ASp

la revue du GERAS

1 | 1993

**Actes du 13e colloque du GERAS et de l'atelier Langue
de spécialité du 32e congrès de la SAES**

Création de dictionnaires bilingues spécialistes : l'épreuve de vérité du retournement informatique

Development of bilingual scientific dictionaries: the file reversal as an acid test

Martine Schuwer

Édition électronique

URL : <http://journals.openedition.org/asp/4404>

DOI : 10.4000/asp.4404

ISSN : 2108-6354

Éditeur

Groupe d'étude et de recherche en anglais de spécialité

Édition imprimée

Date de publication : 1 mars 1993

Pagination : 531-546

ISSN : 1246-8185

Référence électronique

Martine Schuwer, « Création de dictionnaires bilingues spécialistes : l'épreuve de vérité du retournement informatique », *ASp* [En ligne], 1 | 1993, mis en ligne le 04 juin 2014, consulté le 19 avril 2019. URL : <http://journals.openedition.org/asp/4404> ; DOI : 10.4000/asp.4404

Ce document a été généré automatiquement le 19 avril 2019.

Tous droits réservés

Création de dictionnaires bilingues spécialistes : l'épreuve de vérité du retournement informatique

Development of bilingual scientific dictionaries: the file reversal as an acid test

Martine Schuwer

- 1 La présente étude constitue la deuxième partie de la méthodologie de création de dictionnaires bilingues spécialisés¹, rédigée à l'intention d'auteurs indépendants, lexicographes débutants, désireux de réaliser un véritable dictionnaire, c'est-à-dire plus qu'une simple nomenclature bilingue.
- 2 La première partie de cette méthodologie était consacrée à l'élaboration de la version anglais-français, établie à l'aide d'un logiciel de gestion de base de données². Elle a d'ailleurs été appliquée à la création d'un dictionnaire bilingue de l'écologie et de l'environnement (Collin & Schuwer 1992).
- 3 Réaliser la version français-anglais d'un dictionnaire à partir de la version anglais-français peut sembler une opération de routine consistant, en substance, à une réorganisation des données.
- 4 Restituons brièvement les conditions d'élaboration de la première version.
- 5 Le masque de saisie d'un mot-vedette comporte *x* rubriques, dont trois à remplir obligatoirement (mot - catégorie grammaticale - traduction) ; chaque mot-vedette peut avoir un nombre infini de sous-entrées, chacune d'elle ayant son propre masque comportant exactement les mêmes rubriques.
- 6 L'opération de retournement consiste donc à intervertir l'ordre des champs, ceux de la langue de départ étant transférés à la place de ceux de la langue d'arrivée et inversement.

Exemples :

ABSORB *v (of a solid)* absorber
ABSORBABLE *adj* absorbable
ABSORBENT *n* absorbant *m*
ABSORPTION *n* absorption *f*

qui devient dans la version français-anglais :

ABSORBER *v* absorb
 ABSORBABLE *adj* absorbable
 ABSORBANT *m* absorbent
 ABSORPTION *f* absorption

- 7 À ce stade, l'hypothèse selon laquelle le retournement se résume à un simple transfert se trouve confirmée, à l'exception près qu'il faudra modifier l'ordre de présentation des termes pour respecter l'ordre alphabétique : le mot-vedette sera **ABSORBABLE**, suivi de ses sous-entrées **ABSORBANT**, **ABSORBER**, **ABSORPTION**.
- 8 Cependant, un retournement automatique peut générer d'étranges équivalences, et ce pour chaque élément constitutif de l'article (dérivés, lexies composées/expressions, exemples) dont la structure se présente ainsi :

MOT-VEDETTE
 - lexies composées/expressions
 - exemples
 - SOUS-ENTREE
 - lexies composées/expressions
 - exemples
 - SOUS-ENTREE
 - lexies composées/expressions
 - exemples
 - SOUS-ENTREE, etc.

Ainsi se présente **SUN** et ses sous-entrées³ :

SUN *n* soleil *m*
 SUNSET *n* coucher *m* du soleil
 SUNBURN *n* coup *m* de soleil
 SUN PLANT *n* héliophyte *f*
 SUNSTROKE *n* insolation *f*
 SUNRISE *n* lever *m* du soleil
 SUNGLASSES *pl* lunettes *fpl* de soleil
 SUNSPOTS *pl* taches *fpl* solaires

dont le retournement aboutit à la présentation suivante :

SOLEIL *m* sun
 - COUCHER *m* DU SOLEIL : sunset
 - COUP *m* DE SOLEIL: sunburn
 - HELIOPHYTE *f* sun plant
 - INSOLATION *f* sunstroke
 - LEVER *m* DU SOLEIL sunrise
 - LUNETTES *fpl* DE SOLEIL sunglasses
 - TACHES SOLAIRES *fpl* sunspots

- 9 À propos de cet exemple, rappelons qu'en sous-entrée d'un mot-vedette, figurent en principe les dérivés, au sens strict ou au sens large. Par dérivés au sens large, on entend lexies qui figurent en entrée principale d'un dictionnaire unilingue⁴.
- 10 Dans cet article, on constate qu'aucun des mots figurant en sous-entrée n'est un dérivé (strict ou large) de **SOLEIL** : cinq sont des lexies composées, tandis que **HELIOPHYTE** et **INSOLATION** devraient figurer en entrée principale. Le produit du retournement ne peut donc être laissé tel quel. On observe le même phénomène lors du retournement des lexies composées.

Ainsi l'article **OIL** :

OIL *n* pétrole *m*
 OIL POLLUTION *n* pollution *f* aux hydrocarbures

OIL SLICK *n* nappe *f* de pétrole

OIL TANKER *n* pétrolier *m*

OIL RIG *n* plate-forme *f* pétrolière

deviendrait dans la version français-anglais :

PETROLE *m* oil

POLLUTION *f* AUX HYDROCARBURES : oil pollution

NAPPE *f* DE PETROLE oil slick

PETROLIER *m* oil tanker

PLATE-FORME *f* PETROLIERE oil rig

- 11 Seule une des quatre lexies composées françaises (NAPPE DE PÉTROLE) pourrait rester sous le mot-vedette (PÉTROLE) ; la première (POLLUTION AUX HYDROCARBURES) devrait figurer sous POLLUTION ou HYDROCARBURES ; PÉTROLIER, dérivé strict de PÉTROLE, en serait une des sous-entrées ; PLATE-FORME PÉTROLIÈRE, lexie composée, pourrait figurer sous PÉTROLIER, -ÈRE (*adj*) ou sous PLATE-FORME...
- 12 Le retournement des exemples donne également lieu à des associations incohérentes ; ce point sera traité dans la deuxième partie de cette étude.
- 13 Cette analyse démontre que la validité opérationnelle d'un retournement automatique est directement fonction de la quantité et de la diversité des informations à transférer.
- 14 En réalité, afin d'éviter ce type d'incohérences, il est recommandé, au stade de la saisie de la première version, de placer une balise devant chaque terme dont on sait qu'il devra être déplacé.
- 15 Une balise /V/ devant un mot qui devra figurer en **mot-vedette** (tel/V/HELIOPHYTE ou/V/INSOLATION). Une balise /L/ devant le terme-référence d'une **lexie composée** (ainsi /L/ POLLUTION AUX HYDROCARBURES, qui pourrait aussi être POLLUTION AUX /L/ HYDROCARBURES). Une balise /D/ devant les **dérivés**, avec l'indication du mot-vedette sous lequel il devra se placer : ainsi /D/ TERRESTRE -> TERRE. Une balise /E/ devant le mot-clé de l'**exemple** indiquant quel mot-vedette il devra venir illustrer (LES CHAUDIÈRES FONCTIONNENT AU /E/ GAZ NATUREL). Lors du retournement informatique, les mots précédés de balises iront donc se placer à l'endroit indiqué.
- 16 Prendre le temps de placer ces repères au stade préparatoire, c'est accroître la fiabilité du retournement, et limiter ainsi le nombre de manipulations ultérieures. Pour lourdes qu'elles puissent paraître, ces opérations de balisage sont néanmoins obligatoires compte tenu de la microstructure dictionnaire (organisation en sous-entrées, lexies composées, exemples) ; elles présentent cependant l'avantage d'être applicables de façon méthodique et objective, car les problèmes à résoudre, directement observables, se situent au niveau morphologique.
- 17 Cependant, même si on a pris toutes ces précautions, on constatera que la version retournée comporte encore bon nombre d'articles dont le contenu est insatisfaisant, le plus souvent parce qu'il est incomplet.
- 18 Comment se fait-il qu'à partir d'une « bonne » première version, on obtienne un produit aussi approximatif ?
- 19 Afin de tenter de déterminer ce qui, dans la démarche du lexicographe est à l'origine des imperfections constatées, nous examinerons essentiellement deux points : d'une part la validité des équivalences proposées, d'autre part les exemples.
- 20 On conçoit généralement que dans un dictionnaire bilingue spécialisé, à un terme correspond un terme (et un seul) dans l'autre langue. C'est tout d'abord moins

systématique qu'on ne l'imagine et de toute façon, ce n'est pas face à ces mots que le lexicographe, et donc l'utilisateur, rencontre des difficultés, mais plutôt face à ceux utilisés à la fois dans la langue usuelle et dans la spécialité. Rappelons à cet égard que la nomenclature d'un dictionnaire bilingue spécialisé, telle que nous l'avons définie dans notre méthodologie, comprend, outre les termes spécialisés, également ceux de la langue courante qui appartiennent au domaine, comme **environnement**, **pollution**. Ces termes appellent souvent plusieurs traductions proches certes (cependant l'omission de l'une d'elles équivaldrait tout de même à une lacune), mais pas suffisamment différentes pour que soit introduite une subdivision sémantique. Au lexicographe d'estimer s'il a affaire à une véritable polysémie (ce que Vinay et Darbelnet nomment bi- ou tribifurcation) ou à de simples glissements de sens ne justifiant pas de telles sous-catégories.

- 21 On passera sur le cas où la polysémie est clairement établie, pour réfléchir à la signification, aux implications et à la portée du cas où le lexicographe propose plusieurs traductions pour un même terme.
- 22 Soit le mot **GROUND**, pour lequel *Robert & Collins* propose quatre subdivisions sémantiques⁵ :
 - (a) (U) **terre** f, **sol** m
 - (b) (U) **sol** m, **terre** f, **terrain** m
 - (c) (U) **terrain** m, **domaine** m, **terres** fpl, **territoires** mpl, **sol** m
 - (d) **terrain** m
- 23 On remarque d'emblée les répétitions d'une subdivision sémantique à l'autre : quelle est leur raison d'être et comment réagit l'utilisateur face à ces diverses propositions de traduction ?
- 24 Si on devait donner le mode d'emploi d'un dictionnaire bilingue à quelqu'un qui n'en a jamais vu, il est quasiment certain qu'on expliquerait que **GROUND** peut se traduire par **terre** ou **sol**, avec des exemples à l'appui, ou bien par **sol** ou **terre** ou **terrain** dans tel ou tel contexte, etc. C'est probablement ainsi que le lexicographe l'a conçu, et les explications données seraient en partie fondées. Mais comment justifier ces redites ? En effet, toutes traductions confondues, on ne dénombre que 5 termes différents. Pour quelles raisons le lexicographe a-t-il choisi cette configuration ?
- 25 Étant donné que ces traductions ne sont pas interchangeables, et que par ailleurs chacune d'elles est indispensable, cela signifie qu'elles doivent entretenir une relation d'interdépendance et qu'il faut donc peut-être les considérer telles qu'elles sont juxtaposées. En réalité, le véritable message n'est pas « **GROUND** = **sol** » ou « **GROUND** = **terre** » ou « **GROUND** = **terrain** », autrement dit « **A** = **a**, ou **A** = **b** ou encore **A** = **c** ».
- 26 Lorsqu'il a traité le mot **GROUND**, le lexicographe a tout d'abord estimé nécessaire d'établir des subdivisions sémantiques (polysémie) puis, au sein de chacune d'elles de proposer deux, trois, voire quatre traductions afin de couvrir le « sous-champ sémantique » qu'il avait délimité en premier lieu (parasynonymie). En d'autres termes, il a ajouté une traduction chaque fois que le « sous-champ » ne lui semblait pas globalement couvert.
- 27 Comme on sait par ailleurs que **sol**, **terre** ou **terrain** ne se traduisent pas exclusivement par **GROUND**, il serait logique de conclure que ce n'est pas à la totalité du champ sémantique de **sol**, **terre** ou **terrain** que se réfère le lexicographe. Il ne peut cependant matérialiser à quelle partie du champ de **sol** (ou **terre** ou **terrain**) il estime que **GROUND** correspond. Mais puisqu'on sait que ces termes sont parasynonymiques — ce qui implique

que leurs champs sémantiques se recouvrent partiellement — on peut en conclure que le champ de **GROUND** est constitué des parties de champs sémantiques communes aux trois termes présentés.

- 28 C'est ce fonds commun qui constituerait l'équivalent réel, mais « non-nommable » de **GROUND**. Le sens du message du lexicographe serait donc : « A correspond à ce que a, b et c ont en commun ». C'est par l'association de plusieurs termes de la langue cible que le lexicographe a reconstruit le sème du terme-source.
- 29 Le traducteur devra, lui aussi, procéder à sa propre opération de reconstitution à partir des éléments proposés par le lexicographe, afin d'en déterminer le « fonds commun », puis de sélectionner, pour le contexte donné, le terme-cible idoine : ce ne sera pas nécessairement un de ceux proposés, car le lexicographe, soucieux de ne pas noyer son lecteur, ne sera pas tombé dans le piège de l'inventaire — inutilisable et contenant par nature de nombreuses traductions marginales.
- 30 On ne peut manquer d'établir un parallèle entre cette question du « fonds commun » et celle de la définition du dictionnaire unilingue : les équivalences du dictionnaire bilingue constitueraient en fait une sorte de définition du terme-source dans la langue-cible, à ceci près que dans le dictionnaire unilingue, il s'agit de délimiter un champ sémantique et un seul, alors que dans l'autre, on doit, à partir de *n* champs sémantiques, tenter d'en définir un autre. Et tout comme la définition ne peut véritablement donner les limites de l'extension du sens d'un terme, on conçoit que les traductions ne pourront, au mieux, que constituer une « définition » très approximative.
- 31 À ce propos, afin de procurer à l'utilisateur le plus d'informations possible, on pourrait, avec profit, insérer dans un dictionnaire bilingue (spécialisé ou non) la définition du terme-source.
- 32 Après avoir analysé ce qui, dans les équivalences sémantiques, est générateur d'incohérences lors du retournement, examinons ce qui dans la nature même des exemples, peut induire des problèmes. Car le lexicographe n'aura pas manqué d'insérer des « illustrations » afin de donner au lecteur une représentation plus riche du champ sémantique.
- 33 L'actualisation du mot-vedette dans des exemples compense certes la sécheresse des équivalences. Cependant, l'exemple porte en lui-même ses limites. Il n'illustre qu'une facette de la notion en un point choisi du champ sémantique, tout comme un sondage donne une image statique de l'opinion à un moment donné. Proposer un exemple, c'est aussi risquer de faire oublier l'étendue et le continuum du champ sémantique.
- 34 L'exemple est également révélateur du caractère réducteur des équivalences proposées : en effet, dans un cas sur trois environ, on traduira, dans la phrase, le mot-vedette à illustrer par un terme qu'on n'aura pas proposé parmi les traductions possibles. Ce phénomène est d'autant plus fréquent qu'on voudra offrir un énoncé véritablement équivalent, non calqué, qui ne « sent » pas la traduction⁶.
- 35 Ces décalages peuvent évidemment être éludés si on ne choisit que des exemples qui se prêtent à une traduction terme à terme, mais on n'aiderait guère l'utilisateur et ce qu'on pourrait gagner en cohérence interne apparente, on le perdrait en authenticité et en richesse.
- 36 Pour résumer, les exemples sont par nature porteurs d'informations utiles, mais nécessairement parcellaires.

- 37 Ces réflexions sur la démarche du lexicographe lors de l'élaboration de la première version nous apportent quelques réponses aux problèmes évoqués au début de cette étude sur le caractère imparfait du produit du retournement.
- 38 Nous avons montré dans la première partie que tant qu'il s'agissait d'anomalies liées à la morphologie des termes, on pouvait anticiper les transformations et déplacements à effectuer.
- 39 Mais tout ce qui vient d'être exposé sur la démarche du lexicographe laisse deviner que le retournement brut de la première version ne va procurer que le canevas de la deuxième. Car dans la mesure où le lexicographe a été contraint, pour chaque terme, d'élaborer son champ sémantique dans l'autre langue, on ne voit pas comment il pourrait échapper au même travail d'élaboration pour la deuxième version.
- 40 Il avait écrit : **A = a, b, c**
- 41 Le retournement informatique brut lui donnera :
- a = A
b = A
c = A
- 42 On sait d'ores et déjà que ces équivalences sont fausses puisque seule une partie du champ sémantique de **a, b** et **c** correspond à **A**. Le lexicographe sera probablement amené à écrire :
- a = A, B, C**
b = A, C, D
c = A, B, E
- 43 Lors de l'élaboration de la deuxième version, le lexicographe va également intervenir dans la nomenclature. Car il serait erroné de penser que tous les termes traduits de la première version vont nécessairement figurer dans la seconde.
- 44 Ainsi pour **INFINITE**, il propose **inépuisable** et **infini**. Dans le cadre d'un dictionnaire de l'environnement, il est tout à fait justifié d'inclure cet adjectif dans la nomenclature⁷. On ne peut légitimement pas se contenter de donner **inépuisable** comme seule équivalence ; on ajoute donc **infini**, mais lors du retournement, il paraît superflu de le maintenir. Ceci est une fois de plus dû à la recherche du fonds commun que le lexicographe a dû cerner afin de proposer un champ d'équivalence sémantique relativement bien couvert.
- 45 Ainsi, la seconde version n'est pas la première retournée. On pourrait dire qu'elle en est le reflet global mais que ses composantes ne sont pas strictement identiques et qu'elles sont organisées différemment.
- 46 La question ultime est bien sûr : dans quelle mesure un retournement informatique donne-t-il des résultats exploitables ? Vaut-il la peine d'être effectué ?
- 47 Pour apporter une réponse fondée à cette question, il faudrait faire l'expérience d'un retournement « manuel », mais on peut estimer qu'il donne, malgré toutes les imperfections développées au cours de cet exposé, une ébauche de document de travail tout à fait utilisable. Cependant, il ne faut pas s'attendre à ce que la plus grande partie du travail soit réalisée par cette simple opération.
- 48 La question se pose-t-elle différemment selon qu'il s'agit d'un dictionnaire bilingue général ou spécialisé ? Dans la première partie de notre méthodologie, nous avons conclu, avec arguments à l'appui, qu'un dictionnaire bilingue spécialisé méritait d'être aussi bien conçu qu'un dictionnaire bilingue général, que les contraintes étaient les

mêmes si l'on voulait créer un dictionnaire véritablement utile et complet. Ce qui distingue fondamentalement les deux types, c'est l'ampleur de la nomenclature. Le fait qu'il s'agisse d'un dictionnaire spécialisé est un aspect annexe.

- 49 En conclusion, qu'a apporté ce retournement informatique ? En quoi a-t-il été une épreuve de vérité ?
- 50 Il a tout d'abord confirmé ce que nous savions par ailleurs de par notre expérience de linguiste, d'enseignant et de traducteur à propos de la non-correspondance des champs sémantiques. Mais le caractère brut du produit du retournement automatique des données a obligé à une recherche sur les causes profondes des anomalies engendrées. Et cette réflexion sur la démarche du lexicographe, sur les conséquences de ses choix stratégiques nous a conduit à élaborer l'hypothèse du « fonds commun », en fin de compte à en apprendre sur la conception même de la première partie.
- 51 En conclusion, ce retournement informatique a permis d'élucider les raisons pour lesquelles l'élaboration de la seconde version d'un dictionnaire bilingue constitue, exactement au même titre que la première, un authentique travail de création lexicographique.

BIBLIOGRAPHIE

Collin, Peter & Schuwer, Martine. 1992. *Dictionnaire anglais/français - français/anglais de l'environnement et de l'écologie*. Paris : P. Collin Publishing-Larousse.

NOTES

1. En collaboration avec Annette Reumert Kazès, In *Actes du XI^e colloque du GERAS* (1990).
2. PACD (Programme d'aide à la création de dictionnaires) élaboré par Martin Rabenjamina - OI Concept
3. Exemples tirés de l'ouvrage précité.
4. Telles qu'on les trouve dans *Longman Dictionary of Contemporary English* ou dans *Collins Cobuild* et qui se distinguent donc des expressions données à titre d'exemple.
5. In *Robert & Collins* (1987). On a sélectionné les quatre premières subdivisions sémantiques, laissant de côté les sens figurés ou marginaux. Par ailleurs, pour des raisons pratiques, on ne citera pas l'article dans son intégralité, on omettra les lexies composées et exemples qui constituent l'originalité et la « chair » de cet ouvrage, qui permettent justement à l'utilisateur de saisir les nuances et ainsi de choisir parmi les équivalences proposées.
6. La moindre des modifications sera une transposition (changement de catégorie grammaticale d'un terme) : *foggy days are frequent in November* = il y a souvent du **brouillard** en novembre, mais on observe souvent de plus grands changements (utilisation d'un mot d'une autre famille) : *the tree grows a new ring each year* = chaque année, l'arbre a un nouvel anneau de **croissance**.
7. INFINITE RESOURCES = ressources inépuisables

RÉSUMÉS

Cet exposé constitue la seconde partie de la méthodologie de création de dictionnaires bilingues spécialisés consacrée à la réalisation de la version anglais-français, et élaborée à l'aide d'un logiciel spécifique de gestion de données. Lors de la mise en œuvre de la partie français-anglais, réalisée informatiquement, sont apparus de multiples problèmes inattendus : certains, d'ordre technique, peuvent bénéficier d'améliorations en amont, mais pour les autres, il n'existe pas de solution d'ensemble car ils sont en réalité le révélateur sévère de la non-correspondance des champs sémantiques d'une langue à l'autre. Le retournement informatique constitue donc une « épreuve de vérité » en ce sens qu'il met en lumière les « vices cachés » (lacunes/approximations) de la version de départ ; il souligne par ailleurs les efforts délibérés ou inconscients que le lexicographe déploie lors de la mise en équivalence afin de procurer à l'utilisateur un outil aussi fiable et complet que possible.

The first part of this methodology was devoted to the development of bilingual dictionaries of scientific and technical terms; it was carried out with the help of purpose-designed information management software and applied to the development of the English-French version. The computerized implementation of the French-English version lead to numerous and unexpected problems: those linked with morphology have easily been solved, but no overall solution could be adopted for the remainder for lack of correspondence between semantic fields from one language to another. The file reversal finally amounts to an acid test insofar as it throws light on hidden defects (omissions/approximations) of the first version. It also stresses deliberate or unconscious efforts by the lexicographer to provide the user with equivalents as reliable and complete as possible.

AUTEUR

MARTINE SCHUWER

Université Rennes 2 Haute-Bretagne. Martine.Schuwer@uhb.fr