

Phraséologie et lexicographie bilingue: Apports croisés des corpus monolingues et parallèles

Sylviane Granger

Granger, S. (2021). Phraséologie et lexicographie bilingue: Apports croisés des corpus monolingues et parallèles. In Raluca, N. & Hanote, S. (eds) *Opérations prédicatives et énonciatives, contrastivité et corpus*. Presses universitaires de Rennes, 205-224.

Résumé

La perspective élargie du domaine de la phraséologie, qui repose sur l'utilisation de larges corpus de textes informatisés et de techniques d'extraction automatiques puissantes, n'a pas encore eu l'impact qu'elle pourrait avoir en lexicographie bilingue. Dans cette étude, je me propose de décrire certaines des lacunes que présente la couverture phraséologique des dictionnaires bilingues et de suggérer quelques pistes d'amélioration. Un accent particulier est mis sur l'apport des corpus, monolingues et parallèles, au traitement des locutions invariantes, particulièrement adverbiales, qui peuvent être aisément extraites grâce à la technique des segments répétés. Comme ces locutions fonctionnent comme entités unitaires sur le plan syntagmatique, il est suggéré de leur donner le statut d'entrée principale dans les dictionnaires et de les faire bénéficier de la même qualité de description que les lexies simples.

Abstract

The extended view of phraseology, which relies on the use of large electronic corpora and powerful automatic extraction methods, has not yet had the expected impact on bilingual lexicography. The aim of this study is to examine some of the weaknesses in the phraseological coverage of current bilingual dictionaries and to suggest ways of overcoming them. The main focus will be on the contribution of monolingual and parallel corpora to the description of invariant phrases, especially adverbial ones, which can easily be extracted automatically by means of the n-gram method. As these phrases function as unitary syntagmatic entities, it is suggested that they should appear as main entries in dictionaries and receive the same depth of description as single words.

Introduction

La phraséologie peut se définir de manière générale comme « le domaine qui traite les séquences lexicales perçues comme préconstruites » (Legallois & Tutin 2013: 3). Jadis centré sur les unités présentant un haut degré de figement sémantique et syntaxique, le domaine s'est fortement ouvert ces dernières années et a progressivement grignoté du terrain sur le domaine des expressions dites 'libres'. De mon point de vue, il s'agit là d'un développement très positif, qui cadre parfaitement avec l'extrême fluidité de la notion de figement. Comme le précisent à juste titre Klein & Lamiroy (2005: 79), « le figement constitue un phénomène si central dans le fonctionnement du langage qu'on en viendrait à se demander s'il existe des assemblages vraiment libres ».

Cette évolution du domaine de la phraséologie est principalement liée à l'utilisation croissante de larges corpus électroniques et à des méthodes d'extraction automatique qui ont permis de mettre au jour une large palette d'unités phraséologiques présentant des degrés de figement

moins marqués que celles qui étaient jusqu'alors étudiées en priorité et une fréquence d'emploi beaucoup plus élevée que les expressions étudiées traditionnellement.

De manière générale, l'approche sur corpus a démontré que la phraséologie occupe une position bien plus centrale dans le langage qu'on ne l'avait pensé jusqu'alors. En même temps, cette perspective large contribue à brouiller les frontières du champ phraséologique et on pourrait craindre qu'à terme, la phraséologie ne se vide de son sens: si tout est phraséologique, rien ne l'est! Tout comme Legallois & Tutin (2013), je pense qu'il s'agit là d'un risque qui vaut la peine d'être couru.

L'évolution du domaine a pour conséquence que deux approches bien distinctes de la phraséologie coexistent (Granger & Paquot 2008) : une approche que l'on pourrait qualifier de 'lexicologique', étant donné que les critères utilisés pour identifier les unités phraséologiques sont d'ordre linguistique (compositionnalité sémantique, fixité syntaxique, combinatoire lexicale restreinte) et une approche 'quantitative', qui consiste à extraire des combinaisons préférentielles automatiquement à l'aide de méthodes quantitatives basées sur la co-occurrence et la récurrence. Les procédés d'extraction automatique sont essentiellement de deux ordres. D'une part, des tests statistiques qui mesurent la force d'association entre deux co-occurents, tels que le t-score ou l'information mutuelle, permettent d'extraire un large éventail d'unités de natures linguistiques diverses, qui peuvent être contiguës (*pluie torrentielle*) ou non contiguës (*une pluie aussi soudaine que torrentielle*). Parmi celles-ci, on retrouve des unités généralement considérées comme phraséologiques, tels que des idiomes (*voir rouge*), des collocations (*grièvement blessé*) et des locutions (*du coup*), mais aussi des associations libres (*famille formidable*). D'autre part, l'extraction de séquences de *n* mots contigus, répétées à l'identique dans un même corpus de textes, permet de mettre au jour des blocs lexicalisés (« lexical bundles », cf. Biber et al. 1999) qui jouent un rôle important tant à l'oral qu'à l'écrit. Il est important de noter que toutes les séquences extraites ne constituent pas des unités phraséologiques, même dans le sens large du terme, certaines séquences étant d'ailleurs incomplètes (*président de la*). Néanmoins, bon nombre des séquences mises en évidence par cette technique revêtent un réel intérêt phraséologique. C'est le cas des locutions, principalement nominales (*art contemporain*), adverbiales (*au contraire*) et prépositives (*en dépit de*), et de séquences plus longues qui ont souvent une valeur modale (*il y a fort à parier que*).

Bien que les deux courants phraséologiques soient pour l'heure encore fort éloignés l'un de l'autre, ils sont en fait parfaitement complémentaires: l'approche quantitative génère un grand nombre d'unités potentiellement intéressantes qui demandent à être analysées à l'aide de critères linguistiques, tandis que l'approche lexicologique fournit les critères nécessaires à l'analyse (Klein & Lamiroy 2016) et permet, en outre, de prendre en compte des unités phraséologiques qui, en raison de leur faible fréquence d'emploi, ne sont que peu représentées dans les corpus.

La phraséologie étant rarement transposable littéralement d'une langue à l'autre, il est essentiel qu'elle fasse l'objet d'une attention toute particulière dans les dictionnaires bilingues. Or, même si la couverture phraséologique des dictionnaires bilingues s'est élargie ces dernières années, il subsiste une tendance à privilégier les unités sémantiquement non-compositionnelles, qui posent des difficultés de compréhension et qui sont le reflet des cultures d'où elles émanent. Pour Kastberg Sjöblom (2013: 208), « [o]n pourrait même considérer qu'elles constituent les fleurons du dictionnaire, car elles sont souvent imagées, drôles et très belles ». Du point de vue de la production, cependant, ce sont les unités moins

figées sémantiquement, y compris certaines combinaisons dites ‘libres’, qui posent le plus de problèmes. En conséquence, l’adoption d’une définition très restrictive de la notion d’unité phraséologique dans les dictionnaires bilingues est difficilement défendable.

Le présent article adopte cette perspective large et vise à montrer que le recours à des corpus de textes informatisés et l’application de techniques automatiques, en particulier l’extraction de segments répétés, peuvent contribuer à améliorer sensiblement la couverture phraséologique des dictionnaires bilingues. Les dictionnaires visés sont les dictionnaires bilingues généraux. Il sera néanmoins occasionnellement fait référence à des lexiques bilingues orientés machine, qui sont conçus pour des applications en traitement automatique du langage, car des avancées significatives ont été faites dans ce cadre. L’accent est mis sur la paire français-anglais mais l’ensemble des idées émises est transposable à d’autres paires de langues.

La section 1, qui présente un bref état des lieux de la phraséologie dans les dictionnaires bilingues, est suivie d’un aperçu des apports que l’on peut attendre des corpus monolingues et parallèles (section 2). La section 3 aborde la problématique des locutions, particulièrement adverbiales, et contient une analyse de deux locutions (*de la même manière* et *de la même façon*) dans les dictionnaires bilingues et dans les corpus. La section 4 tire les conclusions de l’étude et propose quelques pistes de réflexion pour des recherches futures dans le domaine.

1. La phraséologie dans les dictionnaires bilingues

La phraséologie a toujours occupé une place importante dans les dictionnaires bilingues. Cependant, comme l’expliquent Béjoint & Thoiron (1996), cette attention à la phraséologie dans les dictionnaires bilingues a eu tendance à s’amenuiser au fil du temps. En conséquence, alors qu’au départ, les dictionnaires monolingues anglais faisaient pâle figure par rapport aux dictionnaires bilingues, aujourd’hui, la situation s’est inversée. Ceci n’a en fait rien d’étonnant quand on pense aux nombreux obstacles auxquels est confronté le traitement lexicographique de la phraséologie, particulièrement dans une perspective contrastive. Pecman (2008: 209-210) en identifie pas moins de quinze, dont cinq spécifiques aux dictionnaires bilingues, et conclut: « Presently, the phraseological information included in traditional bilingual dictionaries is often missing or at best incomplete and presented in a disorderly fashion ». Plutôt que de présenter un survol de ces nombreuses difficultés, il m’a semblé plus opportun de décrire de manière plus approfondie deux d’entre elles, à savoir les critères d’inclusion des unités phraséologiques et leur structure d’accès.

1.1. Critères d’inclusion

Le premier problème, déjà évoqué plus haut, est que les dictionnaires bilingues ont tendance à privilégier les unités les plus figées sémantiquement, tels que les idiomes (*jeter l’éponge*), les composés (*hors-bord*), les proverbes (*l’occasion fait le larron*) ou, pour l’anglais, les verbes à particule (*to catch up*). Néanmoins, même pour ce type d’unité, il subsiste de nombreuses lacunes. Ainsi, par exemple, Mogorrón Huerta (2008) constate qu’une large proportion des locutions verbales à sens idiomatique (c’est-à-dire non compositionnelles) est absente des dictionnaires bilingues espagnol-français.

En ce qui concerne les unités dont le sens est compositionnel, du moins en partie, la situation est contrastée. Dans les dictionnaires bilingues actuels, des améliorations notoires ont été apportées au niveau des collocations, qui sont plus présentes que dans les versions antérieures

des dictionnaires et mieux mises en évidence. Néanmoins, il subsiste encore une large marge de progression. Ainsi, Moon (2008 : 330) analyse les mots *river* et *rivière/fleuve* dans les dictionnaires bilingues français-anglais Robert & Collins et Oxford Hachette et déplore la pauvreté de l'information au niveau des collocations et des colligations: Bertels et al. (2009) font le même constat pour les verbes dénotant la notion de hausse en français, en anglais et en néerlandais (par exemple, *augmenter*, *croître*, *progresser*). En outre, afin d'avoir une vue claire des collocations qu'il convient d'inclure, il est indispensable que les lexicographes aient une bonne compréhension des différents types d'unités phraséologiques et des critères clairs pour les distinguer. Or, cela ne semble pas être toujours le cas. Ainsi, De Giovanni (2014) analyse les collocations dans trois dictionnaires bilingues français-italien et constate que le statut et la notion de collocation sont loin d'être clairs. Il conclut que « [s]ans une réglementation de la notion de collocation, les dictionnaires bilingues auront du mal à la traiter au mieux d'un point de vue lexicographique ». Une autre raison qui peut expliquer l'absence de certaines collocations est liée aux limitations inhérentes au support papier. Plus que d'identifier ce qu'ils ont intérêt à inclure dans un dictionnaire, un des soucis majeurs des lexicographes est d'identifier ce qu'ils peuvent se permettre de ne pas inclure (cf. Fillmore & Atkins 1994: 350-351: « But, above all, what can be left out? »). A l'heure où les dictionnaires bilingues électroniques prennent largement le pas sur les dictionnaires « papier », cette contrainte n'est plus vraiment d'actualité. Or, il faut bien constater que, jusqu'à présent, le changement de support n'a pas été de pair avec un accroissement significatif des collocations.

Une autre catégorie qui est traitée de manière très parcellaire dans les dictionnaires bilingues et qui, contrairement aux collocations, ne semble faire l'objet que de très peu d'études est celle des « routines ». Cette notion a des contours assez flous mais elle fait toujours référence à des séquences préconstruites, souvent sémantiquement compositionnelles, qui ont des fonctions rhétoriques ou interactionnelles importantes tant à l'oral qu'à l'écrit. Née et al. (2014: 2119) parlent de « routines discursives » et les définissent de la manière suivante: « Une routine discursive consiste en la mise en relation de séquences linguistiques récurrentes, partiellement figées (...) avec des déterminations discursives et des fonctions textuelles propres à un genre ou une sphère d'activité ». On retrouve souvent le terme de « routines conversationnelles » (Klein & Lamiroy 2011), car ce phénomène est très fréquent à l'oral mais le rôle que ce type de séquence joue à l'écrit fait l'objet d'un nombre croissant d'études (cf. les travaux de Tutin 2014 pour la langue scientifique). Après avoir recherché un certain nombre de ces unités dans des dictionnaires monolingues anglais et bilingues anglais-français et anglais-allemand, Siepmann (2008) conclut que, probablement en raison de leur sens largement compositionnel, les routines sont souvent absentes des dictionnaires. Ceci est très regrettable car une traduction littérale de ces séquences est rarement possible. Heureusement, comme elles sont souvent totalement figées syntaxiquement, le recours à la technique des segments répétés devrait permettre d'en dresser une liste, qui sera sans doute incomplète, mais devrait déjà contribuer à améliorer quelque peu la couverture lexicographique de ce type d'unité.

1.2. Structure d'accès

Même lorsque les unités phraséologiques sont présentes dans les dictionnaires, elles sont souvent difficiles d'accès, car elles peuvent se retrouver à divers endroits et ne sont pas toujours facilement repérables. Lorsqu'elles font partie de la nomenclature, elles apparaissent comme entrées principales et sont faciles à localiser. Cependant, le plus souvent elles n'apparaissent que comme sous-entrées ou sont simplement incluses dans les exemples, ce

qui en fait des entrées de seconde zone, moins facilement accessibles et décrites de manière plus sommaire. Pour bien décrire le problème, il est important de distinguer les dictionnaires bilingues sur support papier et les dictionnaires bilingues électroniques.

Dans la plupart des cas, les dictionnaires « papier » contiennent une préface qui donne à l'utilisateur des renseignements précieux sur le choix et la localisation des unités polylexicales dans le dictionnaire. Comme le fait remarquer Mulhall (2007), c'est loin d'être le cas pour tous les dictionnaires ou pour tous les types d'unités. La localisation des mots composés pose aussi problème. Ainsi, le dictionnaire Robert & Collins (2006) précise que le statut d'entrée principale ou de sous-entrée se fait sur la base d'un critère formel. Si la séquence nominale s'écrit avec un trait d'union, elle fera partie de la nomenclature. C'est le cas, par exemple, de *maître-autel* et *maître-chanteur*. Si, par contre, elle s'écrit sans trait d'union (auquel cas, certains linguistes parleront plutôt de locution, voir section 3), elle se retrouvera sous le premier mot de la séquence, dans une sous-section regroupant les composés formés à partir de ce mot. C'est le cas, par exemple, pour *maître artisan*, *maître de conférences*, *maître d'hôtel*, *maître nageur* et *maître queux*, qui sont listés au bas de l'entrée du lemme *maître*. Cette pratique est problématique à plus d'un titre. D'une part, lorsque l'entrée principale est fort longue, elle rend la recherche particulièrement fastidieuse. D'autre part, un critère purement formel est loin d'être idéal pour les utilisateurs qui ne maîtrisent pas nécessairement l'orthographe des mots recherchés, particulièrement dans un domaine aussi aléatoire que la présence ou non du trait d'union. La situation est d'autant plus déroutante pour l'utilisateur que la pratique adoptée pour l'anglais est différente. Seuls les composés soudés (*mastermind*, *masterpiece*, *masterwork*) ont le statut d'entrée principale, les autres composés, qu'ils comportent un trait d'union (*master-at-arms*) ou non (*master key*), sont listés sous *master*. Pour ajouter à la confusion des utilisateurs, les stratégies adoptées par les différents dictionnaires bilingues peuvent différer très fortement. Ainsi, le dictionnaire Larousse (2007) utilise un critère sémantique et attribue un statut d'entrée à part entière aux composés anglais considérés comme des « unités de sens autonome », un critère bien trop sophistiqué à appliquer pour l'utilisateur qui comprendra difficilement pourquoi *master key* apparaît comme entrée principale, alors que *master switch* et *master copy* sont des sous-entrées.

On pourrait s'attendre à ce que ce problème ne se pose plus pour les dictionnaires électroniques étant donné que « [l]e stockage informatique rend beaucoup plus facile la navigation non linéaire dans la nomenclature et, plus généralement, dans l'ensemble des informations contenues dans le dictionnaire » (Polguère 2008 : 234-5). En effet, pour un grand nombre d'unités, dont toutes celles comprenant *maître* mentionnées ci-dessus, l'utilisateur est directement dirigé au bon endroit, que l'unité ait le statut d'entrée principale ou de sous-entrée. Cependant, pour beaucoup d'unités, le problème subsiste. Ceci est démontré, par exemple, par une recherche des locutions *of course* et *let alone*. Mon choix s'est porté sur ces lexies à la lecture de Klotz & Herbst (2016: 119) qui comparent leur localisation dans quatre dictionnaires monolingues anglais « papier » et constatent de grandes différences. Ainsi, *of course* a le statut d'entrée principale dans deux des dictionnaires et est regroupé avec d'autres locutions idiomatiques au bas de l'article du lemme *course* dans les deux autres. Pour *let alone*, la présentation est encore plus hétérogène.

Il m'a semblé intéressant de rechercher ces deux items dans trois dictionnaires électroniques bilingues (Larousse¹, Reverso-Collins² et Oxford³) pour en mesurer la facilité d'accès. Les

¹ <http://www.larousse.fr/dictionnaires/anglais-francais>

² <http://dictionary.reverso.net/english-french/>

résultats révèlent des pratiques très différentes. Le Larousse exploite parfaitement les possibilités offertes par l'électronique et donne un accès direct aux deux locutions. Le dictionnaire Oxford fait de même pour *let alone*, mais pour *of course*, l'utilisateur est étonnamment renvoyé vers *clerk of the course* (= commissaire/directeur de courses); une recherche sur *course* permet néanmoins d'avoir accès à la locution adverbiale *of course* mais pour y parvenir, il faut passer en revue la très longue liste des autres emplois de *course* en tant que nom, verbe transitif et verbe intransitif. Le dictionnaire Reverso s'en sort le moins bien des trois. Une recherche sur *let alone* mène l'utilisateur directement vers le dictionnaire collaboratif, ce qui est loin d'être idéal étant donné que ce type de dictionnaire ne donne aucune garantie de qualité. Dans le dictionnaire proprement dit, la locution se retrouve, sans aucun surlignement particulier permettant de l'identifier aisément, tout en bas de l'entrée du lemme *alone*. La locution *of course* reçoit un traitement semblable.

Si on inverse la perspective et qu'on recherche des locutions françaises dans ces trois mêmes dictionnaires, on obtient des résultats tout aussi peu satisfaisants. Ainsi, par exemple, une recherche de la locution *bien sûr* dans le dictionnaire Oxford dirige l'utilisateur vers des occurrences surlignées en jaune dans l'entrée de *bien*. Cependant, les deux exemples (*les roses font bien sur la terrasse* = the roses look nice/good on the terrace; *on est bien sur cette chaise* = what a comfortable chair) ne contiennent pas la locution adverbiale *bien sûr* mais l'adverbe *bien* suivi de la préposition *sur*, une erreur due à une mauvaise gestion des accents. En faisant défiler les nombreux emplois de *bien*, l'utilisateur finit par trouver la locution recherchée, si du moins il n'a pas entretemps abandonné sa recherche.

Le constat est clair: dans de nombreux cas, la recherche d'unités phraséologiques relève du parcours du combattant, même dans les dictionnaires électroniques où l'on aurait pu s'attendre à une navigation rapide et aisée. Une des raisons est que le dictionnaire reste très largement basé sur l'analyse des lexies simples. Le nombre d'unités polylexicales faisant partie de la nomenclature est très minoritaire et les critères utilisés pour en décider sont très discutables. Il est important de souligner à ce propos que les dictionnaires en ligne ne contiennent que très peu d'informations d'ordre métalexicographique, sans doute parce que leur utilisation est supposée être intuitive, ce qui aurait pour effet de rendre ces informations superflues. Comme nous avons pu le constater, ce présupposé est infondé en ce qui concerne de nombreuses unités polylexicales.

Il ne faudrait cependant pas donner l'impression que tout est mauvais dans les dictionnaires bilingues. Je rejoins tout à fait Moon (2008) qui, bien que mettant en évidence des lacunes de ces dictionnaires, tient à insister sur leurs indéniables qualités. Il est clair que dans une perspective étendue de la phraséologie, il est particulièrement difficile de décider ce qui mérite d'être intégré dans un dictionnaire bilingue. Même dans un dictionnaire électronique, il n'est ni possible ni souhaitable de tout inclure. Néanmoins, des améliorations sont possibles et, comme suggéré dans la section suivante, une des pistes les plus prometteuses à cet égard est l'exploitation plus systématique de corpus de textes informatisés.

2. L'apport des corpus monolingues et parallèles

Les corpus qui se présentent de la manière la plus évidente comme utiles à l'amélioration des dictionnaires bilingues sont les corpus parallèles, qui contiennent des textes dans une langue, alignés avec leur traduction dans une autre. De nombreuses études ont démontré l'utilité de

³ <http://www.oxforddictionaries.com/>

ce type de ressource pour améliorer la qualité des dictionnaires bilingues. Cependant, le manque de corpus parallèles représentatifs des langues dans toute leur diversité réduit sensiblement leur impact. La bonne nouvelle est qu'en dépit de leurs limites, les corpus parallèles existants peuvent déjà se révéler d'une très grande utilité et qu'en outre, des progrès significatifs peuvent être réalisés en utilisant des corpus monolingues qui sont, eux, plus largement disponibles. Cette section a pour but de présenter brièvement l'apport de chacun de ces deux types de corpus.

2.1. Corpus monolingues

Dans leur article consacré aux dictionnaires phraséologiques bilingues, Lubensky & McShane (2007 : 919-920) identifient les principaux apports des corpus monolingues, dont je retiendrai les trois suivants:

- « detect common phrasemes, even if they do not fit into typical lists of idioms, defined as entities in which the sum of the parts does not equal the whole »
- « give a truly rich picture of the semantic and pragmatic contexts in which a phraseme can be used, as well as its combinatorial properties and co-occurrence patterns »
- « compare the frequency of phrasemes in various types of corpora – e.g. newswires versus literary prose »

Le premier apport se fonde sur la force heuristique des corpus. En appliquant des techniques d'extraction automatique, il est possible d'identifier dans les corpus des unités phraséologiques qui ne sont pas immédiatement accessibles à l'intuition des lexicographes. Une technique particulièrement efficace est l'extraction de segments répétés. Deux études récentes (Granger & Lefer 2012 & 2013) ont démontré que, même en se basant sur des corpus monolingues de petite taille (de l'ordre d'un million de mots), il est déjà possible de repérer un nombre significatif d'unités polylexicales qui présentent un réel intérêt lexicographique, mais ne figurent pas dans les dictionnaires bilingues ou n'y occupent pas la place qu'ils méritent.

Dans la première étude, nous avons extrait toutes les séquences de 2 à 5 mots qui avaient une fréquence d'au moins 20 occurrences dans le corpus Label France, un corpus d'un million de mots contenant des textes d'une revue du Ministère français des affaires étrangères⁴. Un tri manuel réalisé indépendamment par les deux auteurs a permis d'identifier, parmi les 6000 n-grammes extraits automatiquement, 425 séquences potentiellement utiles dans une perspective lexicographique bilingue. Parmi ces séquences figurent des unités qui ont une fonction référentielle (*diversité culturelle*), organisationnelle (*à cet égard*) ou modale (*il est vrai que*). Une recherche effectuée dans deux dictionnaires bilingues en ligne a révélé que 15% d'entre elles étaient totalement absentes de la partie français-anglais des dictionnaires (par exemple, *c'est pourquoi, qu'il s'agisse de, sans oublier*).

Dans la deuxième étude, nous avons extrait les séquences répétées contenant le mot *encore* et constaté à nouveau de grandes lacunes dans les dictionnaires bilingues consultés. Le premier constat que nous avons pu faire est que ce mot a de nombreux emplois phraséologiques: près de 50% des occurrences font partie de segments répétés plus longs. Le deuxième constat est que les dictionnaires ont tendance à inclure des emplois plus typiques de l'oral (*Encore toi!, Encore heureux!, Et encore! Quoi encore?*) et à négliger ceux qui sont plus typiques de l'écrit

⁴ Ce magazine a cessé de paraître en 2008.

(ou encore, là encore, et bien d'autres encore). Bien que ceci reste à confirmer, nous suggérons que cette tendance pourrait être liée au fait qu'on est plus conscient des expressions orales, qui ont souvent une forte composante émotionnelle, que des expressions propres à l'écrit qui, bien que très fréquentes, sont souvent plus banales et ont dès lors tendance à passer inaperçues. Alors que les premières sont accessibles à l'introspection des lexicographes parce qu'elles ont une forte « saillance cognitive » (terme emprunté à Hanks 2013: 343-344), pour les deuxièmes, qui ont une forte « saillance sociale » (ibid), le recours aux corpus semble indispensable. De manière générale, vu le nombre d'unités intéressantes mises en lumière grâce à l'extraction des séquences répétées, on peut se montrer optimiste sur les bénéfices que l'on pourrait retirer de l'application de la méthode à plus grande échelle, avec des corpus plus étendus et plus diversifiés.

Un deuxième bienfait des corpus monolingues est qu'ils permettent de décrire de manière extrêmement précise les contextes et spécificités d'emploi des unités polylexicales. Sans l'apport des corpus, la description tend à rester assez schématique, et ce d'autant plus que les unités phraséologiques sont souvent reléguées au statut de sous-entrée ou simplement introduites sous forme d'exemples dans les dictionnaires, limitant ainsi très fortement la quantité et la qualité des informations incluses (voir Granger & Lefer 2012 pour une analyse de cet aspect). Une chose qu'il est important de garder à l'esprit est que, comme le souligne (2008), les unités polylexicales peuvent aussi être polysémiques. Une analyse de corpus permet d'identifier ces différents sens et les traductions appropriées à chacun d'entre eux.

Comme troisième avantage, Lubensky & McShane mentionnent la possibilité d'extraire des corpus monolingues de précieux renseignements sur la fréquence d'emploi des unités phraséologiques dans différents registres. Pour ne prendre qu'un exemple, une recherche des locutions *d'ailleurs* et *par ailleurs* dans les textes originaux français du corpus PLECI⁵ (voir tableau 1) révèle que *d'ailleurs* est utilisé fréquemment dans les articles de presse et les romans, avec une fréquence deux fois plus élevée dans ces derniers, alors que *par ailleurs* est relativement fréquent dans la presse mais totalement absent du corpus de romans. Ces deux registres ne sont cependant pas suffisants. Pour avoir une vue complète des locutions, il est essentiel d'y adjoindre l'oral, pour lequel les corpus manquent cruellement. Dans ces circonstances, l'expertise du lexicographe reste essentielle car les corpus ne permettent pas encore de fournir toutes les informations nécessaires.

Locution	Corpus	Fréquence absolue	Fréquence relative (par million de mots)
<i>D'ailleurs</i>	PLECI-News	219	223.3
	PLECI-Fiction	174	500.6
<i>Par ailleurs</i>	PLECI-News	127	129.5
	PLECI-Fiction	0	0

⁵ Le corpus PLECI (Poitiers-Louvain Échange de Corpus Informatisés) est un corpus bilingue anglais↔français, compilé conjointement par le laboratoire FoReLL de l'Université de Poitiers et le CECL (Centre for English Corpus Linguistics) de l'Université catholique de Louvain, Belgique, sous la responsabilité d'Hélène Chuquet (Poitiers) et de moi-même. Il s'agit d'un corpus bi-directionnel qui comprend une partie littéraire (PLECI-Fiction) et une partie journalistique (PLECI-News). Je tiens à saisir cette occasion pour remercier vivement Hélène Chuquet pour la belle collaboration que nous avons eue et qui, je l'espère, se poursuivra encore pendant de nombreuses années!

Tableau 1: Fréquence de *d'ailleurs* et *par ailleurs* dans PLECI-News et PLECI-Fiction

2.2. Corpus parallèles

Les unités phraséologiques sont particulièrement sujettes aux calques. Moon (2008: 329-331) note une tendance à privilégier les traductions littérales, souvent non idiomatiques. C'est le cas, par exemple du proverbe *les petits ruisseaux font les grandes rivières* qui est traduit dans le dictionnaire Robert & Collins *Robert* par *little streams make big rivers*.

L'utilisation de corpus parallèles est porteuse d'améliorations significatives au niveau du nombre et de la qualité des équivalents de traduction. Elle permet de passer d'une méthodologie aléatoire que Hartmann (1994) qualifie de « hit-and-miss and trial-and-error » à une approche empirique solide. Une comparaison des équivalents de traduction inclus dans les dictionnaires bilingues avec les traductions réelles attestées par les corpus parallèles fait apparaître des différences très marquées entre les deux ensembles. En fait, la traduction des locutions nécessite souvent le recours à des unités fort éloignées structurellement des unités en langue source. Comme nous l'avons montré dans nos travaux antérieurs (Granger & Lefer 2012, 2013 et 2016), elles correspondent souvent à un mot simple plutôt qu'à une locution en langue cible (*de plus en plus* → *increasingly* ; *ou encore* → *and*). Elles impliquent aussi régulièrement un changement de catégorie grammaticale (*yet another* → *énième*) ou même le recours à des entités d'un autre ordre linguistique (*bel et bien* traduit par l'auxiliaire emphatique *do*, comme dans *la législation a bel et bien fonctionné*: *the legislation did work*). Bertels et al. (2009) constatent également que « les équivalents fournis dans les dictionnaires appartiennent (presque) toujours à la même catégorie grammaticale que le mot de départ » et préconisent le recours aux corpus parallèles afin de découvrir les nombreuses traductions « non couvertes par ce parti pris ». A ce propos, une attention particulière doit être portée aux locutions à valeur pragmatique, car comme l'indique Armstrong (2005), elles se prêtent rarement à une traduction littérale. Dans le cas de ces locutions, les corpus révèlent notamment de nombreux cas d'omission en langue cible. Lorsque ces cas sont fréquents, ils valent la peine d'être signalés dans les dictionnaires bilingues, mais dans les faits, c'est rarement le cas.

L'apport des corpus parallèles est indéniable. Cependant, on aurait tort de croire qu'ils constituent la panacée. Roberts & Montgomery (1996: 463) insistent sur les bienfaits que l'on peut retirer de l'utilisation de ces corpus, mais s'empressent d'ajouter que « corpora, like all lexicographic tools, are a lexicographer's best friend only if they are appropriately used ». Quand on travaille régulièrement avec les corpus parallèles, on ne peut que constater, avec Lubensky & McShane (2007), qu'ils présentent de nombreuses lacunes. Il convient de rester prudent et de vérifier dans des corpus de textes non traduits la fréquence et les contextes d'emploi des équivalents de traduction extraits des corpus parallèles. De telles comparaisons font souvent apparaître des différences très significatives. Ainsi, par exemple, une étude récente (Granger sous presse) réalisée sur la base du corpus Europarl a révélé que, si la locution *on the contrary* a une fréquence de 84.6 par million de mots dans les textes anglais traduits du français, la fréquence n'est que de 12.3 dans les textes du même corpus écrits originellement en anglais, soit sept fois moins.

3. Les locutions

3.1. Définition, catégorisation et traitement lexicographique

Une catégorie d'unités phraséologiques dont il conviendrait de se préoccuper en priorité est celle des locutions. Gaatone (1981 : 49) en donne la définition suivante : « La notion de locution désigne, en règle générale, un groupe de deux mots ou plus, ressenti intuitivement comme équivalant à un mot unique ». Gaatone va même jusqu'à suggérer que pour des séquences telles que *à condition que*, on devrait parler de « mot » plutôt que de « locution » : « pourquoi ne dirions-nous pas, par exemple, ignorant l'aspect purement graphique de la question, à savoir la présence d'espaces blancs séparant des éléments les uns des autres, que *à condition que* est un seul mot plutôt que de lui donner l'appellation de locution conjonctive, comme il est d'usage de le faire? ». Le terme de « locution » est néanmoins resté la norme, même si l'on rencontre parfois « words-with-spaces » pour faire référence aux séquences totalement figées syntaxiquement (Sag et al. 2002).

La définition donnée par Niklas-Salminen (2015) insiste également sur le caractère solidaire des locutions: « ces unités complexes, qui sont inscrites comme unités figées dans le code de la mémoire du sujet parlant et que l'on doit reproduire telles quelles, en bloc, se comportent dans leurs rapports avec les autres éléments de l'énoncé, exactement comme les unités simples ». L'auteur distingue les huit catégories suivantes:

- Locutions nominales: *un m'as-tu-vu, le qu'en dira-t-on*⁶
- Locutions verbales: *faire fi de, chercher noise*
- Locutions adjectivales: *aigre-doux, bon enfant, médico-social*
- Locutions adverbiales: *en effet, sur le champ, tout à coup, d'arrache-pied, à l'improviste*
- Locutions prépositives: *autour de, à cause de, en raison de, grâce à, au fur et à mesure*
- Locutions conjonctives: *afin que, de même que, pour la raison que, à supposer que, étant entendu que*
- Locutions interjectives: *nom de dieu !, bonté divine !, mais enfin !*

Bien que ces locutions se comportent dans les faits comme des mots simples, elles sont traitées de manière tout à fait différente de ceux-ci dans les dictionnaires bilingues, une différence majeure étant qu'elles n'apparaissent que rarement comme entrées principales, avec les conséquences néfastes décrites plus haut. À cet égard, je partage pleinement l'avis de Grezka (2011: 23) qui considère que « c'est à tort qu'on les fait figurer dans les dictionnaires sous l'un ou l'autre des termes. Elles devraient en effet constituer des entrées indépendantes ». Sur ce point, les locutions se distinguent nettement des collocations qui trouvent tout naturellement leur place dans l'entrée de leur base (par exemple, *importance capitale* sous le lemme *importance*).

Si l'on veut rendre mieux compte de ce type d'unité phraséologique dans les dictionnaires, la première chose à faire est d'en dresser la liste la plus complète possible. Cette tâche est moins ardue qu'on pourrait le penser de prime abord étant donné que la technique des segments répétés a précisément comme effet d'extraire les séquences invariantes. Étonnamment, cette technique ne semble guère avoir été utilisée pour la confection des dictionnaires bilingues, peut-être parce que les locutions sont moins au centre des préoccupations des lexicographes que d'autres unités phraséologiques telles que les idiomes purs ou les collocations. Par contre, dans une perspective de traitement automatique du langage (TAL), des ressources électroniques ont été constituées ou sont en cours de constitution et pourraient utilement être exploitées en lexicographie non orientée vers le TAL. Par exemple, Laporte & Voyatzi (2008)

⁶ Contrairement à d'autres linguistes, Niklas-Salminen ne classe que les locutions nominales à trois termes comme locutions ; les unités à deux termes sont rangées parmi les noms composés.

ont constitué un lexique électronique d'adverbes français polylexicaux qui comprend 6800 unités fixes, semi-fixes et flexibles. D'autres travaux pertinents sont ceux de l'équipe de Nancy, dirigée par A. Polguère, qui travaille à la construction du Réseau Lexical du Français, une ressource lexicographique de nouvelle génération qui est fortement centrée sur la phraséologie (Lux-Pogodalla & Polguère 2011, Polguère à paraître), ainsi que ceux de Grezka (2011 : 17), dont l'objectif est de réaliser un inventaire exhaustif des séquences complètement figées (adverbiales, adjectivales, nominales et verbales).

Une fois que la liste des locutions totalement figées syntaxiquement est dressée, les locutions peuvent être incorporées à la nomenclature des dictionnaires bilingues et recevoir une description lexicographique approfondie au même titre que les mots simples. Grâce à l'apport conjoint des corpus monolingues et bilingues, il est alors possible de rendre compte des différents sens qu'elles peuvent exprimer et des différences stylistiques qu'elles peuvent présenter, et de proposer des équivalents de traduction appropriés aux divers contextes d'emploi. Dans la section suivante, la démarche est illustrée pour deux locutions adverbiales: *de la même manière* et *de la même façon*.

3.2. Les locutions adverbiales *de la même manière* et *de la même façon*

Les locutions *de la même manière* et *de la même façon* appartiennent à une catégorie très importante, celle des locutions adverbiales. Grezka (2011: 18) range sous cette appellation « toutes les formes adverbiales non soudées présentant un certain 'degré' de figement, quelle qu'en soit la présentation » et en dénombre pas moins de 5000 locutions adverbiales, tout en reconnaissant que « les constructions complètement figées sont en nombre relativement limité » (p. 26). Il n'est pas possible à ce stade de donner de chiffre précis, mais à titre indicatif, on peut faire référence au Réseau Lexical du Français⁷ en cours de construction (Polguère à paraître), qui dénombre approximativement 500 locutions à fonction adverbiale, ou au dictionnaire en ligne L'Internaute⁸ qui en dénombre aux alentours de 370⁹. Un grand nombre de ces locutions sont complètement figées et auraient donc tout à fait leur place dans la nomenclature des dictionnaires bilingues français-anglais, d'autant plus que bon nombre d'entre elles sont polyfonctionnelles et donc, difficiles à traduire. Le tableau 2 présente un échantillon représentatif de ces locutions.

<i>à cet effet</i>	<i>bien sûr</i>	<i>en général</i>
<i>à contrecœur</i>	<i>coup sur coup</i>	<i>en même temps</i>
<i>à plein régime</i>	<i>d'ailleurs</i>	<i>en outre</i>
<i>à première vue</i>	<i>dans l'ensemble</i>	<i>en particulier</i>
<i>à propos</i>	<i>dans les faits</i>	<i>en premier lieu</i>
<i>à tour de rôle</i>	<i>de la même façon</i>	<i>en réalité</i>
<i>à tous les coups</i>	<i>de la même manière</i>	<i>en somme</i>
<i>à tout hasard</i>	<i>de nos jours</i>	<i>en tout cas</i>
<i>à vue de nez</i>	<i>de plus en plus</i>	<i>en travers</i>
<i>à la fois</i>	<i>depuis toujours</i>	<i>malgré tout</i>
<i>à la longue</i>	<i>en d'autres termes</i>	<i>par mégarde</i>
<i>à nouveau</i>	<i>en quelque sorte</i>	<i>par ailleurs</i>
<i>à toutes fins utiles</i>	<i>en revanche</i>	<i>par chance</i>

⁷ Je tiens à adresser mes vifs remerciements à Alain Polguère qui m'a transmis les données du Réseau Lexical du Français actuellement disponibles concernant les locutions.

⁸ <http://www.linternaute.com/dictionnaire/fr/type/location-adverbiale/1/>

⁹ Le dictionnaire liste 425 locutions adverbiales mais après vérification, il apparaît que plusieurs d'entre elles n'appartiennent pas à cette catégorie.

<i>au demeurant</i>	<i>en temps et heure</i>	<i>par défaut</i>
<i>au fait</i>	<i>en tout point</i>	<i>quand même</i>
<i>au premier abord</i>	<i>en vain</i>	<i>sans encombre</i>
<i>au contraire</i>	<i>en amont</i>	<i>sans façon</i>
<i>au petit bonheur</i>	<i>en conséquence</i>	<i>sous peu</i>
<i>avant tout</i>	<i>en effet</i>	<i>sur place</i>
<i>bel et bien</i>	<i>en fait</i>	<i>un à un</i>

Tableau 2 : Echantillon représentatif de locutions adverbiales

Pour illustrer la démarche, j'ai recherché deux de ces locutions, *de la même manière* et *de la même façon* dans les trois dictionnaires électroniques cités plus haut, à savoir Reverso, Larousse et Oxford. Comme le montre le tableau 3, les dictionnaires font preuve d'une belle unanimité: ils présentent tous trois *in the same way* comme seule et unique traduction des deux locutions. Pour *de la même façon*, et uniquement pour cette locution-là, le dictionnaire Oxford contient un peu plus d'informations que les deux autres dictionnaires. Il mentionne la possibilité d'utiliser la locution comme locution conjonctive en mettant la conjonction *que* entre parenthèses, mais en maintenant la même traduction. Il ajoute aussi une extension de la locution (*à peu près de la même façon*) et inclut un exemple qui donne lieu à une traduction tout à fait différente (*to do the same*). Cette présentation sommaire (un seul équivalent de traduction, pas de distinction entre les différents emplois des locutions et pas d'exemples) est représentative du traitement des locutions dans leur ensemble, particulièrement des locutions adverbiales, prépositives et conjonctives.

	Reverso	Larousse	Oxford
de la même manière	<i>in the same way</i>	<i>in the same way</i>	(travailler) <i>in the same way</i>
de la même façon	<i>in the same way</i>	<i>in the same way</i>	de la même façon: <i>in the same way (que as)</i> à peu près de la même façon : <i>in much the same way (que as)</i> agir de la même façon : <i>to do the same</i>

Tableau 3: Traduction de de la même manière et de la même façon dans trois dictionnaires bilingues français-anglais

Pour la deuxième étape de l'analyse, j'ai recherché les deux locutions dans des corpus de français¹⁰, ce qui m'a permis d'identifier trois emplois différents :

- locution adverbiale de manière (exemples 1 et 2)
- locution adverbiale de liaison (exemples 3 et 4). Il s'agit ici de connecteurs logiques exprimant la mise en parallèle de deux idées ou entités, qui sont souvent en position initiale dans la phrase, suivis d'une virgule.
- locution conjonctive : *de la même façon/manière que*, suivi d'un syntagme nominal ou d'une proposition (exemples 5 et 6).

(1) De face, Jean-Charles n'est plus tout à fait le même – elle ne le voit plus *de la même manière*. (PLECI-Fiction)

(2) Or j'ai tendance, personnellement, à m'exprimer *de la même façon* quel que soit le public. (Label France)

¹⁰ Pour cette recherche, j'ai utilisé les textes français originaux des corpus suivants : Europarl, PLECI-Fiction, PLECI-News, Label France et Scientext.

- (3) *De la même façon*, il ne nous semble pas équitable de suggérer, comme le faisait le texte libéral initial, un boycott européen du pétrole et du diamant congolais. (Europarl)
- (4) *De la même manière*, les dirigeants du Parti travailliste ont été obligés de réinventer une mystique politique pour s'assurer du soutien de l'électorat (PLECI-News)
- (5) Là encore je n'ai pas lu les résultats *de la même façon* que Caroline (Europarl)
- (6) la lexicalisation du terme réduit suppose (...) qu'il soit employé par les locuteurs *de la même manière* que ceux-ci emploient le terme complexe (Scientext)

La troisième étape, qui consiste à extraire les équivalents de traduction des corpus parallèles français-anglais¹¹, démontre qu'il est indispensable de distinguer clairement ces trois usages afin d'identifier les traductions appropriées. Si la traduction *in the same way* se révèle être la traduction la plus courante lorsque les locutions sont employées comme adverbes de manière (exemple 7), d'autres équivalents de traduction existent et sont souvent préférables dans les deux autres emplois.

Pour l'emploi en tant que connecteur logique, par exemple, outre *in the same way*, les corpus mettent notamment en évidence les traductions suivantes: *similarly*, *likewise*, *by the same token* et *also*. Du point de vue de la fréquence, deux traductions se détachent du lot, *in the same way* et *similarly*, qui ont une fréquence équivalente (exemples 8 et 9). On peut se demander si la fréquence relativement élevée de *in the same way*, qui a un nombre d'occurrences équivalant à celui de *similarly* dans les traductions, n'est pas un effet de calque. Pour vérifier cette intuition, j'ai comparé la fréquence des deux lexies orthographiées avec majuscule et suivies d'une virgule dans un sous-corpus académique du British National Corpus de 15 millions de mots. La différence de fréquence entre les deux emplois est frappante: 62 occurrences par million de mots pour *similarly* contre seulement 2.2 occurrences pour *in the same way*. On a donc ici une confirmation éclatante que, si les corpus parallèles sont d'une grande utilité pour découvrir des traductions possibles, il ne faut pas se fier à leur fréquence respective en langue cible car même des traducteurs professionnels dont la langue cible est la langue maternelle peuvent être influencés par la locution présente en langue source. La tentation est particulièrement grande quand, et c'est souvent le cas, une traduction littérale existe en langue cible. Il est donc sage de vérifier les fréquences dans de larges corpus de langue non traduite. Ceci est particulièrement important pour des unités telles que les locutions adverbiales, qui ont tendance à avoir de nombreux équivalents de traduction.

Pour l'emploi de *de la même manière/façon que*, les corpus révèlent, outre la traduction *in the same way as* donnée par les dictionnaires bilingues, plusieurs autres traductions, en particulier *just as* (ou simplement *as*) qui mérite clairement d'être inclus dans le dictionnaire (exemples 10 et 11).

- (7) Elle devrait donc s'appliquer *de la même manière* aux quinze membres de l'Union (Europarl)
It should, therefore, apply *in the same way* to the fifteen members of the Union.
- (8) *De la même façon*, notre proposition d'une taxation des billets d'avion a suscité l'intérêt de plusieurs partenaires européens (Label France)
In the same way, our proposal for a levy on air fares has aroused the interest of several European partners.
- (9) *De la même manière*, vous avez tout à fait raison de souligner un problème essentiel au niveau européen (Europarl)
Similarly, you are quite right to highlight a crucial problem at European level
- (10) ...les Gallois ou les Ecossais (...) ont été sanctionnés *de la même façon que* les producteurs anglais (Europarl)
...the Welsh and the Scots (...) have been punished *in the same way as* English producers

¹¹ Les corpus suivants ont été utilisés : PLECI-News, PLECI-Fiction, Label France et Europarl (uniquement les textes avec le français comme langue source).

- (11) Et, *de la même façon que* les entreprises se battent pour allonger la liste de leurs clients, les bergers évangéliques entrent en concurrence pour attirer le plus de fidèles possibles (PLECI-News)
Just as businesses compete for clients, so do evangelicals for followers.

Les deux locutions sont représentatives d'une tendance plus générale qui caractérise le passage du français à l'anglais, à savoir ce que Chuquet & Paillard (1987: 14-15) appellent la « dilution ». Ce procédé a toute sa place dans les dictionnaires bilingues car, contrairement à un procédé de traduction tel que l'étoffement, la dilution « n'est pas un procédé de traduction mais une simple équivalence lexicale entre un élément simple et un élément composé ». Le phénomène s'applique à de nombreuses locutions prépositives (*de l'autre côté du chemin: across the track ; vols à destination de Delhi: flights to New-Delhi; au large de Land's End: off Land's End*) et est également à l'œuvre dans les locutions adverbiales qui correspondent souvent à des adverbes en *-ly* en anglais (*de plus en plus: increasingly; au fond: basically*) (ibid: 193).

4. Conclusion

L'extension rapide du domaine de la phraséologie est en passe de changer fondamentalement notre perception du fonctionnement du langage. Beaucoup aujourd'hui estiment que ce sont les phraséologismes plutôt que les lexies simples qui constituent le cœur même du langage. Il est essentiel que ce changement profond se répercute dans les ressources langagières les plus utilisées, parmi lesquelles le dictionnaire bilingue occupe une place de choix. Même si les dictionnaires monolingues sont aujourd'hui, à plus d'un égard, plus sophistiqués d'un point de vue phraséologique, de nombreuses études prouvent que c'est vers les dictionnaires bilingues que les utilisateurs préfèrent se tourner lorsqu'ils doivent écrire dans une langue étrangère (Lew 2004).

Une remise en question de la notion même d'unité phraséologique dans les dictionnaires bilingues s'impose. Pour les rendre plus efficaces, particulièrement pour les activités d'encodage, il convient d'adopter une notion plus large car, sans être nécessairement phraséologiques au sens strict, de nombreuses unités polylexicales posent de grandes difficultés de traduction et méritent d'être décrites avec le même degré de précision que les lexies simples. Il en va ainsi des locutions qui ne reçoivent souvent qu'un traitement sommaire dans les dictionnaires bilingues et qui, en outre, sont souvent difficiles à repérer, même dans les dictionnaires électroniques.

Plus que de longues discussions pour savoir si une unité est ou non phraséologique, ce dont on a le plus besoin dans une perspective lexicographique bilingue, ce sont des analyses fines des unités dans les deux langues concernées (Gross 2005). Les corpus de textes informatisés constituent la ressource par excellence pour atteindre cette précision, même si, on ne le soulignera jamais assez, l'expertise bilingue de lexicographes chevronnés reste essentielle.

Alors qu'on peut à la rigueur se passer des corpus pour traduire des expressions de sens clairement non compositionnel, tels que les idiomes purs ou les proverbes, pour les unités partiellement ou totalement compositionnelles, le recours aux corpus est indispensable. Sans cet apport, les collocations incluses ne sont qu'un pâle reflet de la réalité; une présentation beaucoup plus complète peut être obtenue par le recours systématique à des outils d'extraction automatique de collocations tels que WordSketch¹², qui ont fait leurs preuves

¹² <https://www.sketchengine.co.uk/user-guide/user-manual/word-sketch/>

pour l'élaboration des dictionnaires monolingues anglais. En ce qui concerne les locutions, on observe une tendance à ne pas prendre en compte leur polyfonctionnalité et à privilégier un nombre très restreint de traductions, souvent très proches syntaxiquement des unités en langue source. Comme j'ai tenté de le démontrer, les corpus monolingues permettent de mettre en évidence les spécificités des locutions en matière de sens, de fréquence et de restrictions, notamment au niveau du registre. Les corpus parallèles, quant à eux, révèlent une riche gamme de traductions qui, contrairement aux traductions privilégiées dans les dictionnaires, se révèlent souvent être assez éloignées formellement des unités en langue source.

Le chantier est énorme et il est impossible de tout réaliser, et en tout cas, pas tout de suite. Que ce soit au niveau de la sélection des unités phraséologiques, de leur localisation dans la macrostructure et la microstructure du dictionnaire ou du choix des traductions à inclure, il y a un énorme travail d'analyse à réaliser. Comme première étape, je suggère de commencer par les locutions invariantes, particulièrement adverbiales, adjectivales, prépositives et conjonctives¹³, qui fonctionnent comme des mots simples et devraient, comme eux, avoir le statut d'entrée principale. Pour ce faire, outre l'application de la technique des segments répétés à de larges corpus monolingues, il convient de promouvoir l'interaction avec les linguistes spécialisés en TAL qui travaillent à l'élaboration de ressources lexicales électroniques potentiellement très utiles aux lexicographes. Une collaboration avec les phraséologues serait également très bénéfique afin d'assurer une meilleure appréhension des phénomènes phraséologiques. Ce n'est qu'au prix d'une étroite synergie entre lexicographes, phraséologues et talistes que de réelles avancées se feront jour dans le domaine passionnant mais éminemment complexe de la phraséologie.

Dictionnaires

Dictionnaire Larousse français-anglais anglais-français. 2007. Paris: Larousse.

Le Robert & Collins. Dictionnaire français-anglais anglais-français. 8e édition. 2006. Paris: Dictionnaires Le Robert

References

ARMSTRONG, Nigel, 2005, *Translation, Linguistics, Culture: A French-English Handbook*, Clevedon, Multilingual Matters.

BEJOINT, Henri & THOIRON, Philippe, 1996, « Introduction » in H. BEJOINT & P. THOIRON (éds.), *Les dictionnaires bilingues*, Louvain-la-Neuve, Duculot, 5-15.

BERTELS, Anne, FAIRON, Cédric, TIEDEMANN, Jörg & VERLINDE, Serge, 2009, « Corpus parallèles et corpus ciblés au secours du dictionnaire de traduction », *Cahiers de Lexicologie* 94-1, 199-219.

BIBER, Doug, JOHANSSON, Stig, LEECH, Geoffrey, CONRAD, Susan & FINEGAN, Edward, 1999, *Longman Grammar of Spoken and Written English*, Harlow, Pearson Education.

CHUQUET, Hélène & PAILLARD, Michel, 1987, *Approche linguistique des problèmes de traduction anglais-français*, Paris/Gap, Ophrys.

DE GIOVANNI, Cosimo, 2014, « Pragmatique et didactique du dictionnaire : Quelques réflexions terminologiques », *Verbum* 1, 27-36.

¹³ Les locutions nominales doivent faire l'objet d'un traitement à part conjointement aux mots composés, de même que les locutions verbales, qui sont d'une grande complexité et dont le statut-même en tant que locution est remis en cause par Gaatone (1981).

- FILLMORE, Charles & ATKINS, Beryl, 1994, « Starting where the dictionaries stop: the challenge of corpus lexicography », in B.T.S. ATKINS & A. ZAMPOLLI (éds.), *Computational Approaches to the Lexicon*, Oxford, Oxford University Press, 349-393.
- GAATONE, David, 1981, « Les 'locutions verbales': Pour quoi faire? », *Revue Romane* 16, 49-73.
- GRANGER, Sylviane, sous presse, « Tracking the third code: A cross-linguistic corpus-driven approach to metadiscursive markers » in A. CERMAKOVA & M. MAHLBERG (éds.), *Corpus as Discourse*, Amsterdam, John Benjamins.
- GRANGER, Sylviane & LEFER, Marie-Aude, 2012, « Towards more and better phrasal entries in bilingual dictionaries », *Euralex 2012 Proceedings*, August 2012, Oslo, 682-692.
- GRANGER, Sylviane & LEFER, Marie-Aude, 2013, « Enriching the phraseological coverage of high-frequency adverbs in English-French bilingual dictionaries » in K. AJMER & B. ALTENBERG (éds.), *Advances in Corpus-based Contrastive Linguistics*, Amsterdam & Philadelphia, Benjamins, 157-176.
- GRANGER, Sylviane & LEFER, Marie-Aude, 2016, « From general to learners' bilingual dictionaries: Towards a more effective fulfilment of advanced learners' phraseological needs », *International Journal of Lexicography*, doi: 10.1093/ijl/ecw022.
- GRANGER Sylviane & PAQUOT, Magali, 2008, « Disentangling the phraseological web » in S. GRANGER & F. MEUNIER (éds.), *Phraseology: An Interdisciplinary Perspective*, Amsterdam & Philadelphia, Benjamins, 27-49.
- GREZKA, Aude, 2011, « La base de données Figement », *Neophilologica* 23, 15-28.
- GROSS, Gaston, 2005, « Réflexions sur le figement » in C. BOLLY, J. KLEIN & B. LAMIROY (éds.), *La phraséologie dans tous ses états*, Louvain-la-Neuve, Cahiers de Linguistique de Louvain 31 2-4, 45-61.
- HANKS, Patrick, 2013, *Lexical Analysis. Norms and Exploitations*. Cambridge & London, The MIT Press.
- HARTMANN, Reinhard, 1994, « The use of parallel text corpora in the generation of translation equivalents for bilingual lexicography », *Euralex 1994 Proceedings*, 291-297.
- KASTBERG SJÖBLUM, Margareta, 2013, « Collocations et cooccurrences dans le dictionnaire bilingue: étude lexicométrique », *Ela. Etudes de linguistique appliquée* 170, 207-225.
- KLEIN, Jean & LAMIROY, Béatrice, 2005, « Relations systématiques entre expressions verbales figées à travers quatre variétés de français » in C. BOLLY, J. KLEIN & B. LAMIROY (éds.), *La phraséologie dans tous ses états*, Louvain-la-Neuve, Cahiers de Linguistique de Louvain 31 2-4, 77-92.
- KLEIN, Jean & LAMIROY, Béatrice, 2011, « Routines conversationnelles et figement » in J.-C. ANSCOMBRE & S. MEJRI (éd.), *La parole entravée: études sur le figement*, Paris, Champion, 195-213.
- KLEIN, Jean & LAMIROY, Béatrice, 2016, « Le figement : unité et diversité. Collocations, expressions figées, phrases situationnelles, proverbes », *L'information grammaticale* 148, 15-20.
- KLOTZ, Michael & HERBST, Thomas, 2016, *English Dictionaries. A Linguistic Introduction*, Berlin, Erich Schmidt Verlag.
- LAPORTE, Eric & VOYATZI, Stavroula, 2008, « An Electronic Dictionary of French Multiword Adverbs. Language Resources and Evaluation Conference », Workshop Towards a Shared Task for Multiword Expressions, 2008, Marrakech, Morocco, 31-34.
- LEGALLOIS, Dominique & TUTIN, Agnès, 2013, « Présentation: Vers une extension du domaine de la phraséologie », *Langages* 189, 3-25.
- LEW, Robert, 2004, *Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-Bilingual Dictionaries by Polish Learners of English*, Poznań, Motivex.

- LUBENSKY, Sophia & MCSHANE, Marjorie, 2007, « Bilingual Phraseological Dictionaries » in H. BURGER, D. DOBROVOL'SKIJ, P. KÜHN & N.R. NORRICK (éds.), *Phraseology. An International Handbook of Contemporary Research*, Berlin and New York, de Gruyter, 919-928.
- LUX-POGODALLA Veronika & POLGUÈRE Alain, 2011, « Construction of a French Lexical Network: Methodological Issues », Proceedings of the International Workshop on Lexical Resources (WoLeR 2011), ESSLLI, Ljubljana, 1-5 août 2011.
- MOGORRON HUERTA, Pedro, 2008, « Compréhension et traduction des locutions verbales », *Meta : journal des traducteurs / Meta: Translators' Journal* 53-2, 378-406.
- MOON, Rosamund, 2008, « Dictionaries and collocation » in S. GRANGER & F. MEUNIER (éds.), *Phraseology. An Interdisciplinary Perspective*, Amsterdam & Philadelphia, Benjamins, 313-336.
- MULHALL, Chris, 2007, « Semantics and the lemmatisation of idiomatic expressions: an analysis of bilingual Italian-English dictionaries ». Proceedings of LangUE 2007. University of Essex. http://www.essex.ac.uk/langling/documents/langue/langue_2007_proceedings.pdf
- NEE, Emilie, SITRI, Frédérique & VENIARD, Marie, 2014, « Pour une approche des routines discursives dans les écrits professionnels », *SHS Web of Conferences* 8, 113-2124.
- NIKLAS-SALMINEN, Aino, 2015, *La lexicologie* - 2e édition, Armand Colin.
- PECMAN, Mojca, 2008, « Compilation, formalisation and presentation of bilingual phraseology: problems and possible solutions » in F. MEUNIER & S. GRANGER (éds.), *Phraseology in Foreign Language Learning and Teaching*, Amsterdam & Philadelphia, Benjamins, 203-221.
- POLGUERE, Alain, 2008, *Lexicologie et sémantique lexicale: Notions fondamentales*, Les Presses de l'Université de Montréal.
- POLGUERE, Alain, à paraître, « Non-compositionnalité: ce sont toujours les locutions faibles qui trinquent ». *Verbum*.
- ROBERTS, Roda & MONTGOMERY, Catherine, 1996, « The use of corpora in bilingual lexicography » in M. GELLERSTAM *et al.* (éds.), *Euralex '96 Proceedings Part II*, Göteborg, Göteborg University, 457-464.
- SAG, Ivan, BALDWIN, Timothy, BOND, Francis, COPESTAKE, Ann & FLICKINGER, Dan, 2002, « Multi-word expressions: A pain in the neck for NLP », Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (*CICLING 2002*), Mexico City, 1-15.
- SIEPMANN, Dirk, 2008, « Phraseology in learners' dictionaries. What, where and how? » in S. GRANGER & F. MEUNIER (éds.), *Phraseology. An Interdisciplinary Perspective*, Amsterdam & Philadelphia, Benjamins, 185-202.
- TSCHICHOLD, Cornelia, 2008, « A computational lexicography approach to phraseologisms » in S. GRANGER & F. MEUNIER (éds.), *Phraseology. An Interdisciplinary Perspective*, Amsterdam & Philadelphia, Benjamins, 361-376.
- TUTIN, Agnès, 2014, « La phraséologie transdisciplinaire des écrits scientifiques: des collocations aux routines sémantico-rhétoriques » in A. TUTIN & F. GROSSMANN (éd.), *L'écrit scientifique: du lexique au discours. Autour de Scientext*, Rennes, Presses Universitaires de Rennes, 27-44.