatical pour donner une construction qui est fonctionnellement équivalente à une verbalisation de L :

<b>Oper<sub>1</sub></b> (coup)	= asséner, donner, fam flanquer [ART ~] [asséner/donner/flanquer un coup ≡ frapper]
<b>Oper<sub>2</sub></b> (coup)	= fam encaisser, prendre, recevoir [ART ~] [encaisser/prendre/recevoir un coup ≡ être frappé]

4. **Func<sub>i</sub>** (du latin (de cuisine) *functionare* 'fonctionner') associé à une lexie L un verbe sémantiquement vide qui prend L comme sujet grammatical et le i<sup>ème</sup> actant de L comme premier complément ; i = 0 si **Func** ne prend pas de complément :

<b>Func<sub>0</sub></b> (pluie)	= tomber
<b>Func<sub>1</sub></b> (éloge)	= provenir, venir [de N <sub>X</sub> ]

5. **Labor<sub>ij</sub>** (du latin *laborare* 'travailler') associé à une lexie L un verbe sémantiquement vide qui prend le i<sup>ème</sup> actant de L comme sujet grammatical, le j<sup>ème</sup> actant de L comme premier complément et L comme second complément :

<b>Labor<sub>12</sub></b> (aversion)	= avoir, tenir [N <sub>Y</sub> en ~]
<b>Labor<sub>21</sub></b> (effroi)	= remplir, saisir < glacer [N <sub>X</sub> d'~]

Les **Oper<sub>i</sub>**, **Func<sub>i</sub>** et **Labor<sub>ij</sub>**, qui sont sémantiquement vides, s'opposent aux verbes dits de « réalisation », modélisés par les FL **Real<sub>i</sub>**, **Fact<sub>i</sub>** et **Labreal<sub>ij</sub>**. Ces verbes contrôlent les mêmes structures syntaxiques que les FL de verbes supports, mais ont un contenu sémantique ; celui-ci peut être caractérisé de façon très vague comme étant la réalisation, la satisfaction de la situation dénotée par le mot-clé de cette FL.

6. **Real<sub>i</sub>** (du latin *realis* 'réel') est le verbe de réalisation correspondant (syntaxiquement) à **Oper<sub>i</sub>** :

<b>Real<sub>1</sub></b> (projet)	= accomplir, concrétiser, réaliser [ART ~]
<b>Real<sub>2</sub></b> (avertissement)	= <sup>⌈</sup> prendre en considération <sup>⌋</sup> , <sup>⌈</sup> prendre en compte <sup>⌋</sup> , suivre [ART ~], <sup>⌈</sup> tenir compte <sup>⌋</sup> [de ART ~]; écouter, entendre [ART ~]
<b>Real<sub>1</sub></b> (vêtement)	= porter [ART ~]

7. **Fact<sub>i</sub>** (du latin *factum* 'fait') est le verbe de réalisation correspondant (syntaxiquement) à **Func<sub>i</sub>** :

<b>Fact<sub>0</sub></b> (stratagème)	= marcher, réussir, fonctionner
<b>AntiFact<sub>0</sub></b> (stratagème)	= échouer
<b>Fact<sub>0</sub></b> (remède)	= agir
<b>Fact<sub>1</sub></b> (remède)	= guérir, soigner [N <sub>X</sub> ]

**Fact**<sub>2</sub>(remède) = *aider* [contre N<sub>Y</sub>], *combattre* [N<sub>Y</sub>]

8. **Labreal**<sub>ij</sub> (nom hybride de **Labor** et **Real**) est le verbe de réalisation correspondant (syntaxiquement) à **Labor**<sub>ij</sub> :

**Labreal**<sub>12</sub>(gratitude) = *assurer* [N<sub>Y</sub> de ART ~]

**Labreal**<sub>12</sub>(pelle) = *creuser* [N<sub>Y</sub> avec ART ~]; *remuer* [N<sub>Y</sub> avec ART ~]; *ramasser* [N<sub>Y</sub> à la ~ <avec ART ~>] // *pelleter* [N<sub>Y</sub>]

**Labreal**<sub>21</sub>(excuse) = *donner*, *faire valoir*, *invoquer* [N<sub>X</sub> comm/en ~] < brandir [N<sub>X</sub> comme ~]; saisir [N<sub>X</sub> comme ~] // *alléguer* [que PROP<sub>X</sub>]

## Références

Une adresse de site Internet est donnée pour les textes dont une version électronique peut être téléchargée.

Daille B. (2001), Extractions de collocations à partir de textes, Actes de *TALN 2001*, Tome 2, Tours, pp. 3-8.

Firth J. (1957), Modes of Meaning (1951), in *Papers in Linguistics 1934-1951*, Oxford, Oxford University Press, pp. 190-215.

Kahane S. (2001), Grammaires de dépendance formelles et théorie Sens-Texte, Actes de *TALN 2001*, Tome 2, Tours, pp. 17-76.

Kahane S., Polguère A. (2001), Formal foundation of lexical functions, Actes de *COLLOCATION: Computational Extraction, Analysis and Exploitation*, 39<sup>th</sup> Annual Meeting and 10<sup>th</sup> Conference of the European Chapter of the Association for Computational Linguistics, Toulouse, pp. 8-15.

[<http://www.fas.umontreal.ca/ling/olst/FrEng/SKAPolColloc.pdf>]

Lim G., Polguère A. (1992) Can Lexical Functions Help to Characterize Singapore English?, in A. Pakir (éd.) : *Words in a Cultural Context*, Singapour, UniPress, pp. 154-171.

Mel'čuk I. (1988), Paraphrase et lexique dans la théorie linguistique Sens-Texte — Vingt ans après, *Revue internationale de lexicologie et lexicographie*, Vol. 52/53, pp. 5-50/5-53.

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

Mel'čuk I. (1996), Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon, in L. Wanner (éd.) : *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam/Philadelphia, Benjamins, pp. 37-102.

Mel'čuk I. (1997), *Vers une linguistique Sens-Texte. Leçon inaugurale (faite le Vendredi 10 janvier 1997)*, Collège de France, Chaire internationale.

[<http://www.fas.umontreal.ca/ling/olst/FrEng/MelcukColldeFr.pdf>]

Mel'čuk I. (2001), *Communicative Organization in Natural Language: The semantic-communicative structure of sentences*, Amsterdam/Philadelphia, Benjamins.

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

Polguère A. (1998), La théorie Sens-Texte, *Dialangue*, Vol. 8-9, pp. 9-30.

[<http://www.fas.umontreal.ca/ling/olst/FrEng/PolgIntroTST.pdf>]

Polguère A. (2000a), Une base de données lexicales du français et ses applications possibles en didactique, *Revue de Linguistique et de Didactique des Langues (LIDIL)*, Vol. 21, pp. 75-97.

Polguère A. (2000b), Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French, Actes de *EURALEX'2000*, Stuttgart, pp. 517-527.

[<http://www.fas.umontreal.ca/ling/olst/FrEng/APoleURALEX.pdf>]

Polguère A. (à paraître), *L'étude du lexique : Système de notions de base en lexicologie et sémantique*.

Tesnière L. (1959), *Éléments de syntaxe structurale*, Paris, Klincksieck.

Williams G. (2001), Sur les caractéristiques de la collocation, Actes de *TALN 2001*, Tome 2, Tours, pp. 9-16.

#### Dictionnaires cités :

Beauchesne J. (2001), *Dictionnaire des cooccurrences*, Montréal, Guérin.

Benson M., Benson E., Ilson R. (1997), *The BBI dictionary of English word combinations*, Amsterdam/Philadelphia, Benjamins.

Kirk-Greene C. (1997), *French Idioms and Expressions*, Londres et al., Foulsham.

*Nouveau Petit Robert* (1993), Paris, Dictionnaires Le Robert.

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

Rey A., Chantreau S. (1997), *Dictionnaire des expressions et locutions*, Les usuels du Robert, Paris, Dictionnaires Le Robert.