



Collocations et fonctions lexicales : pour un modèle d'apprentissage

Alain Polguère

OLST — Département de linguistique et de traduction

Université de Montréal

alain.polguere@umontreal.ca

<http://www.fas.umontreal.ca/ling/olst/polguere>

Résumé

Au vu des résultats théoriques et descriptifs obtenus en lexicologie explicative et combinatoire, on pourrait postuler que le problème de la modélisation des collocations a déjà trouvé une solution, dans ses grandes lignes, grâce à la notion Sens-Texte de fonction lexicale. Une telle perspective ne fait cependant que déplacer le problème dans la mesure où le système formel des fonctions lexicales lui-même s'avère, dans la pratique, très difficile à enseigner, acquérir et manipuler. Après avoir tenté de justifier le point de vue adopté ici — focalisation sur la notion de fonction lexicale —, je vais montrer comment les fonctions lexicales peuvent et doivent être appréhendées indépendamment de leur formalisation habituelle. Je proposerai ensuite des pistes pour l'enseignement des fonctions lexicales aux spécialistes de la langue (linguistes, étudiants en langue et linguistique, enseignants et futurs enseignants de langues, etc.). La conclusion mettra l'accent sur l'importance du lien unissant fonctions lexicales et paraphrase linguistique ainsi que sur l'utilisation possible de ce lien dans un contexte d'enseignement des fonctions lexicales.

Mots-clés : fonction lexicale, collocation, dérivation sémantique, lexicologie explicative et combinatoire, théorie Sens-Texte.

1 Introduction

Avant toute chose, puisque le thème du présent ouvrage est la modélisation des collocations, je tiens à préciser ce que je vais entendre dans la suite du texte par les deux termes *collocation* et *modélisation*.

Je vais considérer la notion de collocation telle que postulée dans le cadre de la théorie Sens-Texte, c'est-à-dire une notion qui se définit non sur des bases statistiques (haute fréquence d'apparition de deux cooccurrents dans des corpus de référence) mais sur des bases strictement fonctionnelles et sémantiques, en tant qu'expression semi-idiomatique. Pour une définition de cette notion, voir l'article d'I. Mel'čuk dans le présent volume.

Passons maintenant à ce qu'il faut entendre par *modélisation*. La modélisation d'un phénomène linguistique tel que la collocation est une représentation (formelle ou non) qui permet d'expliquer, anticiper ou reproduire ce phénomène. La notion de modélisation est somme toute assez vague et il est souhaitable de ne pas chercher à spécifier plus les finalités possibles d'une modélisation linguistique.

Ma contribution à l'étude du problème de la modélisation des collocations (tel que spécifié ci-dessus) va être légèrement biaisée, et je m'en excuse par avance auprès du lecteur. En effet, je vais prendre pour acquis que ce problème a déjà trouvé une solution satisfaisante dans ses grandes lignes : les *FONCTIONS LEXICALES* de la théorie Sens-Texte. Je pourrais bien entendu tenter de démontrer ici pourquoi il me semble que les fonctions lexicales (dorénavant, *FL*) sont une

réponse adéquate au problème de la modélisation des collocations. Mais, d'une part, cette tâche a été effectuée dans nombre de publications antérieures et, d'autre part, je pense que seule la pratique de la description systématique de collocations, et non une « démonstration » logique, peut parvenir à mettre véritablement en lumière l'efficacité descriptive des FL. J'ai dit que ma réponse à la question posée allait être biaisée parce que, considérant les FL comme une réponse adéquate à la modélisation des collocations, je vais plutôt me concentrer ici sur le problème de l'enseignement et de la diffusion de la notion de FL elle-même. Il me semble en effet que nous sommes dans une situation où, une question ayant été posée (comment modéliser les collocations), une réponse satisfaisante a été fournie (les FL) mais dans une langue que presque personne ne sait utiliser. Ma contribution à la résolution du problème de la modélisation des collocations va donc consister à examiner le problème de la diffusion et de la maîtrise de la notion de FL.

On pourrait objecter à ce qui vient d'être dit que les FL sont un outil général permettant de modéliser aussi bien les liens paradigmatiques, appelés *dérivations sémantiques*, que les liens de cooccurrence au sein de collocations. En traitant des FL, je déborderais donc du seul problème de la modélisation des collocations. Cependant, l'outil des FL permet justement de mettre en évidence le lien conceptuel profond qui existe entre dérivations sémantiques (synonymie, antonymie, conversivité, nom d'actants ou de circonstants, etc.) et collocations. Ces deux types de phénomènes relèvent de liens lexicalement contrôlés et leur standardisation fait appel aux mêmes universaux linguistiques. Prenons un exemple pour illustrer ce fait. Il serait faux de prétendre que le problème de la modélisation du lien existant entre la lexie WAGON et la lexie LOCOMOTIVE ne relève pas, de façon indirecte, de la modélisation des collocations. Voici pourquoi :

- WAGON, comme la plupart des lexies dénotant des artefacts, est la base de collocations où elle est sujet d'un verbe (le collocatif) signifiant 'fonctionner' : *Le wagon roule/circule*. En terme de FL, il s'agit de la relation **Fact₀**.
- Le sens de causation s'exprime souvent au sein de collocations, notamment en se combinant avec d'autres sens collocationnels. Dans le cas de WAGON, 'causer que le wagon fonctionne' peut s'exprimer par *pousser/tirer un wagon*. Nous sommes ici en présence de la relation **Caus_{engin}Fact₀**.
- Pour chaque sens prédicatif, il est légitime de s'attendre à ce que la langue nous offre un ou plusieurs noms permettant de désigner chacun des actants du prédicat en question. Il existe en français au moins deux noms pour désigner un engin dont la fonction est de « causer qu'un wagon fonctionne » : LOCOMOTIVE et MOTRICE.
- Le lien sémantique paradigmatique entre WAGON et LOCOMOTIVE, qui peut se paraphraser par 'engin dont la fonction est de faire fonctionner un wagon', s'appuie donc sur deux significations s'exprimant couramment au sein de collocations ('causer' et 'fonctionner') et sur une relation paradigmatique (nom du premier actant).
- La modélisation du lien entre WAGON et LOCOMOTIVE faite au moyen des FL

S₁Caus_{engin}Fact₀(wagon) = locomotive, motrice

rend explicite ce lien fonctionnel entre dérivations sémantiques et collocations, montrant par là-même que les deux types de notions ne devraient sans doute pas être considérés séparément.

En conclusion, il serait tout à fait artificiel de se limiter ici aux FL dites syntagmatiques (celles destinées à encoder les collocations) sous prétexte que nous nous intéressons à la modélisation

des collocations. La notion de FL nous montre justement la voie à suivre en nous forçant à envisager de façon intégrée tous les liens lexicaux.

Cet article n'a pas pour finalité de présenter une énième introduction à la notion de FL. Je présume que les FL (Mel'čuk *et al.*, 1995 :125-152 ; Mel'čuk, 1996) ainsi que les principes de base de la théorie Sens-Texte sont déjà connus du lecteur. Mon propos est ici de tenter de démontrer les deux faits suivants :

1. Les FL ne sont pas une fiction théorique développée dans le cadre de la théorie Sens-Texte, mais correspondent bien à un fait linguistique universel, observable, dont la maîtrise est une partie essentielle de notre connaissance linguistique.
2. Pour être bien comprise, la notion de FL doit être considérée indépendamment de sa formalisation actuelle, car cette dernière n'est pas appropriée en tant que support pour l'apprentissage de la notion et pour son utilisation.

On l'aura compris : mon but est ici fondamentalement pédagogique. Il s'agit d'examiner comment on peut rendre la notion très technique de FL accessible à tous. Par *tous*, j'entends toute personne intéressée à mener une étude approfondie de la langue, qui va au-delà du simple apprentissage du code linguistique lui-même. Il s'agit donc notamment des étudiants en linguistique, sciences du langage mais aussi, des futurs enseignants de langues (langue maternelle ou langue seconde). Dans la mesure où les FL devraient pouvoir être appréhendées sans qu'il soit nécessaire de recevoir une formation véritable en linguistique Sens-Texte, il est clair qu'il me faudra aborder ici le problème de la vulgarisation des notions linguistiques Sens-Texte. Notons toutefois que je vais devoir limiter mon propos et ne pas couvrir tous les contextes envisageables d'enseignement des FL. On peut ainsi identifier deux contextes majeurs :

1. l'enseignement des FL — en tant qu'outil descriptif — aux étudiants de linguistique, aux professeurs ou futurs professeurs de langue maternelle et langue seconde, aux traducteurs, etc. ;
2. l'enseignement des FL sous une forme vulgarisée ou « déguisée » — en tant que notion à caractère explicatif permettant d'améliorer la compréhension et la maîtrise d'une langue — aux élèves du primaire ou du secondaire.

Dans ce qui suit, je me focaliserai sur la première problématique : l'enseignement des FL aux « professionnels » ou futurs professionnels de la langue.

Cet article se divise en deux grandes sections. Premièrement, je vais introduire la notion de FL en la dégageant du formalisme auquel elle est habituellement associée. Ensuite, je proposerai des pistes pour présenter, enseigner et utiliser les FL dans des contextes autres que celui de l'étude linguistique Sens-Texte « pure et dure ». La conclusion mettra l'accent sur l'importance du lien unissant FL et paraphrase linguistique ainsi que sur l'utilisation possible de ce lien dans un contexte d'enseignement du système des FL.

2 Les fonctions lexicales toutes nues

2.1 Pourquoi distinguer les fonctions lexicales de leur formalisation ?

Par *fonctions lexicales toutes nues*, j'entends ici les FL en tant que notion linguistique fondamentale, dépouillée de l'appareillage formel d'encodage auquel elle est habituellement associée. Je pense en effet que les FL peuvent être enseignées, comprises et même utilisées comme

outil conceptuel indépendamment de leur formalisme d'encodage traditionnel. De plus, puisque les FL ont une existence non formelle propre, il est essentiel qu'elles soient appréhendées de cette façon. Je vais tenter d'expliquer ce point de vue.

Il est possible de dire, en simplifiant un peu, que toute modélisation formelle d'un phénomène naturel peut mettre en jeu deux grandes familles de notions. Il y a tout d'abord les notions « primaires », qui sont des artifices formels liés directement à certains phénomènes naturels. Ainsi, en statistique, la notion de courbe, qui permet de représenter de façon visuelle une fluctuation statistique, est un appareillage formel — la courbe — associé à un fait naturel — la variation quantitative d'un fait à travers une « population » statistique donnée : la note *A+* a été obtenue par 3 étudiants, la note *A* par 5, la note *A-* par 7, la note *B+* par 12, etc. À partir de ces notions formelles primaires, on peut développer des notions formelles que l'on pourrait appeler « dérivées », qui ne se comprennent que par rapport aux premières. Ainsi, toujours en statistique, la notion d'aire sous la courbe, qui est à la base de nombreux calculs, est entièrement dérivée de la notion de courbe elle-même, choisie pour représenter une fluctuation dans les données. Prenons maintenant un exemple emprunté à la linguistique. Les formalismes d'arbre syntagmatique et d'arbre de dépendance, utilisés pour modéliser la structure syntaxique des phrases, peuvent être considérés comme des notions primaires ; ce sont des métaphores formelles d'un phénomène naturel. Cependant, les théories linguistiques utilisant les arbres syntaxiques ont toutes développé des notions qui ne peuvent se comprendre qu'en référence au formalisme d'arbre, même si, bien entendu, elles visent en bout de ligne à rendre compte de phénomènes linguistiques « réels » ; par exemple : la notion de C-commande (Reinhart, 1983), associée aux arbres syntagmatiques, ou la notion de (non) projectivité (Lecerf, 1960), associée aux arbres de dépendance. Les arbres syntaxiques sont des métaphores visuelles de la structure syntaxique des énoncés ; on trouvera dans Polguère (2002) une réflexion sur les métaphores visuelles, appliquées à la représentation du sens linguistique.

Les FL appartiennent clairement à la première famille de notions décrite ci-dessus. Les FL ne sont pas une métaphore formelle développée à partir d'une autre métaphore formelle. Elles visent à rendre compte d'un phénomène linguistique naturel de façon brute, directe. Or, je suis convaincu qu'une telle notion doit absolument être enseignée et assimilée de façon brute, sans la médiation du formalisme servant à la représenter et à la manipuler. Il faut d'abord mettre en évidence et faire observer la « réalité » du phénomène naturel dont il est question, et ensuite seulement proposer un outil formel, ou semi-formel, pour le modéliser.

Malheureusement, les FL sont la plupart du temps enseignées comme le serait une suite de notions mathématiques. Une présentation traditionnelle des FL, à peine caricaturée, ressemble à peu près à ceci :

- Il existe une soixantaine fonctions lexicales servant à modéliser les liens qu'entretiennent les lexies de la langue.
- On distingue deux groupes de fonctions lexicales : les *FL PARADIGMATIQUES* vs les *FL SYNTAGMATIQUES*.
- La première fonction lexicale paradigmatique est *Syn*, pour la relation de synonymie...

Vient alors une longue énumération des FL, agrémentée d'exemples, suivie d'une présentation des notions de *FL COMPLEXE* et de *CONFIGURATION DE FL*. Selon cette méthode, l'essentiel du travail de compréhension se ramène à une tâche de mémorisation des formalismes introduits

pour encoder les différentes FL. Il faut ainsi d'abord mémoriser une liste assez longue de FL simples :

Syn, Anti, Conv_{ij}, Contr, Epit, Gener, Figur, S₀, A₀, V₀, Adv₀, S_i, S_{instr}, S_{med}, S_{mod}, S_{loc}, S_{res}, Able_i, etc.

Même si certains regroupements sont proposés — FL paradigmatiques vs syntagmatiques, FL de verbes supports, FL de verbes de « réalisation », etc. —, il s'agit avant tout de mémoriser une liste, plutôt qu'un système structuré. Il faut ensuite assimiler plusieurs modes de combinaison des FL simples, fondés sur des opérations approximativement définies d'un point de vue formel. Je ne m'étendrai pas ici sur la critique de la validité formelle du mode d'encodage traditionnel des FL ; le lecteur intéressé pourra consulter Kahane & Polguère (2001). Ce qui est clair, c'est que le nombre de toutes les FL pouvant résulter de la combinaison de FL simples est suffisamment énorme pour qu'il soit vain de tenter d'en donner une liste exhaustive. On doit donc la plupart du temps se contenter de se familiariser avec les FL complexes et les configurations de FL les plus courantes :

AntiMagn, AntiBon, AntiVer, IncepPredPlus, IncepPredMinus, MagnReal₁, Magn+Real₁, etc.

Je reviendrai plus loin sur des propositions concrètes pour enseigner les FL. Pour l'instant, contentons-nous de noter que les FL forment un système, qui doit être appris comme tel. De plus, comme ce système est trop complexe pour être appréhendé d'un bloc, il faut parvenir d'abord à une compréhension en profondeur d'un fragment du système des FL. Cela implique une compréhension intuitive claire des FL en tant que phénomène linguistique « naturel » et de leur nature profonde.

Finalement, les FL telles que formalisées dans la théorie Sens-Texte ne représentent pas une métaphore « évidente » des phénomènes dont elles permettent de rendre compte. Le formalisme sur lequel elles reposent se distingue clairement, de par son fonctionnement, d'un formalisme tel que celui des arbres syntaxiques. Ainsi, lorsque l'on associe la Figure 1 ci-dessous à la phrase *Il fait un temps de chien*, on offre une métaphore visuelle explicite, relativement évidente, de la structure syntaxique qui hiérarchise les éléments de cette phrase :

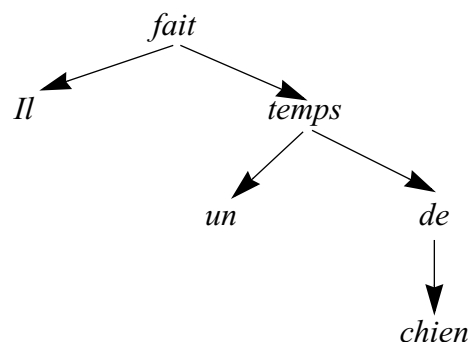


Figure 1 Arbre représentant la hiérarchie syntaxique des éléments lexicaux de *Il fait un temps de chien*.

Il en va tout autrement avec le formalisme des FL. Le fait de représenter le collocatif *de chien* apparaissant dans notre phrase d'exemple par la formule « **AntiBon**(*temps*) » ne correspond qu'à la traduction d'une expression linguistique dans un autre langage linéaire qui, de façon circulaire, ne peut lui-même être interprété qu'en tant que paraphrase approximative des expressions linguistiques qu'il modélise : **AntiBon** ne se comprend comme modélisation de [*temps*] *de chien*

que si on lui associe sa signification linguistique ‘qui est le contraire de bon’. La modélisation formelle offerte par les FL est somme toute une pseudo-langue : son principe de fonctionnement sémiotique n’est pas externe à la langue elle-même. Attention cependant ! Il ne faut pas conclure de ces remarques que les bases formelles de la modélisation des FL sont mauvaises. Il est fort probable qu’il n’existe pas de meilleure façon de modéliser un lien lexical que de le paraphraser au moyen d’une pseudo-langue. Simplement, il faut être conscient du fait qu’une formule linéaire d’encodage de FL aura rarement une interprétation évidente et devra toujours être ramenée à son équivalent linguistique ; par exemple :

- **Magn** se traduit par *intense* si le mot-clé est un nom de sentiment, par *grand* si le mot-clé est un nom d’objet, etc.
- **Oper₁** se traduit par *faire* si le mot-clé est un nom d’action, par *éprouver* si le mot-clé est un nom de sentiment, etc.

Chaque FL peut être associée à une ou plusieurs expressions linguistiques qui en sont la traduction par défaut, cette traduction variant selon la nature sémantique du mot-clé. J’aurai l’occasion de revenir sur ce point à la fin de l’article (Section 4).

2.2 Liens lexicaux universaux

Une FL est avant tout une entité informationnelle correspondant à un lien lexical très général. Par *entité informationnelle*, je veux dire une entité qui n’a pas d’existence physique bien qu’elle se manifeste physiquement de façon indirecte. Elle est apprise par l’humain et, une fois apprise, devient un élément de connaissance, plus particulièrement, une partie de la connaissance linguistique de chaque personne. Le *SIGNE LINGUISTIQUE* saussurien est un exemple type d’entité informationnelle linguistique. C’est une information, une connaissance possédée par le locuteur de la langue, qui peut être observée indirectement à travers la manifestation physique (sonore ou graphique) du *SIGNIFIANT* du signe. Je rappelle que pour Ferdinand de Saussure (Saussure, 1972), le signifiant du signe n’est pas lui-même de nature physique ; c’est un patron sonore (phonémique) qui se réalise dans chaque énonciation par une suite de sons donnée.

De façon parallèle, une relation lexicale comme la synonymie est un patron général de connexion lexicale connu de tout locuteur de la langue. Toute personne parlant une langue naturelle sait, de façon consciente ou non consciente, que la synonymie existe et que toute lexie de la langue est potentiellement liée à d’autres lexies dont le sens peut être considéré comme étant *grosso modo* équivalent. Un locuteur du français sait donc non seulement qu’un lien d’équivalence sémantique approximative unit les lexies MAISON et DEMEURE ; il sait aussi que ce lien est récurrent dans le lexique. Un lien de synonymie donné n’est pas un « accident » : c’est la manifestation d’un patron standard de connexion lexicale.

Les liens lexicaux tels que la synonymie sont particulièrement intéressants du fait que ce sont des *UNIVERSAUX LINGUISTIQUES* : toute langue naturelle possède la synonymie et apprendre une langue naturelle consiste aussi à apprendre (encore une fois, de façon plus ou moins consciente) que la synonymie existe. Il est essentiel de distinguer un lien de synonymie donné, qui lie deux lexies particulières de la langue, et la synonymie elle-même, en tant qu’entité informationnelle, en tant que type de lien universellement présent dans toutes les langues.

Lorsqu’on a appris que la synonymie existe, on n’a pas seulement acquis une connaissance à propos de notre langue, comme le serait l’apprentissage d’un signe linguistique particulier : on a intégré une notion qui vaut pour toutes les langues naturelles.

Ce que je viens de dire à propos de la synonymie s'applique à d'autres liens bien connus en lexicologie, comme l'antonymie ou la conversivité. Cela s'applique aussi, de façon plus générale, à toutes les FL standard.

La notion d'*UNIVERSEL LINGUISTIQUE* est centrale dans toute théorie linguistique, de façon plus ou moins explicite. Un universel linguistique est une unité d'expression linguistique ou une propriété des énoncés qui se retrouve dans chaque langue. Sans l'identification d'universaux, il n'y a pas de théorie linguistique possible puisque une théorie linguistique est un système de notions universelles qui permet de développer des modèles particuliers pour chaque langue naturelle.

Un universel linguistique peut énoncer une propriété partagée par toutes les langues en tant que systèmes : toute langue distingue au moins les deux parties du discours nom et verbe ; une phrase standard est, dans toutes les langues, gouvernée par un élément prédicatif ; un dépendant syntaxique peut être un actant syntaxique ou un modificateur ; etc. Mais un universel linguistique peut aussi être une règle descriptive, comme le sont les règles de paraphrasage impliquées dans des équivalences du type *soupirer* = *pousser un soupir*, *regretter* = *éprouver un/du regret*, etc. Pour une présentation des règles universelles de paraphrasage, voir Mel'čuk (1988). Les FL standard, en tant que moyen de description des dérivations sémantiques et des collocations, appartiennent à la seconde catégorie d'universaux linguistiques.

Il est important de noter qu'un universel linguistique associé à un contenu sémantique particulier n'est pas nécessairement un primitif sémantique. Ainsi, on peut voir que le sémantème 'commencer', associé à la FL **Incep**

IncepOper₁(*silence*) = *s'enfoncer, tomber* [dans ART ~]

IncepFunc₀(*dialogue*) = *s'engager, s'établir, s'installer, s'instaurer, se nouer* [entre N et N]

n'appartient pas au lexique du *natural semantic metalanguage* d'Anna Wierzbicka (Wierzbicka, 1996), puisqu'il peut se décrire en terme d'autres sémantèmes plus simples. Sa définition ressemblerait à quelque chose comme 'X commence' = 'Il n'y a pas X, après il y a X'. Même si ce n'est pas un primitif sémantique, on peut tout de même postuler que 'commencer' a de bonnes chances d'être un universel linguistique, c'est-à-dire que toutes les langues naturelles doivent posséder des moyens lexicaux ou grammaticaux privilégiés pour l'exprimer.

Les FL standard postulées par la théorie Sens-Texte sont toutes des universaux linguistiques, qui ont été isolés de façon empirique à travers l'examen des différents types de liens lexicaux apparaissant de façon récurrente dans un très grand nombre de langues, de familles très diverses. Par exemple, la théorie Sens-Texte postule que dans toutes les langues naturelles on trouve un nombre important de paires de lexies L_1 et L_2 , telles que L_2 signifie *grosso modo* 'qqch./qqn. qui est le X [= le premier actant] de L_1 '. Il s'agit de la FL **S**₁ :

S₁(*faute* [Ex. : *Il a commis une faute grave.*]) = *coupable*

S₁(*ensemble*_N) = *élément*

Ma caractérisation sémantique de **S**₁ est sans doute assez maladroite et peu intuitive. Pour mieux « définir » cette FL, il faut utiliser une autre FL, syntagmatique celle-là : **Oper**₁. Le sens de **S**₁(L_1) est ainsi définissable par la formule 'qqch./qqn. qui **Oper**₁ L_1 ' :

Oper₁(*faute*) = *commettre*, donc 'coupable' = 'qqn. qui a commis une faute'

Oper₁(*ensemble*_N) = *appartenir*, donc 'élément' = 'qqch. qui appartient à un ensemble'

Cette présentation est bien entendu une simplification, qui laisse de côté de nombreux problèmes. Par exemple, dans le cas d'une lexie L_1 dénotant un artefact, $S_1(L_1)$ se paraphrase via la FL **Real**₁, et non **Oper**₁ :

$S_1(\text{véhicule}) = \text{conducteur}$

Real₁(*véhicule*) = *conduire*, donc 'conducteur' = 'qqn. qui conduit un véhicule'.

D'autres problèmes demanderaient à être examinés pour faire un traitement véritablement rigoureux de la spécification sémantique des FL. Par exemple, on voit que $S_1(\text{voiture})$ est aussi *conducteur*, mais il serait trop restrictif de définir 'conducteur' par 'qqn. qui conduit une voiture'.

Résumons maintenant ce qui vient d'être dit. Toutes les FL sont des universaux linguistiques, dénotant des liens lexicaux récurrents. Elles sont, de plus, toutes associables à un contenu sémantique donné de nature très vague et qui n'est pas nécessairement un primitif sémantique, au sens où l'entend par exemple Anna Wierzbicka. Nous allons maintenant voir que les FL doivent en fin de compte être considérées comme étant des « métalexies ».

2.3 Les FL en tant que métalexies

Imaginons une situation de parole « standard ». Que s'est-il passé dans l'esprit d'un locuteur francophone qui a produit la phrase suivante ?

(1) *Igor a poussé un gros soupir.*

L'énonciateur de la phrase ci-dessus a voulu dire quelque chose à propos d'un individu donné et a identifié ce qui était pour lui la meilleure façon d'y référer, dans le présent contexte d'énonciation. Il aurait pu dire *Igor Mel'čuk*, ou *Mel'čuk*, ou *le roi du chachlik*, mais il a préféré désigner cette personne par son prénom et a donc fait usage de l'information suivante : « la personne dont je veux parler peut être désignée au moyen de la lexie IGOR ». Même chose pour le fait dont il est question. Il s'agissait de parler d'un soupir et, entre le nom SOUPIR et le verbe SOUPIRER, l'énonciateur de (1) a sélectionné le premier, sachant du même coup qu'il se devait de trouver un verbe pratiquement vide de sens qui permettrait de « faire tenir » syntaxiquement la phrase. Je veux dire par là que l'énonciateur n'a sûrement pas voulu dire que Igor a poussé quelque chose. Le choix de la lexie POUSSER répond ici à un besoin très particulier, celui de combler un vide syntaxique, puisque toute phrase française complète doit être gouvernée par un verbe. On peut donc considérer que l'énonciateur a choisi POUSSER parce qu'il avait besoin d'une lexie ayant au moins les trois propriétés suivantes :

1. elle est verbale ;
2. elle se combine avec SOUPIR sans ajouter de sens à cette dernière mais plutôt en opérant ainsi une translation nom→verbe, au sens de Tesnière (1959), puisque *pousser un soupir* est équivalent à *soupirer* ;
3. elle permet d'effectuer la translation en question en connectant syntaxiquement SOUPIR à son premier actant de la façon la plus directe possible et en donnant à ce premier actant la position communicative type du thème et à SOUPIR la position communicative type du rhème :

*Igor*_[sujet grammatical, thème] *a poussé un gros soupir*_[CObjet, rhème].

L'énonciateur n'a donc pas accédé directement à POUSSER. Il a accédé à cette lexie par l'intermédiaire d'un patron très général de propriétés, qui ne peut trouver son expression linguistique que de façon conditionnelle au choix d'une autre lexie devant apparaître dans la phrase. On peut caricaturer de la façon suivante ce qui distingue le processus de choix des lexies SOUPIR et POUSSER dans mon exemple :

1. Besoin : expression du sémantème 'soupir'
→ SOUPIR
2. Besoin : support syntaxique verbal vide pour SOUPIR de type $X_{\text{thème}} + \text{SOUPIR}_{\text{rhème}}$
→ POUSSER

C'est cette différence dans la procédure de choix — choix direct vs choix conditionnel à un autre choix — qui justifie la modélisation des universaux linguistiques qui nous intéressent ici en tant que fonctions (lexicales). Pour une présentation des FL mettant l'accent sur la nature particulière de la procédure de choix lexical qu'elles impliquent, voir Mel'čuk (1995 ; 1998).

Je ne vais pas développer plus cette notion de choix lexical, laquelle apparaît dans la définition même de la collocation telle que définie dans la théorie Sens-Texte. Le lecteur intéressé pourra trouver une proposition de modèle de lexicalisation tenant compte de ce type de phénomènes dans Polguère (1998), en français, ou Polguère (2000a), en anglais. L'idée centrale que je voudrais mettre en avant ici est qu'une FL *f* est bien une entité en soi. C'est une information autonome codée dans la mémoire du locuteur et qui lui sert à accéder à un petit ensemble de lexies données, retourné par l'application *f*(L).

On peut ainsi dire que la clé d'accès à POUSSER était une *MÉTALEXIE*, que nous appellerons **Oper₁** — **Oper₁**(*soupir*) = *laisser échapper, pousser* [ART ~] — et qui peut être décrite de la façon suivante :

| | |
|--------------------------|---|
| Oper₁ | |
| Sens | : Ø |
| Combinatoire | : verbe, FL à argument L nominal, sujet = actant 1 de L, complément d'objet = L |
| Signes lexicaux associés | : voir les signes lexicaux associés à Oper₁ (L) |

La description donnée ci-dessus est certainement très grossière. Elle nous indique que **Oper₁** n'est pas associée à un sémantème ; cette FL est vide de sens. Son « contenu » se ramène donc à sa combinatoire, qui va justifier son utilisation. On sait que toute lexie est associée à un ensemble de signes linguistiques, qui sont en fait les formes fléchies de la lexie. Par exemple, SOUPIR est associée au signe singulier *soupir* et au signe pluriel *soupirs*¹. Dans le cas d'une FL, on n'a bien entendu pas de signes lexicaux associés. Ceux-ci ne sont accessibles qu'une fois effectuée l'application *f*(L) et une fois opéré le choix d'une lexie particulière permettant d'exprimer la FL *f*.

Une FL *f* donnée peut donc être conceptualisée comme une entité linguistique à part, une *MÉTALEXIE*, existant dans l'esprit du locuteur de façon indépendante des multiples applications possibles *f*(L) de cette fonction². La particularité des FL en tant que métalexies est qu'elles n'ont pas de signes lexicaux associés, donc pas de signifiants. La lexie POUSSER est associée aux signifiants *pousser, poussé, poussant, [je] pousse, [tu] pousses*, etc. Mais la métalexie **Oper₁** n'a

1 Pour une présentation de la notion de signe linguistique Sens-Texte, qui est une extension du signe saussurien, voir Mel'čuk (1982 :40-41 ; 1993 :109-133).

pas de signifiants linguistiques propres associés ; ses signifiants sont ceux obtenus par l'application de **Oper₁** à des lexies nominales L données. Bien entendu, une FL considérée en tant qu'élément d'un langage formel possède un signifiant non linguistique, qui est le nom que l'on utilise pour la désigner : « **Syn** », « **Oper₁** », etc.

Ce qui vient d'être dit ici à propos de **Oper₁** et de la façon dont l'énonciateur de (1) a choisi la lexie **POUSSER**, s'applique de façon peut-être encore plus convaincante à la FL **Magn** et au choix du collocatif **GROS** dans l'expression *gros soupir*. En effet, contrairement à **Oper₁**, **Magn** peut être directement associée à un sémantème et sa description en tant que métalexie semble donc encore plus naturelle :

| | |
|--------------------------|--|
| Magn | |
| Sens | : 'intense/très/beaucoup/...' |
| Combinatoire | : FL à argument L, modificateur de L |
| Signes lexicaux associés | : voir les signes lexicaux associés à Magn(L) |

Comme on le voit, **Magn** est tellement vague sémantiquement que l'on doit en fait considérer qu'elle est associée en bloc à une famille de sémantèmes (d'intensification) plutôt qu'à un sémantème unique.

Je récapitule maintenant ce qui a été dit dans cette section. Pour être bien comprises, les FL doivent avant tout être appréhendées comme des métalexies : elles sont caractérisées par des propriétés de sens et de combinatoire, mais ne sont pas associées de façon directe à des signes, et donc à des signifiants. L'association à des signifiants particuliers se fait, de façon indirecte, par l'application de la métalexie (en tant que fonction) à une lexie particulière, préalablement identifiée par le locuteur. Ces métalexies universelles devraient être enseignées comme telles. La FL **Magn** existe, même si c'est une notion « abstraite ». Elle existe au même titre que la notion de partie du discours, ou de genre grammatical. Si elle est enseignée de la bonne façon, il en résultera nécessairement une meilleure compréhension du fonctionnement de la langue pour qui maîtrisera cette notion de façon consciente. Par exemple, si je suis conscient de phénomènes tels que **Oper₁**, je sais qu'il ne me suffit pas d'apprendre la traduction d'un nom prédicatif dans la langue que j'étudie. Je dois aussi apprendre alors quelle est la bonne façon d'en « dériver » une expression verbale. Si j'apprends la lexie anglaise **SIGH_N** 'soupir', je dois aussi apprendre que l'on dit *to breathe/heave/let out a sigh* ; si j'apprends **BURP_N**, je dois aussi apprendre que l'on dit soit *to let out* soit *to make a burp*. (Comparer avec **faire/laisser échapper/pousser un soupir* vs *faire/laisser échapper/*pousser un rot*.)

Pour que ma démonstration de la nature métalexicale des FL soit véritablement convaincante, il me faudrait couvrir bien plus que les deux seules FL **Magn** et **Oper₁**. Il faudrait notamment démontrer que la présentation des FL en tant que métalexies s'applique aussi aux FL paradigmatiques (**Syn**, **Anti**, **S₀**, etc.). Je dois malheureusement arrêter là cette tentative de caractérisation

2 Il me semble important d'utiliser le terme *métalexie* et non *pseudolexie* car la FL n'est pas « moins » qu'une lexie. C'est une entité englobante [= méta] vis-à-vis des lexies, qui synthétise un ensemble de propriétés servant de clé d'accès à des ensembles spécifiques de lexies ou expressions linguistiques (les applications **f(L)**). Pour bien comprendre l'usage qui est fait ici de ce terme, il faut aussi garder à l'esprit que, selon moi, une métalexie de ce type est un élément de connaissance du locuteur de toute langue, et non une simple construction théorique ne fonctionnant que dans le cadre restreint de la théorie Sens-Texte. En s'inspirant d'un concept de la grammaire générative, on pourrait dire que les FL standard font partie du « lexique universel » (cf. la grammaire universelle). Qu'il soit de nature innée ou acquise, ce lexique est possédé par tout être humain parlant au moins une langue.

brute, non formelle, de la notion de FL et passer au second volet de mon exposé : le problème de l'enseignement des FL.

3 Comment enseigner les FL ?

Je vais me limiter dans cette section à deux aspects de la problématique de l'enseignement des FL :

1. la stratégie globale d'enseignement, qui doit introduire les FL en tant que système plutôt que sous la forme d'une liste de notions ;
2. l'identification des notions linguistiques de base dont la compréhension est présupposée par la notion de FL.

Je rappelle (voir ce qui a été dit à la fin de la Section 1) que le public visé ici est essentiellement celui des spécialistes ou futurs spécialistes de l'étude et de l'enseignement des langues, qui étudient la notion de FL pour mieux comprendre comment fonctionnent les systèmes linguistiques. Les stratégies dont il va être question ne sont vraisemblablement pas applicables lorsqu'il s'agit non d'enseigner la notion de FL, mais de l'utiliser pour enseigner une langue donnée.

3.1 Enseigner les FL comme un système et non comme une liste de notions

Nous avons vu dans la Section 2.1 que les FL sont souvent introduites dans la littérature de façon très linéaire, comme une liste de notions. Cette façon de procéder se heurte à deux problèmes sérieux.

Tout d'abord, du fait du très grand nombre de combinaisons formelles possibles, l'ensemble des FL s'assimile difficilement de façon linéaire. Il faut présenter une structuration des FL, un véritable système où les connexions entre FL sont explicitées. Il faut aussi que la transition vers une méthode d'encodage se fasse de façon très intuitive. Une vague caractérisation des propriétés sémantiques et formelles de chaque FL ne peut suffire. La nécessité de mettre en évidence l'organisation interne du système des FL a déjà été soulignée dans la littérature Sens-Texte — voir notamment Grimes (1990) et Alonso Ramos & Tutin (1996). Mais une stratégie « systémique » d'enseignement des FL reste encore à développer. Elle devrait s'appuyer sur une claire identification de la nature linguistique profonde des FL. Pour cela, je pense que la notion de métalexie reste une bonne piste. Une stratégie d'enseignement devra aussi prendre en compte de façon élégante et pédagogique les liens « incestueux » qu'entretiennent les deux groupes de FL, paradigmatiques et syntagmatiques, notamment à travers les notions d'équivalence paraphrastique et de fusion (symbolisée dans le langage formel des FL par l'opérateur //). Je fais ici référence à des phénomènes tels que :

- $L_V \equiv \text{Oper}_i(\text{S}_0(L_V)) + \text{S}_0(L_V)$ [Ex. : *soupirer* \equiv *pousser un soupir*]
- Un **Magn**(L) fusionné est un **Syn** \supset (L), un synonyme plus riche de L — voir par exemple l'article de dictionnaire de VICTOIRE, où deux pointeurs vers TRIOMPHE peuvent être introduits

Syn \supset : *triomphe*
Magn : *grande* | antépos; *claire, complète, éclatante, écrasante, retentissante* // *triomphe*

Je rappelle que l'interconnexion entre FL paradigmatiques et syntagmatiques a aussi été démontrée au tout début de cet article, lors de l'examen du lien unissant WAGON à LOCOMOTIVE (Section 1). Dans un contexte d'enseignement des FL, il faut donc absolument rendre compte du fait que la notion de FL permet de faire apparaître, de façon très naturelle, un lien logique entre le phénomène de *DÉRIVATION SÉMANTIQUE* (FL paradigmatiques) et celui de *COLLOCATION* (FL syntagmatiques) :

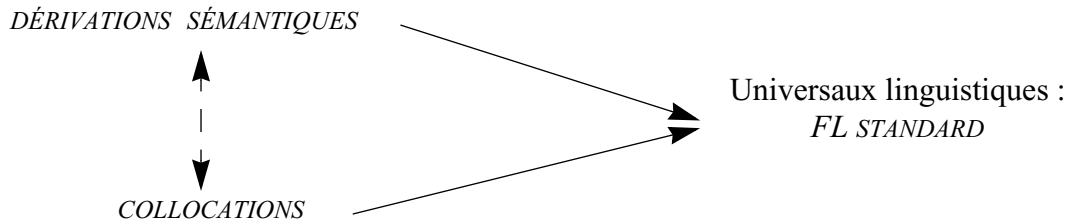


Figure 2 Liens notionnels rendus possibles par le recours aux FL

Le second problème auquel se heurte une présentation « à plat » des FL vient de la **diversité** des liens à modéliser et de la fréquente présence de liens peu récurrents. Si l'on part des données linguistiques elles-mêmes, on constate dans la pratique que la plupart des articles des dictionnaires qui encodent les liens lexicaux au moyen des FL vont contenir un nombre plus ou moins élevé de liens lexicaux standard mais rares. Par exemple, pour la lexie BATEAU :

| | |
|---|---|
| FinAble₁Fact₀ | : (s') <i>échouer</i> [<i>sur</i> N] [Ex. : <i>Le bateau s'est échoué <a échoué> sur les récifs.</i>] |
| S₀FinAble₁Fact₀ | : <i>échouage</i> |
| CausFinAble₁Fact₀ | : <i>échouer</i> [ART ~] [Ex. : <i>Ils ont échoué leur bateau.</i>] |
| LiquFinAble₁Fact₀ | : <i>remettre à flot, renflouer</i> [ART ~] |

Plus fréquemment encore, on trouvera des liens lexicaux totalement ou partiellement non standard. Voici, toujours pour BATEAU, une partie seulement des liens de FL de ce type :

| | |
|---|--|
| Qui Fact₀ grâce à N | : <i>à</i> [N] [Ex. : <i>à moteur/rames/vapeur/voile</i>] [Noter la combinaison FL non standard et FL standard !] |
| Qui sert à la pêche | : <i>de pêche</i> |
| Qui sert à faire des croisières | : <i>de croisière</i> |
| Qui a un usage non commercial | : <i>de plaisance</i> |
| Qui sert à transporter Y d'un côté à l'autre de l'Atlantique | : <i>transatlantique</i> _{Adj} // <i>transatlantique</i> _N |
| Qui est utilisé par des pirates | : <i>pirate</i> _{Adj} |
| Ce qui reste d'un B. après FinFunc₀ | : <i>épave</i> [un autre exemple de combinaison de FL non standard et standard] |
| B. oscille | : <i>se balancer, tanguer</i> |
| B. se retourne sur lui-même | : <i>chavirer</i> |
| Quantité de marchandise qu'un B. peut transporter | : <i>spéc tonnage</i> [<i>de</i> ART ~]; <i>capacité</i> [<i>de</i> ART ~] |
| Mesure de la taille d'un B. | : <i>spéc calaison, spéc tirant d'eau</i> [<i>de</i> ART ~] |

Il ne faut pas sous-estimer l'importance des liens non standard : ils sont omniprésents. Bien sûr, ils gênent les esprits formalisateurs et ceux qui voudraient implanter des logiciels de traitement automatique des données lexicales ; mais aussi, et surtout, ils viennent parasiter l'apprentissage du système des FL standard : lorsqu'il est confronté aux données, l'apprenti ne sait pas si le lien qu'il doit encoder est standard ou (partiellement) non standard. Seule la parfaite maîtrise du système des FL standard va lui permettre d'établir le bon diagnostic. En gros, un lien qu'on ne peut encoder par des FL standard est non standard. On voit donc qu'en phase d'apprentissage, il ne faut pas exiger de l'apprenti qu'il formalise toutes les données rencontrées. Il doit formaliser celles qui correspondent aux quelques FL dont il a la maîtrise et se contenter d'établir une *PARAPHRASE*, comme pour une FL non standard, pour tous les autres liens qu'il rencontre, l'apprentissage se faisant par couches successives³.

On notera que ce qui vient d'être dit présuppose que l'apprentissage des FL ne peut se faire sans une pratique intensive de l'encodage, de la description, des liens lexicaux. On ne peut parvenir à une bonne compréhension du phénomène en question sans être exposé à un très grand nombre de cas de figure, aussi variés que possible. Au cours de ce processus de formation, c'est la graduelle réapparition de cas de figure présentant un certain caractère de ressemblance qui va faire prendre conscience à l'apprenti de l'existence des FL en tant que phénomène linguistique.

Pour résumer, il faut donc une stratégie générale d'enseignement des FL qui

- vise l'apprentissage d'un système de notions connectées,
- soit progressive,
- reconnaisse dès le départ l'importance du rôle joué par les liens lexicaux non standard,
- soit orientée vers la pratique de la description des liens lexicaux.

Mais toute stratégie d'enseignement des FL, aussi pédagogique et élaborée soit-elle, échouera si l'on ne prend pas soin d'enseigner d'abord quelques notions linguistiques de base.

3.2 Notions présupposées

Je n'ai pas de liste exhaustive de notions préliminaires à proposer. Je me contenterai d'une première version d'une telle liste. Elle est fondée, bien entendu, sur la reconnaissance de nécessités logiques. Par exemple, comment parler de liens entre lexies sans conceptualiser au préalable la notion de lexie elle-même ? Cette liste est aussi fondée sur mon expérience de l'enseignement des FL, qui m'a souvent forcé à relativiser certaines idées préconçues que j'avais à propos de la stratégie à adopter pour inculquer une bonne compréhension des FL.

L'énumération ci-dessous suit plus ou moins un ordre chronologique (et logique) dans l'apprentissage. Il ne s'agit nullement d'aller des notions les plus importantes aux notions les moins importantes, ou *vice versa*.

LEXIE ou **UNITÉ LEXICALE** Cette notion est le point de départ logique de la description des liens lexicaux ; noter qu'il faut à tout prix éviter d'y référer en utilisant le terme trop ambigu de *mot*. Il faudra aussi introduire la notion de *VOCABLE* (« mot polysémique »), qui désigne un regroupement de lexies. Comparer les acceptions (les lexies) d'un même vocable sous l'angle des

3 Pour une étude du rapport entre FL standard et non standard, voir Polguère (à paraître).

liens de dérivation sémantique et des phénomènes collocationnels est d'ailleurs un exercice particulièrement éclairant pour ce qui est de la compréhension de l'importance des FL.

SENS La plupart des FL se caractérisent avant toute chose par la nature du lien sémantique qu'elles modélisent. Il est donc fondamental de présenter la notion de *SENS* en tant que propriété centrale des énoncés et des éléments dont ils sont constitués.

PARAPHRASE et DÉFINITION La meilleure façon d'explicitier le sens d'un élément linguistique, et notamment d'une lexie, est d'avoir recours à la *PARAPHRASE*. C'est non seulement ainsi que procèdent les dictionnaires et bon nombre de théories sémantiques, mais c'est aussi la façon naturelle de procéder dans la vie courante, lorsque une personne vous demande de lui expliquer le « sens d'un mot ». Pour garder une certaine rigueur dans l'utilisation de la notion de paraphrase, il faut absolument avoir recours, de façon explicite ou implicite, à celle de *DÉFINITION PAR GENRE PROCHAIN ET DIFFÉRENCES SPÉCIFIQUES*. Toutes les paraphrases du sens d'une lexie devraient être construites selon ce modèle, car c'est celui qui permet le mieux d'élaborer des descriptions véritablement falsifiables et donc testables.

PRÉDICAT et ACTANT L'expérience montre que ces deux notions sont absolument incontournables si l'on veut parvenir à une bonne maîtrise des FL, même sous une forme vulgarisée. Bon nombre de phénomènes de dérivation sémantique (S_i , A_i , etc.) et de collocation (**Oper_i**, **Func_i**, etc.) ne se comprennent qu'en référence à la notion d'*ACTANT* et, donc, de *PRÉDICAT SÉMANTIQUE*. Je suis bien entendu conscient du fait qu'il n'est pas facile de vulgariser de telles notions, mais il faut le faire parce qu'il serait en revanche **impossible** de comprendre sans elles les FL. De plus, introduire ces deux notions est un excellent investissement qui permet de mieux appréhender d'autres notions linguistiques essentielles, notamment celles de sens, de fonction syntaxique et de partie du discours.

COMBINATOIRE La théorie Sens-Texte a ajouté à la notion saussurienne de signe linguistique, association entre un signifiant et un signifié, un composant appelé *SYNTAXIQUE*, qui correspond à la *COMBINATOIRE* propre au signe — voir Mel'čuk (1982 :40-41 ; 1993 :109-133), déjà mentionnés dans la Section 2.3. La notion de signe linguistique, qui est tellement fondamentale, s'avère cependant très difficile à vulgariser de façon satisfaisante. À moins de présenter le signe linguistique en utilisant une langue presque dénuée de phénomènes morphologiques (comme l'est le mandarin), on se trouve obligé de s'aventurer dans un réseau fort complexe d'autres notions (racine, affixe, flexion, etc.). Je ne recommande donc pas d'introduire le signe linguistique dans un contexte autre que l'étude avancée de la langue. Il faut cependant faire référence à la *COMBINATOIRE*, puisque, en plus du sens et des formes associées, c'est une des propriétés fondamentales des lexies, dont justement les FL permettent de rendre compte. La notion de collocation (voir plus bas) est ainsi indissociable de celle de (propriétés de) combinatoire.

FONCTION SYNTAXIQUE Il faut un minimum de connaissances de la structure syntaxique des phrases pour avoir quelques chances de comprendre la notion de FL. De ce point de vue, savoir identifier la *FONCTION SYNTAXIQUE* de chacun des éléments de la phrase est un prérequis pour aborder l'étude des FL. J'aimerais aussi inclure ici la notion de *DÉPENDANCE SYNTAXIQUE* puisqu'elle est de toute façon enchâssée de façon implicite dans celle de fonction syntaxique telle qu'enseignée en grammaire traditionnelle.

IDIOMATICITÉ, COLLOCATION et LOCUTION C'est peut-être en appliquant les FL à la modélisation des *COLLOCATIONS* que l'on perçoit le mieux leur importance. Je pense cependant qu'il convient de partir de la notion plus générale d'*IDIOMATICITÉ*, puisque celle-ci est la clé d'accès

non seulement à la notion de collocation, mais aussi à celle de *LOCUTION* (ou *PHRASÈME COMPLET*). Il faut se rappeler que l’idiomaticité est un phénomène graduel, qui devra être présenté comme tel. Le caractère graduel de l’idiomaticité est illustré par la série d’exemples ci-dessous — de la plus faible à la plus forte idiomaticité :

- (2) *Sylvain a levé la charge avec un treuil.*
Expression libre.
- (3) *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 (*QUASI-PHRASÈME*) dans la théorie Sens-Texte.
- (4) *Sylvain a levé la séance.*
Collocation, c’est-à-dire expression linguistique semi-idiomatique non lexicalisée, encodable dans le langage formel des FL par la formule $\text{LiquFunc}_0(\textit{séance})+\textit{séance}$. (Les collocations sont aussi appelées *SEMI-PHRASÈMES*.)
- (5) *Sylvain travaille trop. Il devrait lever le pied un peu.*
Expression figée signifiant ‘ralentir, en faire moins’, qui est une lexie (*PHRASÈME COMPLET*) du français.

Il est intéressant de remarquer que, contrairement à ce qui est souvent affirmé, la transition du moins au plus idiomatique ne se fait pas de façon strictement parallèle à la transition du moins au plus lexical. Les quasi-phrasèmes sont moins idiomatiques que ne le sont généralement les collocations (puisque leur sens incorporent le sens de toutes les lexies qui les constituent), mais ce sont bien des unités lexicales.

DÉRIVATION SÉMANTIQUE La *DÉRIVATION SÉMANTIQUE* peut être introduite avant ou après la notion de collocation. Je n’ai pas vraiment d’opinion à ce sujet. Ce dont je suis convaincu cependant, c’est qu’il est préférable qu’elle ait été présentée avant les FL elles-mêmes. Ce point de vue découle logiquement de la nécessité de présenter les FL comme un système et non de façon linéaire (cf. Section 3.1).

LIEN LEXICAL Une fois introduites toutes les notions précédentes, on peut enfin entrer dans le vif du sujet et introduire les notions de *LIEN LEXICAL* et de *FL*. L’apprentissage de ces notions se fera d’autant plus vite que les notions précédentes auront été bien intégrées. On constate souvent que la courbe d’apprentissage des FL est abrupte au départ puis stagne très vite, les performances dans la capacité d’encoder de nouveaux phénomènes allant même en se dégradant au bout d’un moment. Cela est dû en grande partie à une mauvaise préparation du « terrain notionnel ». Les premiers exercices d’encodage de liens (collocationnels notamment, avec des **Magn**, **Oper_i**, **Real_i**, etc.) se passent très bien et enthousiasment l’étudiant, qui a vraiment l’impression d’apprendre quelque chose d’utile. Cependant, la langue est un système incroyablement riche et complexe ; les cas simples et bien tranchés de liens lexicaux sont somme toute en minorité. De plus, un lien lexical difficile à encoder dans l’article de dictionnaire d’une lexie peut tout à fait être perçu comme étant beaucoup plus « important » qu’un autre lien très courant et évident. Par exemple, il est moyennement utile d’indiquer que GRAND, GROS et ÉNORME sont des **Magn** de BATEAU, alors qu’il paraît très intéressant de mentionner que cette lexie possède un **FinAble₁Fact₀** : S’ÉCHOUER.

VARIATION INDIVIDUELLE ET DIALECTALE L'expérience montre que l'apprentissage des FL, qui doit se faire comme je l'ai dit sur la base d'une pratique intensive de l'encodage de liens lexicaux, se heurte rapidement au problème de la **VARIATION INDIVIDUELLE ET DIALECTALE**. Je sais qu'il s'agit d'une problématique qui n'est pas vraiment en odeur de sainteté dans le cadre de la théorie Sens-Texte, où il est courant de faire l'autruche face aux problèmes posés par la variation. Cependant, ceux qui rédigent des articles de dictionnaire et les étudiants à qui on veut apprendre à le faire, ou, plus modestement, à qui on veut faire comprendre la notion de FL, sont des êtres vivants et non des « locuteurs idéalisés ». Leur intuition linguistique va sans cesse se trouver être en conflit avec celle d'autres locuteurs, avec des données trouvées dans les corpus ou avec les faits de langue qu'ils pourront directement observer. Il est bien entendu nécessaire de continuer de parler de faute linguistique ou d'énoncé incorrect, mais il faut aussi accepter la réalité de la variation individuelle et dialectale. Ces notions doivent donc être abordées de front. Il convient de faire prendre conscience de ce phénomène aux personnes à qui l'on enseigne les FL. Il faut aussi absolument élaborer de bonnes stratégies pour justement tirer avantage des FL afin de mieux rendre compte de ces phénomènes. Pour un pas timide dans cette direction, voir par exemple Lim & Polguère (1992).

Ceci termine ma rapide présentation des notions de base présupposées par les FL. Je vais maintenant conclure en mentionnant quelques points importants liés au problème de l'encodage des liens de FL.

4 Conclusion : vers un encodage paraphrastique des liens de FL

Il a été montré dans Kahane & Polguère (2001) que le formalisme traditionnellement utilisé pour encoder les FL doit être amélioré. Il est défini d'une façon trop floue, sans la rigueur qui convient à un véritable langage formel adapté, notamment, au traitement automatique. Mais ce n'est pas tout. Kahane & Polguère (2001) mentionne aussi le fait qu'il faut disposer d'autres types d'encodages, adaptés à différents contextes d'utilisation des FL. C'est un peu le travail qui a été commencé dans le cadre de la rédaction (toujours en cours) de la base de données *DiCo* et du dictionnaire vulgarisé *Lexique Actif du Français* (Polguère, 2000b ; 2000c). (Pour un projet lexicographique ayant une finalité proche de celle du *Lexique Actif du Français*, voir Verlinde *et al.* dans le présent volume.)

Dans un contexte d'enseignement des FL, il faut être capable d'encoder les liens de FL en utilisant une sorte de langue « contrôlée », dans laquelle on peut paraphraser de tels liens. En effet, comment un lexicographe du *DICIONNAIRE EXPLICATIF ET COMBINATOIRE* (Mel'čuk *et al.*, 1984 ; 1988; 1992; 1999) encode-t-il le lien unissant BATEAU à S'ÉCHOUER ? Il va tout simplement tenter de paraphraser [*Le bateau*] *s'échoue*. Une bonne paraphrase approximative serait :

(6) *Le bateau se retrouve immobilisé.*

Puis, en fouillant dans son sac à universaux sémantiques, le lexicographe va regarder s'il n'est pas possible de normaliser cette paraphrase. Ainsi, qu'est-ce que cela veut dire pour un bateau que de se retrouver immobilisé ? Cela signifie que le bateau cesse, = **Fin** !, d'être capable, = **Able₁** !, de naviguer. Tout cela s'annonce très bien, mais peut-on paraphraser *naviguer* pour obtenir un encodage entièrement basé sur les FL standard ? Oui, car *naviguer* c'est « fonctionner », = **Fact₀** !, pour un bateau. Le lexicographe peut donc substituer la paraphrase suivante à (6) :

(7) *Le bateau arrête d'être capable de fonctionner.*

Cela permet de dégager la formule de FL **FinAble₁Fact₀**. On notera que cette formule peut être débattue et que j'ai notamment pris soin d'ignorer le fait que S'ÉCHOUER possède un second actant, qui n'est pas directement connectable à BATEAU (*s'échouer sur N*) et qui peut donc éventuellement poser un problème au niveau de l'encodage en terme de FL. De plus, on pourrait vouloir considérer que le fait de s'échouer est avant tout, pour un bateau, une sorte d'accident : l'opposé d'un fonctionnement normal, donc un **AntiFact₀**. Mais tout cela se ramène à une discussion de la justesse de la paraphrase en langue naturelle initialement proposée. C'est cette paraphrase qui doit être affinée si l'on n'est pas satisfait par notre description, et une nouvelle modélisation en terme de FL en découlera, suivant le même processus de « traduction » que celui qui vient d'être décrit.

Ce processus met en évidence deux faits importants :

1. Encoder un lien de FL consiste à paraphraser ce lien, que ce soit au moyen d'un véritable langage formel ou au moyen d'une langue contrôlée fondée sur les universaux sémantiques associés aux FL.
2. Dans le cas de $f(L)$ qui sont des lexies pleines (comme ici S'ÉCHOUER), la modélisation du lien revient en fait à faire une ébauche de définition lexicographique de la lexie liée à L.

Il me semble donc, en conclusion, que le problème le plus urgent à résoudre, pour ce qui est de la diffusion de la notion de FL et pour une amélioration de l'utilisation qui en est faite, est la mise au point de bonnes techniques de paraphrasage sémantique, utilisant des langues (français, anglais, etc.) contrôlées. Les formules en langue contrôlée — qui pourraient ressembler aux paraphrases utilisées dans le *Lexique Actif du Français* (mentionné au début de la présente section) — devront être structurées directement à partir du système des FL. Cela présuppose bien entendu que ce système aura été clairement mis en évidence au préalable.

Remerciements

Merci infiniment à Francis Grossmann, Sylvain Kahane, Igor Mel'čuk, Ophélie Tremblay et Agnès Tutin pour leurs commentaires, qui m'ont permis de corriger un nombre impressionnant d'erreurs et d'imprécisions. Je m'excuse par avance auprès d'eux si leurs remarques ont en même temps soulevé plus de problèmes que je n'avais la possibilité d'en résoudre dans le cadre de ce court texte.

Bibliographie

- Alonso Ramos M. & Tutin A., "A Classification and Description of Lexical Functions for the Analysis of their Combinations" dans Wanner L. éd., *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam/Philadelphie, Benjamins, 1996, p. 147-167.
- Grimes J., "Inverse Lexical Functions", dans Steele J. éd., *Meaning-Text Theory: Linguistics, Lexicography and Implications*, Ottawa et al., University of Ottawa Press, 1990, p. 350-364.
- Kahane S. & Polguère A., "Formal foundation of lexical functions", dans *Proceedings of COLLOCATION: Computational Extraction, Analysis and Exploitation*, 39th Annual Meeting and 10th Conference of the European Chapter of the Association for Computational Linguistics, Toulouse, 2001, p. 8-15.

- Lecerf Y., “ Programme des conflits, modèle des conflits (2^e partie) ”, *La Traduction Automatique*, 5, 1960, p. 17-36.
- Lim G. & Polguère A., “ Can Lexical Functions Help to Characterize Singapore English? ”, dans Pakir A. réd., *Words in a Cultural Context*, Singapour, UniPress, 1992, p. 154-171.
- Mel’čuk I., *Towards a language of linguistics: A system of formal notions for theoretical morphology*, Munich, Wilhelm Fink Verlag, 1982.
- Mel’čuk I., “ Paraphrase et lexique dans la théorie linguistique Sens-Texte — Vingt ans après ”, *Revue internationale de lexicologie et lexicographie*, 52 et 53, 1988, p. 5-50 et p. 5-53.
- Mel’čuk I., *Cours de morphologie générale*, Volume 1, Montréal/Paris, Les Presses de l’Université de Montréal/CNRS Éditions, 1993.
- Mel’čuk I., “ Phrasemes in Language and Phraseology in Linguistics ”, dans Everaert M., van der Linden E.-J., Schenk A. & Schreuder R. réd., *Idioms. Structural and Psychological Perspectives*, Hillsdale, NJ/Hove, UK, Lawrence Erlbaum Associates, 1995, p. 167-232.
- Mel’čuk I., “ Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon ”, dans Wanner L. réd., *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam/Philadelphie, Benjamins, 1996, p. 37-102.
- Mel’čuk I., “ Collocations and Lexical Functions ”, dans Cowie A. réd., *Phraseology. Theory, Analysis, and Applications*, Oxford, Clarendon Press, 1998, p. 23-53.
- Mel’čuk I., Clas A. & Polguère A., *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve, AUPELF-UREF/Duculot, 1995.
- Mel’čuk I. et al., *Dictionnaire explicatif et combinatoire du français contemporain: Recherches lexico-sémantiques I, II, III, IV*. Montréal, Presses de l’Université de Montréal, 1984, 1988, 1992, 1999.
- Polguère A., “ Pour un modèle stratifié de la lexicalisation en génération de texte ”, *Traitement Automatique des Langues (T.A.L.)*, 39 :2, 1998, p. 57-76.
- Polguère A., “ A “natural” lexicalization model for language generation ”, *Proceedings of the Fourth Symposium on Natural Language Processing (SNLP’2000)*, Chiangmai, 2000a, p. 37-50.
- Polguère A., “ Une base de données lexicales du français et ses applications possibles en didactique ”, *Revue de Linguistique et de Didactique des Langues (LIDIL)*, 21, 2000b, p. 75-97.
- Polguère A., “ Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French ”, *Proceedings of EURALEX’2000*, Stuttgart, 2000c, p. 517-527.
- Polguère A., “ Le sens linguistique peut-il être visualisé ? ”, dans Lagorgette D. & Larrivée P. réd., *Représentations du sens linguistique*, Munich, Lincom Europa (Lincom Studies in Theoretical Linguistics 25), 2002, p. 89-103.
- Polguère A., “ Lexical function standardness ”, dans Wanner L. réd., *Festschrift in Honour of Igor Mel’čuk*, Amsterdam/Philadelphie, Benjamins, à paraître.
- Reinhart T., *Anaphora and semantic interpretation*, Londres, Croom Helm, 1983.
- Saussure F. de, *Cours de linguistique générale*, Paris, Payot, 1972.
- Tesnière L., *Éléments de syntaxe structurale*, Paris, Klincksieck, 1959.
- Wiezbicka A., *Semantics: Primes and Universals*, Oxford, Oxford University Press, 1996.