



**HAL**  
open science

# Les apports possibles de la phraséologie à la didactique des langues étrangères

Mojca Pecman

► **To cite this version:**

Mojca Pecman. Les apports possibles de la phraséologie à la didactique des langues étrangères. ALSIC - Apprentissage des Langues et Systèmes d'Information et de Communication, 2005, 08 (1), pp.109-122. edutice-00109615

**HAL Id: edutice-00109615**

**<https://edutice.hal.science/edutice-00109615>**

Submitted on 25 Oct 2006

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

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

## Les apports possibles de la phraséologie à la didactique des langues étrangères

Mojca PECMAN  
Université de Nice-Sophia Antipolis, France

**Résumé :** *L'article explore les avantages d'une analyse phraséologique du discours scientifique face aux difficultés de la maîtrise des formules conventionnelles de ce genre dans une langue étrangère, en l'occurrence en anglais. Notre recherche postule en effet l'existence d'une langue commune aux scientifiques, une langue qui transcende les disciplines, et désormais, grâce à l'anglais, les continents. Pour observer les caractéristiques de cette Langue Scientifique Générale un corpus parallèle a été mis en place recueillant des textes de trois domaines connexes : la chimie, la physique et la biologie. Le dépouillement de ce corpus a abouti à une base de données phraséologiques dont les principes qui ont guidé la construction assurent sa réutilisabilité, notamment pour la réalisation d'applications concrètes à destination des apprenants. L'article présente un projet de construction d'un dictionnaire phraséologique multifonction offrant un double accès aux unités phraséologiques (UP), un accès via la forme des UP (fondé sur une approche sémasiologique de la langue) et un accès via le sens des UP (fondé sur une approche onomasiologique de la langue).*

- 1. Introduction
- 2. Phraséologie et TAL
- 3. Phraséologie du discours scientifique
- 4. Un outil d'aide à la maîtrise du niveau phraséologique de la LSG
- 5. Conclusion
- Références

### 1. Introduction

En Europe occidentale et en France tout particulièrement, la phraséologie constitue un domaine de recherche généralement encore assez peu connu. Son évolution, directement liée au développement de la linguistique de corpus et du traitement automatique du langage (TAL), va toutefois grandissant depuis les années quatre-vingt-dix. Dans cet article, la phraséologie est exploitée pour la construction d'outils d'aide à l'apprentissage des langues. L'étude explore le cas du discours scientifique et des difficultés de la maîtrise des formules conventionnelles de ce genre dans une langue étrangère, en l'occurrence en anglais. Par ailleurs, un postulat se trouve au centre de cette étude : l'existence d'une langue commune aux scientifiques à travers les disciplines, et désormais, grâce à l'accession de l'anglais au statut de *lingua franca* des sciences, à travers le monde. Pour

observer les caractéristiques de cet univers linguistique un corpus a été mis en place recueillant des textes de domaines scientifiques variés. Le dépouillement de ce corpus a abouti à une base de données phraséologiques réutilisable.

Le premier chapitre explore l'interaction entre le TAL et le domaine de la phraséologie. Le deuxième chapitre s'articule autour de la définition du concept de Langue Scientifique Générale (LSG) et du traitement de la phraséologie relative à cet univers linguistique particulier. Ce chapitre donne par ailleurs lieu à une discussion sur les intérêts des études phraséologiques de la LSG et il présente les travaux déjà existants sur ce sujet. Le troisième chapitre explore l'une des possibilités d'exploitation des données phraséologiques obtenues au terme de la collecte pour la réalisation d'applications concrètes.

## 2. Phraséologie et TAL

### 2.1. Une histoire parallèle

La linguistique de corpus et le TAL en général ont joué un rôle décisif dans l'évolution de la phraséologie, domaine à la frontière de la lexicologie et de la syntaxe qui se consacre à l'étude des combinaisons de mots récurrentes et arbitraires souvent appelées "collocations" ou "unités phraséologiques" (UP)<sup>[1]</sup>. Bien qu'elles constituent des éléments constants du discours, les combinaisons de mots lexicalisées n'en demeurent pas moins des éléments subliminaux du langage se refusant à tout repérage facile. Souvent absentes des dictionnaires traditionnels, elles parsèment le discours d'effets de "déjà entendu" et ne se présentent pas à l'esprit distinctement comme des unités minimales de la langue. La prolifération de corpus textuels informatisés, ainsi que de programmes pour leur exploitation, a apporté une réponse à cette difficulté en rendant possible le repérage des séquences de langue récurrentes :

*The increasing use of computer-held text corpora containing many millions of words has allowed linguists to establish lexico-grammatical patterns in language that were previously unavailable to observers. ([Gledhill00] : 115)*

Alors que dans les autres disciplines de la linguistique, l'apport des corpus et du TAL consiste principalement en un gage d'objectivité, en phraséologie les corpus apparaissent comme les piliers même de la recherche, puisque eux seuls permettent en effet de confirmer le statut phraséologique de certaines constructions polylexicales. Comme Habert *et al.* le remarquent :

*(...) étant donné une expression jugée "contrainte" quant à ses possibilités de transformation, les corpus permettent de chercher si ses réalisations effectives confirment ce jugement. ([Habert97] : 57)*

Par exemple, il n'est pas toujours aisé de déterminer si une séquence aussi banale que **to start a car** constitue une combinaison libre d'éléments lexicaux ou une combinaison restrictive. À première vue, il semble s'agir d'une construction libre étant donné que ses éléments constitutifs, le verbe **start** et le nom **car**, possèdent tous les deux des paradigmes combinatoires très étendus (**to start [a lesson/ a play/ a book/ studying/ cooking/ singing/...]**, **to [buy/ drive/ wash/ lend/ break/ look at/...] a car**). Toutefois, grâce au TAL et aux méthodes de la statistique lexicale, il est désormais possible de connaître la probabilité de leur apparition mutuelle et d'affirmer leur statut phraséologique. Par exemple, selon les résultats fournis par le *Collins Cobuild corpus*, la probabilité d'apparition de

l'unité **start** au voisinage de l'unité **car** est significativement forte : 8,6 %.<sup>[2]</sup>

Souvent, l'observation d'autres langues permet de confirmer le statut avéré d'une UP moyennant une interrogation de corpus, puisque dans le cas des vraies UP la traduction mot à mot est souvent impossible : **\*commencer une voiture** est effectivement inacceptable en français, de même que l'est **\*započeti auto** (traduction littérale) en croate. Le choix arbitraire de la collocation devient évident puisqu'en français on dira **démarrer une voiture** et en croate **upaliti auto** (litt. "allumer la voiture").

## 2.2. Limites du TAL au niveau phraséologique de la langue

Les techniques pour le traitement automatique de corpus servant à l'identification des UP utilisent des méthodes statistiques qui s'appuient sur des modèles mathématiques solides. Malgré leur rigueur scientifique, les résultats obtenus au moyen de ces techniques pointent trois défauts majeurs de ce type de méthode lexicographique : l'exclusion des UP à fréquence faible, le recensement possible des amalgames ne constituant visiblement pas des UP et l'absence de toute catégorisation des ressources linguistiques obtenues.

Dans le cadre du traitement automatique de corpus textuels, il existe un grand risque de négligence des modèles qui font visiblement (ou intuitivement) partie de la phraséologie, mais dont la fréquence n'est pas significative. Par exemple, dans sa quête des collocations à partir d'un corpus de textes du domaine des sciences et technologies de l'information, Curado ([Curado01] : 112) relève les unités suivantes comme ayant une fréquence très faible : **first and foremost**, **diskette drive** et **ticket booth**. Les lexicologues ont l'habitude de gérer ce type de problème de façon à réhabiliter ces unités dans leurs listes d'étude afin qu'elles reçoivent le même traitement que les unités de haute fréquence. De manière analogue, les unités polylexicales à haute fréquence qui ne constituent visiblement pas des UP (par ex. **we accurately**, **access which**) sont régulièrement éradiquées de ces mêmes listes.

Enfin, les techniques pour le traitement automatique de corpus offrent comme seul résultat des combinaisons de mots qui sont vraisemblablement des collocations. Elles ne sont pas capables de classer ces combinaisons suivant une typologie. Par exemple, des séquences aussi diverses que **to result in**, **quite possible** et **so as to** apparaîtront toutes dans une liste sans aucune information discriminatoire supplémentaire et sans aucune mise en relation des UP sémantiquement ou autrement proches. Or, une classification, par exemple selon le rôle joué par les UP dans le discours (cf. exemple 1) ou encore selon la notion à laquelle elles renvoient (respectivement "aboutissement", "probabilité" et "objectif") faciliterait grandement la manipulation des données recueillies par la suite, notamment dans le cadre de la création des applications à fins didactiques.

### Exemple 1

- **avoir pour résultat** -> **to result in** (fonction prédicative) ;
- **fort possible** -> [**very / easily / quite**] **possible** (fonction adjectivale) ;
- **en vue de <qch>** -> **so as to <do sth>**, **with a view to <(doing) sth>** (fonction adverbiale) ;

Par conséquent, tout traitement automatique de la phraséologie nécessite une relecture manuelle des résultats obtenus au terme de l'exploitation du corpus.

## 3. Phraséologie du discours scientifique

### 3.1. Langue Scientifique Générale : un postulat

Notre recherche défend l'idée de l'existence d'une langue commune aux scientifiques indépendamment de leurs spécificités disciplinaires et de leur origine linguistique, langue que nous avons décidé d'appeler "Langue Scientifique Générale" (LSG). Dans notre quête de la définition de cet univers linguistique particulier, nous avons préféré, à l'instar de Williams [Williams99], remplacer les notions de "sous-langage", de "langues de spécialité" et de "plurisystèmes" (telles qu'elles sont définies par ([Lerat95] : 11, 28) et ([Habert97] : 148-152)) par la notion de "communauté de discours", introduite par Swales ([Swales90] : 1 *et passim*). Alors que les premières sont souvent précisées selon des paramètres sociolinguistiques, Swales décrit une communauté de discours sur des bases socio-rhétoriques. En ce sens, la LSG peut être définie en tant que : **pratique langagière spécifique à une communauté de discours composée de chercheurs en sciences exactes dont les objectifs communicatifs poursuivis émanent des préoccupations partagées par des scientifiques à travers le monde et indépendamment de leurs spécificités disciplinaires.** L'emploi de plus en plus répandu d'une seule langue, l'anglais, à travers le monde pour les besoins communicatifs de cette communauté de discours, est la preuve indéniable de sa cohésion.

Par conséquent, notre étude se situe dans la lignée des travaux menés sur le "Vocabulaire Général d'Orientation Scientifique" (VGOS) par Phal. Sa définition du VGOS nous conforte dans notre point de vue :

*Le vocabulaire scientifique général est (...) commun à toutes les spécialités. Il sert à exprimer les notions élémentaires dont elles ont toutes également besoin (mesure, poids, rapport, vitesse, etc.) et les opérations intellectuelles que suppose toute démarche méthodique de la pensée (hypothèse, mise en relation, déduction et induction, etc.)* ([Phal71] : 9)

Rappelons que dans la tradition linguistique anglo-saxonne, on réfère parfois à l'univers linguistique visé dans cette étude en termes de *English for Science and Technology* (EST) ou encore en termes de *English for Academic Purposes* (EAP).

### 3.2. Intérêt des études phraséologiques de la LSG

La maîtrise de la phraséologie scientifique, de même que celle de la terminologie, apparaît comme un facteur non négligeable de la pertinence de la communication scientifique. Pourtant le respect des conventions en matière de phraséologie scientifique n'est pas toujours aisé. Dans le cadre de la LSG, ce constat est d'autant plus valable que pour un grand nombre de scientifiques il s'agit de manipuler une langue étrangère. En effet, de nos jours l'anglais joue le rôle de *lingua franca* de la communication scientifique internationale. Dans les sciences dites dures, les chiffres le confirment : 85,7 % de publications en anglais pour la biologie, 84,6 % pour la physique, 66,9 % pour la chimie ([Baldauf83] : 99).

### 3.3. Historique des travaux en LSG

Le nombre de travaux, dans les milieux anglophone et francophone, qui s'intéressent aux langues de spécialité et au discours scientifique plus particulièrement est conséquent. Toutefois ce nombre

diminue considérablement si l'on évalue leur intérêt pour la phraséologie. L'analyse systématique de la phraséologie scientifique est au cœur d'une petite poignée de travaux, ceux de [Gréciano97], [Gledhill00], [Williams99] et [LuzonMarco00] principalement. Il est important de noter que ces travaux n'abordent pas la phraséologie de la LSG par le biais d'une analyse croisée de plusieurs disciplines scientifiques. Les travaux de Gledhill, Williams et Luzon Marco traitent un seul domaine ou thème scientifique à la fois (respectivement, la recherche sur le cancer, les plantes parasites et le domaine médical). En ce qui concerne les travaux de Gréciano, ils sont davantage orientés vers la phraséologie des langues de spécialité de manière générale, phraséologie qui se trouve régulièrement abordée en regard de la langue commune<sup>[3]</sup>. Bien que les études à thème impliquent une certaine pluridisciplinarité, elles n'abordent pas la LSG de manière directe et explicite. Leur apport pour ce genre de recherche est toutefois de taille puisque, de par la nature même des phénomènes collocatifs, leurs observations sont dirigées vers les assemblages lexicaux les plus récurrents qui peuvent apparaître dans d'autres thèmes, d'autres disciplines et qui peuvent, de ce fait, faire partie de la LSG.

### 3.4. "Traquer" les UP de la LSG ?

Pour localiser les UP relatives à la LSG, un corpus bi-textuel a été mis en place. Ce corpus, d'une taille de 82 800 mots, rassemble des textes de trois domaines connexes : la chimie, la biologie et la physique. Un des critères principaux de sélection de nos textes était leur disponibilité dans les deux versions, anglaise et française. En définissant ce critère, nous avons voulu rendre possible l'étude contrastive de la langue scientifique dans son usage. La taille relativement petite du corpus est un fait voulu. En effet, le corpus devait rester à l'échelle d'une exploitation manuelle des données. Celui-ci a pu ainsi subir une double lecture : une première lecture automatique par le biais du logiciel *ZText* [Zinglé98] et une deuxième lecture, manuelle. Les résultats des ces deux lectures ont été confrontés. Les UP retenues ont été celles qui répondaient aux critères de sélection préétablis, à savoir la fréquence, la répartition (à travers les domaines) et la réexploitabilité des unités relevées. Notre hypothèse est la suivante : si une combinaison de mots connaît un emploi réitéré à travers les trois domaines scientifiques, elle est susceptible de faire partie de la LSG. Quant au critère de réexploitabilité, il est lié directement à notre objectif de recherche, à savoir la création d'outils d'aide à l'apprentissage des langues étrangères. L'étape de la collecte des données phraséologiques a été suivie par une étape d'étiquetage des données (cf. *infra*) ainsi que par une étape de vérification et d'enrichissement de la base (consistant en une évaluation des données par des locuteurs natifs et en une confrontation des données à diverses ressources linguistiques telles que les dictionnaires phraséologiques ou autres existants).

Une unité de traduction (UT) consiste en une UP anglaise ou une série d'UP anglaises à laquelle se trouvent associés son ou ses équivalents en français. La présentation des UT reflète le processus de leur collecte avec l'anglais comme langue source et le français comme langue cible (ex. [this work is an outgrowth of](#) -> [ce travail découle de](#)) mais, de temps à autre, l'anglais prend la place de la langue cible pour mettre en avant la perspective d'aide à la rédaction (ex. [ce travail découle de](#) -> [this work is an outgrowth of](#)).

À l'issue de ces deux étapes, la base de données actuelle comporte quelque 2000 unités de traduction (UT), sans la prise en compte des variations paradigmatiques possibles à l'intérieur des UT qui augmenterait sensiblement le nombre total des UT minimales<sup>[4]</sup>. Elle sert de point de départ à notre étude du système phraséologique de la LSG et à notre réflexion sur les possibilités de création d'un dictionnaire phraséologique bilingue.

### 3.5. Stocker les UP de la LSG

Les UP de la LSG sont stockées dans une base de données dont la réutilisabilité est assurée grâce à l'étiquetage des données. Un jeu de six étiquettes a été développé pour les besoins de la recherche : étiquette fonctionnelle, étiquette qualificative, étiquette notionnelle, étiquette du domaine scientifique, étiquette de l'exemple et étiquette de la référence bibliographique. Celles qui répondent directement aux besoins applicatifs de cette recherche sont l'étiquette fonctionnelle et l'étiquette notionnelle (voir à ce sujet [Zinglé03]).

L'étiquette fonctionnelle est intéressante du point de vue de l'utilisation de la langue parce qu'elle apporte une information sur le rôle joué par l'UP dans le discours (cf. exemple 2).

#### Exemple 2

- valeur de quantifieur (*a wide range of* -> *une vaste gamme de*) ;
- valeur qualitative (*well-known* -> *bien connu*) ;
- valeur adverbiale (*within the framework of* -> *dans le cadre de*) ;
- valeur de connecteur (*as expected* -> *comme prévu, comme on s'y attendait*) ;
- valeur prépositionnelle (*with respect to* -> *en ce qui concerne, par rapport à*) ;
- valeur prédicative (*to address a question* -> *soulever une question*) ;
- etc.

Parallèlement, l'étiquette notionnelle apporte une information sur le sens de l'UP et peut constituer un accès aux UP, facilitant ainsi la formulation du sens dans la langue seconde. Au total, 125 étiquettes notionnelles ont été mises en place, apportant une preuve de l'homogénéité de l'univers linguistique observé (seulement 6,6 % du corpus phraséologique n'a pas pu être classé dans des grands thèmes identifiés à ce stade de la recherche).

### 3.6. Spécificités de la phraséologie scientifique

L'analyse des données phraséologiques collectées selon la méthode décrite dans la section précédente indique que l'univers phraséologique relatif à la LSG est caractérisé par un type d'UP bien déterminé. Si l'on se réfère au modèle dominant de la typologie des UP [Pecman04], proposé par des linguistes comme Howarth [Howarth96], Cowie [Cowie98], Mel'čuk [Melcuk93] ou encore Hausmann [Hausmann84], il s'avère que la plupart des unités phraséologiques collectées entrent dans la catégorie des "collocations restrictives", catégorie à laquelle de nombreux linguistes réfèrent d'ailleurs sous le terme simple de "collocations".

Les idiomes purs tels que *vendre la mèche*, *casser sa pipe*, *avoir du pain sur la planche*, etc., n'apparaissent jamais dans le discours scientifique. Ce type d'unités est d'ailleurs d'une manière générale plutôt rare dans la langue écrite. De plus, on ne trouve guère dans notre corpus d'idiomes figuratifs comme *sortir du lot*, *tourner au vinaigre*, *c'est dans la poche*, etc., si ce n'est quelques UP lexicalisées dont l'image est devenue complètement opaque, comme celles qui gravitent autour des thèmes de *light* ou *step* (cf. exemple 3).

### Exemple 3

- in the light of -> à la lumière de, compte tenu de ;
- to shed further light on <sth> -> éclairer (<qn>) sur <qch> ;
- to spotlight <sth> -> mettre <qch> en lumière, souligner ;
- to bring to light -> découvrir ;
- as a first step -> comme première étape ;
- to take steps -> prendre des mesures ;
- to be a key / an important step -> être une étape importante / clé ;
- to remain in step with <sth> -> suivre le rythme de <qch>.

Ainsi, la phraséologie de la LSG consiste principalement en un établissement d'associations privilégiées entre les unités lexicales (UL), associations observées dans des paradigmes (*clusters*) tels que, par exemple, ceux qui gravitent autour de la notion d'|hypothèse|, dont un extrait est donné dans l'exemple 4.

### Exemple 4

- [bâtir / construire] une hypothèse -> to construct a hypothesis ;
- adopter une hypothèse -> to [accept / agree with] a hypothesis ;
- confirmer une hypothèse -> to [confirm / prove / support] a hypothesis ;
- [corroborer / renforcer] une hypothèse -> to [consolidate / strengthen] a hypothesis ;
- dans l'hypothèse de / où -> in case of, in the event of ;
- [démentir / contredire] une hypothèse -> to [invalidate / refute] a hypothesis ;
- écarter l'hypothèse de -> to rule out the possibility of ;
- [émettre / formuler / avancer / suggérer] une hypothèse -> to [put forward / put forth / advance / formulate / suggest] a hypothesis ;
- faire des hypothèses (sur) -> to speculate (about) ;
- proposer une hypothèse -> to [invoke / propose] a hypothesis ;
- hypothèse directrice -> main hypothesis ;
- hypothèse plausible -> plausible hypothesis .

Toutefois, les limites entre les différentes catégories d'UP ne sont pas toujours faciles à établir, notamment celles qui séparent les collocations restrictives des combinaisons libres. Les UP de notre base de données s'approchent ainsi parfois sensiblement de ce que l'on pourrait appeler, à l'instar de Gross, des "classes d'objets" :

*Les domaines d'arguments de prédicats (verbaux, nominaux ou adjectivaux), quand ils sont libres, peuvent être exprimés en termes de classes d'objets (...) et non d'unités lexicales. (...) Le verbe **ressembler** a comme objet un élément de la classe des chaussures (soulier, sandale, botte, etc.). Le verbe **lire** est caractérisé par deux types de classes : celle des textes (récit, dissertation, roman, rédaction, poème, article, etc.) et celle des supports de textes (livre, journal, cahier, brochure, etc.) ([Gross96] : 78-79)*

Après avoir exposé la méthode de collecte des UP qui constituent le fond phraséologique commun des sciences expérimentales et après avoir décrit leurs spécificités fondamentales, nous présenterons une réflexion sur les possibilités d'exploitation de la base de données phraséologiques pour la construction d'outils d'aide à l'apprentissage des langues étrangères.

## **4. Un outil d'aide à la maîtrise du niveau phraséologique de la LSG**

Un des outils qui pourrait répondre le mieux aux difficultés que connaissent les locuteurs non natifs face aux constructions de mots récurrentes et arbitraires se présente, à notre avis, sous forme d'un dictionnaire phraséologique multifonction. Celui-ci se compose de deux volets : un volet qui propose une approche sémasiologique du vocabulaire phraséologique (le dictionnaire phraséologique proprement dit) et l'autre qui offre une approche onomasiologique du même vocabulaire (le thésaurus phraséologique).<sup>[5]</sup>

Un tel projet tend ainsi à fournir des réponses concrètes qui pourraient faciliter l'apprentissage des langues étrangères. En ce sens, notre travail partage les préoccupations des chercheurs en didactique des langues étrangères ([Cohen90] ; [Ellis94] ; [Lessard97]) et plus spécifiquement celles des chercheurs qui abordent la question de la production en langue étrangère dans l'univers des connaissances scientifiques ([Bowers95] ; [Histon03]). En effet, les connaissances phraséologiques sont souvent vues comme présentant un obstacle davantage dans le processus d'encodage que dans le processus de décodage de la langue.

### **4.1. Dictionnaire phraséologique de la LSG**

Le dictionnaire phraséologique de la LSG est sous-tendu par une approche sémasiologique de la langue permettant de connaître la traduction d'UP recherchées dans la langue cible (cf. Figure 1). En effet, toute mise en relation des unités lexicales de deux ou plusieurs langues passe par la compréhension du contenu sémantique spécifique des UP et de leur distribution dans la langue donnée. Par conséquent, l'appariement des UP bilingues équivaut au processus de définition des UP dans la perspective monolingue.

Le dictionnaire phraséologique bilingue devrait permettre en outre de consulter les données à partir de n'importe quel mot contenu dans l'expression recherchée, afin de faciliter l'accès aux UP [Zinglé03].

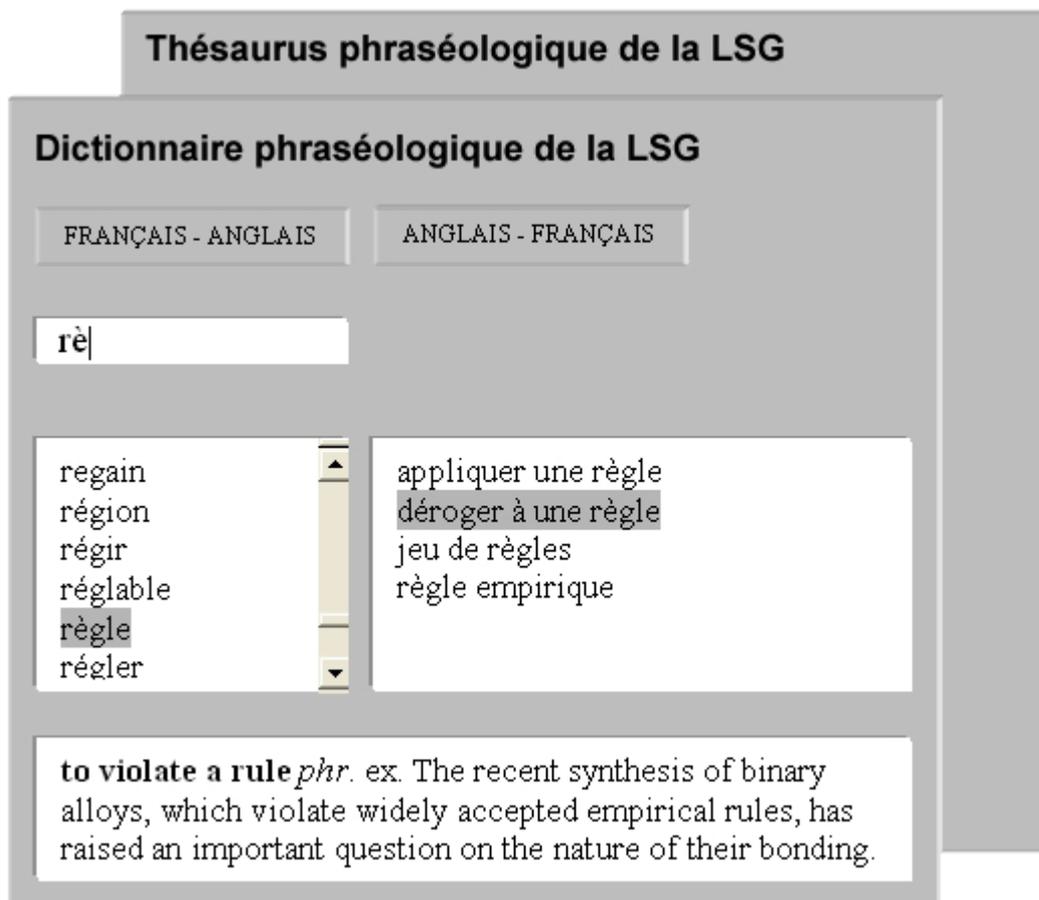
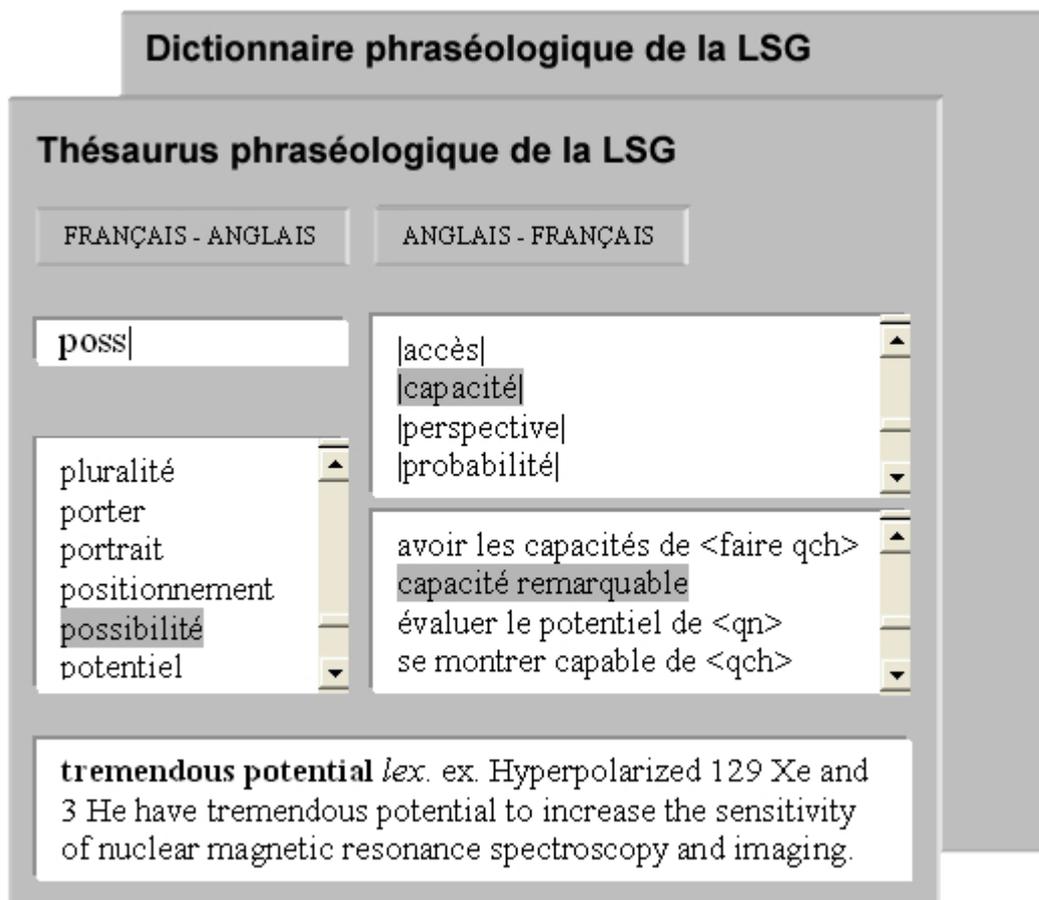


Figure 1 - Maquette pour la création d'un dictionnaire phraséologique.

## 4.2. Thésaurus phraséologique de la LSG

L'approche onomasiologique de la langue apparaît comme une réponse aux problèmes de formulation de sens dans la langue seconde. Un des principaux obstacles auxquels se trouvent confrontés les locuteurs non natifs est la difficulté que l'on a parfois à trouver une forme linguistique précise pour exprimer une idée [Zinglé96].

Notre thésaurus est sous-tendu par une ontologie de la LSG élaborée pour les besoins de la recherche. Étant donné que le seul point d'accès au sens possible demeure les UL, la recherche dans le thésaurus est assurée *via* les mots-clés (cf. Figure 2).



**Figure 2** - Maquette pour la création d'un thésaurus phraséologique.

Un tel outil pourrait trouver des applications pédagogiques intéressantes non seulement en tant que simple répertoire bilingue des unités lexicales complexes permettant de retrouver facilement les équivalents dans les deux langues, mais aussi en tant que source potentielle de données utiles pour la création d'exercices destinés à l'enseignement systématique des paradigmes combinatoires des UL dans la langue seconde. D'ailleurs, le regroupement des UP formellement différentes mais ayant un sens proche, réalisé grâce à l'approche onomasiologique des UP, permet d'ores et déjà d'envisager la construction de fiches de vocabulaire thématiques qui pourraient constituer une aide utile aux apprenants dans le processus d'acquisition de la compétence phraséologique de la langue seconde.

## 5. Conclusion

La présente étude montre que l'approche phraséologique de la langue offre de nombreuses perspectives pour la création d'outils d'aide à l'apprentissage des langues. Face à la difficulté que présente le maniement des collocations restreintes pour les locuteurs non natifs, les outils recensant les UP bilingues offrent des solutions concrètes et pertinentes. Dans le cadre du discours scientifique, de tels outils sont d'autant plus bienvenus que l'efficacité de la communication scientifique repose grandement sur le respect des conventions en matière de phraséologie. Le projet de création du dictionnaire phraséologique multifonction répond aussi à une absence générale à l'heure actuelle de dictionnaires phraséologiques bilingues. Enfin, bien que l'investigation de la phraséologie scientifique présentée dans cet article repose sur l'analyse d'un corpus textuel relatif aux sciences expérimentales, les UP relevées semblent dépasser largement ce cadre. Le caractère très général des UP recensées suggère leur pertinence dans le cadre global des sciences exactes,

voire dans le cadre des sciences humaines et sociales, corroborant ainsi notre hypothèse de l'existence d'une LSG.

## Références

*Les liens externes étaient valides à la date de publication.*

[Baldauf83]

Baldauf, R. B. & Jernudd, B. H. (1983). "Language of Publications as a Variable in Scientific Communication". *Australian Review of Applied Linguistics*, vol. 6, 1. pp. 97-108.

[Benson89]

Benson, M. (1989). "The Structure of the Collocational Dictionary". *International Journal of Lexicography*. vol. 2, 1. pp. 1-14.

[Bowers95]

Bowers, R. (1995). "A Computer-Mediated Scientific Writing Program". *TESL-EJ*, vol. 1, 3, A-3. <http://www.kyoto-su.ac.jp/information/tesl-ej/ej03/a3.html>

[Cohen90]

Cohen, A. (1990). *Language Learning: Insights for Learners, Teachers, and Researchers*. New York: Newbury House.

[Cowie98]

Cowie, A. P. (1998). "Introduction". In Cowie, A. P. (dir.) *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press. pp. 1-20.

[Curado01]

Curado, F. A. (2001). "Lexical Behaviour in Academic and Technical Corpora: implications for ESP development". *Language Learning & Technology*, vol. 5, 3. pp. 106-129.

[Ellis94]

Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

[Gledhill00]

Gledhill, Ch. J. (2000). *Collocations in Science Writing*. Language in performance: 22. Tuebingen: Gunter Narr Verlag.

[Gréciano97]

Gréciano, G. (1997). "Collocations rythmologiques". *Meta*, vol. 42, 1. pp. 33-44.

[Gross96]

Gross, G. (1996). *Les expressions figées en français : noms composés et autres locutions*. Paris / Gap : Ophrys, coll. "L'essentiel français".

[Habert97]

Habert, B., Nazarenko, A. & Salem, A. (1997). *Les linguistiques de corpus*. Paris : Armand Colin, coll. U.

[Halliday66]

Halliday, M. A. K. (1966). "Lexis as a linguistic level". In Bazell, Ch. E., Catford, J. C. Halliday, M. A. K. & Robins, R. H. (dir.). *In Memory of J. R. Firth*. London/Harlow: Longmans. pp. 148-162.

[Hausmann84]

Hausmann, F. J. (1984). "Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortverbindungen". *Praxis des neusprachlichen Unterricht*, vol. 31. pp. 395-406.

[Histon03]

Histon, C. (2003). "Writing Scientific Project Applications for Peer Review". *The Internet TESL Journal*. vol. 9, 8. <http://iteslj.org/Articles/Histon-ProjectApplications.html>

[Howarth96]

Howarth, P. A. (1996). *Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making*. Tübingen: Niemeyer.

[Lerat95]

Lerat, P. (1995). *Les langues spécialisées*. Paris : PUF, coll. "Linguistique nouvelle".

[Lessard97]

Lessard-Clouston, M. (1997). "Language Learning Strategies: An Overview for L2 Teachers". *The Internet TESL Journal*. vol. 3, 12. <http://www.aitech.ac.jp/~iteslj/Articles/Lessard-Clouston-Strategy.html>

[LuzonMarco00]

Luzón Marco, M.ºJ. (2000). "Collocational frameworks in medical research papers: a genre-based study". *English for Specific Purposes*, vol. 19, 1. pp. 63-86.

[Melcuk93]

Mel'čuk, I. A. (1993). "La phraséologie et son rôle dans l'enseignement/apprentissage d'une langue étrangère". *Études de Linguistique Appliquée*, vol. 92. pp. 82-113.

[Pecman04]

Pecman, M. (2004). "L'enjeu de la classification en phraséologie". *EUROPHRAS 2004*, 26-29 août 2004, université de Bâle, Suisse. Baltmannsweiler : Schneider Hohengehren Verlag. pp. 127-146.

[Phal71]

Phal, A. (1971). *Vocabulaire général d'orientation scientifique (V.G.O.S.) : part du lexique commun dans l'expression scientifique*. Paris : Crefid.

[Sinclair66]

Sinclair, J. M. (1966). Beginning the Study of Lexis. In Bazell, Ch. E. *et al.* (dir.). *In Memory of J. R. Firth*, London/Harlow: Longmans. pp. 410-430.

[Smadja93]

Smadja, F. (1993). "Retrieving collocations from text: Xtract". *Computational Linguistics*, vol. 19, 1. pp. 143-177.

[Swales90]

Swales, J. M. (1990). *Genre Analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

[Williams99]

Williams, G. C. (1999). *Les réseaux collocationnels dans la construction et l'exploitation d'un corpus dans le cadre d'une communauté de discours scientifique*. Thèse de doctorat, université de Nantes.

[Zinglé96]

Zinglé, H. (1996). "ZART : un logiciel d'aide à la rédaction scientifique et technique en langue étrangère". *Travaux du Lilla*, n° 1, publication de la faculté des lettres, arts et sciences humaines de l'université de Nice-Sophia Antipolis. pp. 111-113.

[Zinglé98]

Zinglé, H. (1998). "ZTEXT: un outil pour l'analyse de corpus". *Travaux du Lilla*, n° 3, publications de la faculté des lettres, arts et sciences humaines de l'université de Nice-Sophia Antipolis. pp. 69-78.

[Zinglé03]

Zinglé, H. (2003). "Constitution et exploitation d'une base de données phraséologiques en français". *Actes du Congrès international des linguistes*, 24-29 juillet 2003, Prague.

## Notes

[1] Bien que le domaine de la phraséologie souffre encore de nos jours d'une terminologie assez floue et d'une difficulté définitoire de son propre objet d'étude (due principalement à la complexité de celui-ci), les termes "unité phraséologique" et "collocation" commencent à être unanimement admis et reçoivent de plus en plus souvent la définition de "*an arbitrary and recurrent word combination*" ([Benson89] : 3).

[2] Cette méthode d'investigation des amalgames de mots lexicalisés est sous-tendue par l'interprétation statistique de la collocation en tant qu'entité linguistique composée d'éléments monolexicaux réapparaissant à une distance *n* de manière significative au sein d'un corpus représentatif d'une langue (cf. [Halliday66] ; [Sinclair66] ; [Smadja93]).

[3] Dans le cadre du projet *CONPHRAS-PROCOP* par exemple, Gréciano ([Gréciano97] : 34) analyse à la fois les domaines de la médecine, de l'économie, du sport, de l'administration et de la politique.

[4] En effet, une unité comme : *description [détaillée / précise / minutieuse / exhaustive]* -> *[detailed / accurate /thorough / exhaustive] description* est comptée comme une seule UT. Elle consiste toutefois en plusieurs UT minimales (par ex. *description détaillée* -> *detailed description*, *description précise* -> *accurate description*, etc.).

[5] Ce projet est actuellement en cours de réalisation à l'université de Nice-Sophia Antipolis.

## À propos de l'auteure

**Mojca PECMAN** a soutenu sa thèse en linguistique à l'université de Nice en 2004 sur le thème "Phraséologie contrastive anglais-français : analyse et traitement en vue de l'aide à la rédaction scientifique". Auteure de plusieurs articles portant sur les divers aspects de la recherche en phraséologie bilingue, elle enseigne la linguistique et l'informatique depuis cinq ans au sein du département des sciences du langage en qualité d'ATER et précédemment de monitrice.

**Courriel :** pecman@unice.fr

**Adresse :** Laboratoire d'Ingénierie Linguistique et de Linguistique Appliquée (LILLA), université de Nice-Sophia Antipolis, 98 bd Édouard Herriot, BP 209, 06204 Nice Cedex 3, France.

Ce texte fait partie des textes de la journée Atala 2005 qui font l'objet d'un numéro spécial d'Alsic.

Date de réception de l'article 1er novembre 2004 ; date d'acceptation, 30 mai 2005.

### Référence de l'article :

Pecman, M. (2005). "Les apports possibles de la phraséologie à la didactique des langues étrangères". *Apprentissage des langues et systèmes d'information et de communication (ALSIC)*, vol. 8, n° 1. pp. 109-122. [http://alsic.u-strasbg.fr/v08/pecman/alsic\\_v08\\_15-rec9.htm](http://alsic.u-strasbg.fr/v08/pecman/alsic_v08_15-rec9.htm), mis en ligne le 15/12/2005.



**ALSIC | Sommaire | Consignes aux auteurs | Comité de rédaction | Inscription**

© *Apprentissage des Langues et Systèmes d'Information et de Communication*, décembre 2005