

# Acquisition automatique de traductions de termes complexes par comparaison de « mondes lexicaux » sur le Web

Stéphanie Léon

Université de Provence – Équipe DELIC  
leon.stephanie@gmail.com

## Résumé

Nous présentons une méthode de traduction automatique de termes complexes pour la construction de ressources bilingues français/anglais, basée principalement sur une comparaison entre « mondes lexicaux » (ensemble de co-occurents), à partir du Web. Nous construisons les mondes lexicaux des termes français sur le Web. Puis, nous générons leurs traductions candidates via un dictionnaire bilingue électronique et constituons les mondes lexicaux de toutes les traductions candidates. Nous comparons enfin les mondes lexicaux français et anglais afin de valider la traduction adéquate par filtres statistiques. Notre évaluation sur 10 mots français très polysémiques montre que l'exploitation des mondes lexicaux des termes complexes sur le Web permet une acquisition automatique de traductions avec une excellente précision.

**Mots-clés** : termes complexes, traduction automatique, mondes lexicaux, World Wide Web.

## Abstract

We present a method for automatic translation (French/English) of complex terms using a comparison between French and English “lexical worlds” (*i.e.* the set of cooccurring words) on the Web. We first generate the French lexical worlds. We then generate their potential translations by means of an electronic dictionary and extract the lexical worlds corresponding for English. Finally we compare the French and English lexical worlds in order to select the adequate translation. Our evaluation on 10 highly polysemous French words shows that a very high precision can be achieved using this technique.

**Keywords**: complex terms, machine translation, lexical worlds, World Wide Web.

## 1. Introduction

Les performances des systèmes en recherche d'informations interlingue, tout comme celles des systèmes en traduction automatique, sont fortement freinées par le problème de l'ambiguïté lexicale des mots polysémiques ou homographes. Ainsi, la traduction anglaise du terme français *caisse* sera différente selon que l'usage concerne, entre autres, l'INSTRUMENT DE MUSIQUE (*drum*), la BANQUE (*fund*) ou encore la VALISE (*case*). Le manque de désambiguïtation lexicale pour la traduction automatique conduit à des résultats qui gênent souvent considérablement la compréhension. Par exemple, le système de traduction automatique *Systran*<sup>1</sup> traduit le terme complexe<sup>2</sup> *caisse centrale* par *central case*, ce qui est non seulement incompréhensible pour un anglophone, mais qui génère des requêtes erronées dans le cadre de recherches interlingues, lors desquelles la requête est reformulée en

---

<sup>1</sup> <http://www.systran.fr/index.html>

<sup>2</sup> Nous parlons de *terme complexe* afin de désigner une unité lexicale complexe caractérisée par des critères morpho-syntaxiques et sémantiques (Daille, 1994).

diverses langues<sup>3</sup>. Pourtant, la co-occurrence *caisse/centrale* constitue un indice désambiguïseur très fort, qui, si elle était correctement enregistrée et désambiguïcée dans une base de données bilingue, voire multilingue, pourrait servir à générer des traductions correctes (*caisse centrale* > *central fund*). La polysémie est rendue très faible dès que l'on envisage les mots-clés selon leur co-occurent (Yarowsky, 1993 ; Shütze, 1998 ; Véronis, 2003).

Le fort multilinguisme du Web génère des besoins considérables en traduction qui nécessitent d'acquérir des stratégies de traduction automatique de termes complexes, utiles pour la recherche interlingue ou pour informer l'utilisateur sur la traduction de termes complexes sur le Web ou dans des bases textuelles<sup>4</sup>. De nombreux travaux ont proposé des méthodes d'acquisition de terminologie bilingue à partir de corpus parallèles (pour un état de l'art, voir (Véronis, 2000)) ou comparables (citons, entre autres, (Rapp, 1999 ; Fung et McKeown, 1997 ; Fung et Yee, 1998 ; Morin *et al.*, 2004)). Toutefois, de tels corpus constituent des ressources rares et concernent des domaines restreints. Les dictionnaires bilingues existant se contentent, quant à eux, de rares indications ponctuelles sur la traduction des termes complexes.

Notre étude propose une acquisition automatique de traductions français-anglais de termes complexes, pour la construction de ressources bilingues, à partir du Web, via l'exploitation de mondes lexicaux. Nous parlerons de « monde lexical » afin de désigner les co-occurrences fréquentes d'un mot ou d'un terme complexe, à la suite des travaux de Véronis (2003). De tels voisinages, plus larges que le co-occurent immédiat, peuvent se situer au niveau du paragraphe, ou même de la phrase. Par exemple, le monde lexical de la requête « *caisse centrale* » sur Yahoo (*agricole, social, mutualité, crédit, banque, assurance, gestion, etc.*) est proche de celui de sa traduction correcte « *central fund* » (*money, pay, budget, insurance, management, social, etc.*), contrairement à la traduction erronée « *central case* » (*study, law, policy, enterprise, university, etc.*). Nous avons présenté dans Léon et Millon (2005) une méthode de traduction semi-automatique de relations lexicales, basée sur la fréquence sur le Web des traductions candidates, dans la lignée de travaux tels que Grefenstette (1999) et Cao et Li (2002). Toutefois, la fréquence ne permet pas de désambiguïser systématiquement les cas de polysémie, tandis qu'une comparaison des mondes lexicaux permet de lever un grand nombre d'ambiguïtés lexicales. Certains travaux ont montré que l'exploitation des mondes lexicaux permet une désambiguïsement lexical d'un point de vue monolingue (Pichon et Sébillot, 1999 ; Véronis, 2003). En traduction, des recherches ont souligné que les co-occurrences immédiates sont les mêmes d'une langue à l'autre (entre autres (Rapp, 1999)), mais aussi un entourage linguistique plus large<sup>5</sup> (Fung et Yee, 1998 ; Kikui, 1998 ; Tanguy, 1999 ; etc.). Toutefois, les stratégies ont été appliquées essentiellement sur des corpus terminologiques, ce qui limite la quantité des usages, et uniquement à partir de corpus parallèles ou comparables.

Nos travaux présentent des similitudes avec ceux de Lafourcade *et al.* (2004) qui proposent la création de ressources monolingues et bilingues par la construction de vecteurs conceptuels : les concepts sont donnés a priori via des thesaurus et sont reliés à des items lexicaux. Pour

---

<sup>3</sup> La recherche interlingue est à distinguer de la recherche multilingue. Dans ce dernier cas, des traitements d'indexation spécifiques à chaque langue sont appliqués.

<sup>4</sup> Parmi les ressources existantes, le projet Papillon (<http://www.papillon-dictionary.org/>) vise à créer un environnement multilingue de recherche dictionnaire en ligne. Il s'appuie sur des ressources existantes, tandis que nous visons à la création d'une base à partir d'usages sur le Web.

<sup>5</sup> Certains parlent de *vecteurs de co-occurrences* (Kikui, 1998). Une notion proche est celle d'*isotopie* (Greimas, 1986) afin de désigner « la récurrence d'une unité sémantique le long d'un texte » (Tanguy, 1999).

nous, le monde lexical est construit uniquement à partir de données textuelles. À notre connaissance, aucune expérience sur la comparaison des mondes lexicaux n'a été menée en langue générale, sur l'immense base de données que constitue le Web. Nous présentons un protocole de traitement à titre expérimental, susceptible d'être appliqué sur un plus grand nombre de termes complexes, et transposable sur d'autres langues que le français et l'anglais. La section 2 consiste en une description des étapes de traitements. La section 3 constitue un exemple chiffré des processus décrits à la section 2. Nous analysons les difficultés et abordons les perspectives d'évolution dans la section 4.

## 2. Méthodologie et traitement des données

Nous procédons à une acquisition de termes complexes français par extraction des « termes associés » retournés par *Exalead*<sup>6</sup>. Nous constituons les mondes lexicaux sur le Web de chacun de ces termes à l'aide de requêtes sur *Yahoo*. Puis, nous générons leurs traductions candidates via un dictionnaire bilingue électronique et constituons les mondes lexicaux de toutes les traductions candidates. Nous comparons enfin les mondes lexicaux français et anglais, par filtres statistiques, afin de valider la traduction adéquate.

### 2.1. Acquisition de termes complexes en langue source

Notre première étape consiste à acquérir des termes complexes en langue source, dont la tête sémantique présente un degré de polysémie élevé. Nous nous fondons sur le nombre de traductions possibles en langue cible pour chaque entrée française (mono-terme) au sein du dictionnaire électronique bilingue *Collins Pocket français-anglais* : « Les différents sens d'un mot sont déterminés par les différentes traductions de ce mot dans une autre langue. » (Dagan, 1991) Nous avons obtenu en tête de liste les dix noms suivants : *appareil, caisse, tour, éclat, disposition, parc, rapport, reprise, tenue* et *provision*. Ils atteignent un seuil de difficulté de traduction élevé, de par la prolifération de leurs traductions candidates dont le nombre moyen est de 8. Nous utilisons le moteur de recherche *Exalead* afin de récupérer les termes complexes en langue source contenant ces dix noms. *Exalead* propose une organisation monolingue des résultats selon les « termes associés » à la requête, comme dans l'exemple de *caisse* :

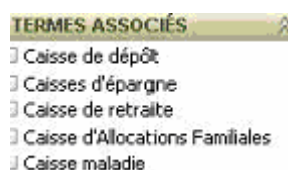


Figure 1. Extrait de termes associés à la requête *caisse* sur *Exalead*

Nous produisons une requête pour chaque terme complexe, au singulier et au pluriel, pour la langue française et collectons manuellement les termes qui leur sont « associés ». Nous ne conservons que ceux qui contiennent exactement deux mots pleins<sup>7</sup>, c'est-à-dire que *tour de scrutin* est conservé, tandis que *tour de France à la voile* est éliminé. Les termes qui ne contiennent pas le nom source (mais qui lui sont associés thématiquement) sont éliminés, comme *soin du visage* pour la requête *éclat*, ainsi que les termes « bruités » (comme par exemple, *disposition du présent*). Nous obtenons 132 termes complexes du type de *parc du*

<sup>6</sup> <http://www.exalead.fr/search>

<sup>7</sup> Prendre en compte plus de deux mots demanderait une analyse comparative plus approfondie.

*château, appareil électrique, caisse d'épargne*, etc. (le nombre moyen de termes associés par nom source est de 13).

## 2.2. Construction automatique de mondes lexicaux français

Le moteur de recherche *Yahoo* est interrogé automatiquement par script via l'interface de programmation d'applications API<sup>8</sup> (*Application Programming Interface*) afin de récupérer les 1000 premiers titres et résumés<sup>9</sup> renvoyés pour chaque requête des termes complexes. Ces dernières sont encadrées par des guillemets afin d'être considérées comme une expression exacte, et sont élargies à leur forme au singulier et au pluriel, pour la langue française :

« *caisse centrale* » -« *caisses centrales* »  
 « *caisses centrales* » -« *caisse centrale* »  
 « *caisse centrale* » +« *caisses centrales* »

Les résumés sont nettoyés automatiquement par script, afin de rétablir certains problèmes de codage des caractères accentués ou de supprimer des adresses Internet, etc. Pour chaque terme, nous sélectionnons de façon automatique les cinquante mots les plus fréquents parmi les résumés. Ces mots constituent leur monde lexical. Le choix de longueur du monde lexical s'est opéré par observation, pour déterminer un seuil représentatif. Pour calculer la fréquence, nous ne prenons en compte qu'une occurrence par résumé, afin de limiter des cas de répétitions multiples liées en particulier aux spams. Un « anti-dictionnaire » est utilisé pour supprimer les mots non pertinents tels que des mots liés au Web (comme par exemple *lien*, *blog*, etc.), des verbes supports, des déterminants, etc. Nous présentons les vingt premiers mots du monde lexical de *caisse centrale* :

*agricole, sociale, mutualité, crédit, banque, banques, sociales, coopératif, nationale, union, gestion, coopération, populaires, réassurance, garantie, assurance, agricoles, société, mutuel, maladie*

## 2.3. Traductions candidates et mondes lexicaux anglais

Nous décrivons les étapes de génération des traductions candidates, puis de production des mondes lexicaux pour la langue cible.

### 2.3.1. Génération automatique des traductions candidates

Nous générons automatiquement toutes les traductions candidates des termes complexes français, via le *Collins Pocket*, sur le même modèle que Léon et Millon (2005). Par exemple, à partir des traductions de *caisse* et de *centrale* du *Collins Pocket*, nous générons toute la combinatoire :

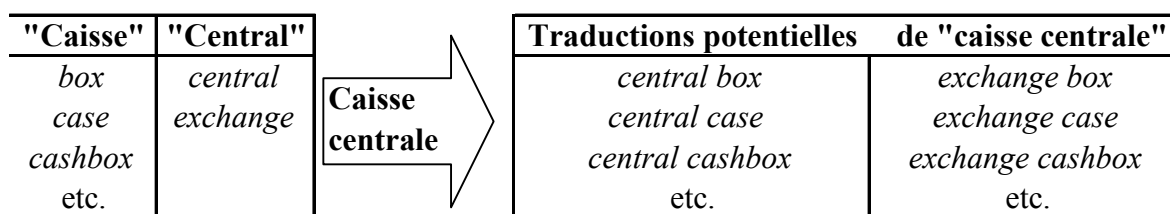


Figure 2. Exemple de traductions candidates de *caisse centrale*

<sup>8</sup> <http://developer.yahoo.net/>

<sup>9</sup> Le nombre de requêtes générées étant très conséquent, il serait fastidieux d'extraire les pages Web.

### 2.3.2. Construction automatique des mondes lexicaux anglais

À partir des traductions candidates, nous interrogeons le Web (via l'API *Yahoo*), pour la langue anglaise. Les traductions sont considérées comme des expressions exactes (encadrés par des guillemets). Les requêtes booléennes permettent de traiter divers types de structures syntaxiques<sup>10</sup> (Chuquet et Paillard, 1987 ; Daille, 1994), ainsi que des variations d'article<sup>11</sup> :

- NOM ADJECTIF > "*the ADJ NOM*" OR "*a ADJ NOM*"
- NOM1 de/d' NOM2 > "*NOM1 of NOM2*" OR "*NOM2 NOM1*"
- NOM1 du/de la/des/NOM2 > "*NOM1 of the NOM2*" OR "*NOM2 NOM1*" OR "*NOM1 of NOM2*"
- NOM1 NOM2 > "*NOM2 NOM1*"

Le monde lexical des traductions candidates est créé de la même façon que pour les résumés français, dont voici des extraits pour des traductions relatives au nom *appareil* :

*electronic device* : *power, technology, control, mitsubishi, computer, digital, parent, battery, electric, apple*

*domestic appliance* : *spares, vacuum, spare, repair, machines, washing, religious, electric, repairs, juicer*

*mobile phone* : *cell, technology, celular, pc, computer, device, nokia, motorola, driving, digital*

*respiratory apparatus* : *anatomy, human, illness, tuscan, air, lungs, splanchnology, arthritises, function, skin*

*judicial apparatus* : *law, government, code, administration, reform, civil, political, guidelines, police, power*

## 2.4. Désambiguïstation lexicale et validation semi-automatique des noms cibles

### 2.4.1. Traduction des mondes lexicaux français

Afin de comparer les mondes lexicaux en langue source et en langue cible, les listes des mondes lexicaux français sont traduites automatiquement, via le dictionnaire *Collins Pocket*. Les listes (françaises et anglaises) sont préalablement « nettoyées » pour que les unités lexicales correspondent aux entrées du dictionnaire : les formes les plus communes sont lemmatisées, par script, comme par exemple la suppression de la marque « s » du pluriel. Un filtre morpho-syntaxique est appliqué : nous ne conservons que les entrées correspondant à un nom, un adjectif ou un verbe. Les mots inconnus sont traduits tels quels, comme les noms propres, les noms de marque, etc. Nous obtenons les listes filtrées des mondes lexicaux français assortis de la (ou les) traduction(s) de chaque mot, comme dans l'exemple :

<sup>10</sup> Il existe des régularités de traduction des structures syntaxiques des termes complexes (Daille, 1994).

<sup>11</sup> Nous ne traitons pas la structure « ADJ-NOM » sans déterminant, ce qui permet de réduire des problèmes d'ambiguïté catégorielle, bien que des cas de silence soient observés.

Fréquence	Mot	Traduction(s)	Fréquence	Mot	Traduction(s)
314	<i>agricole</i>	<i>agricultural</i>	131	<i>crédit</i>	<i>credit/funds</i>
276	<i>social</i>	<i>social</i>	73	<i>banque</i>	<i>bank/banking</i>
266	<i>mutualité</i>	<i>mutual</i>	63	<i>coopératif</i>	<i>cooperative</i>

Tableau 1 : Extrait de traductions du monde lexical français de caisse centrale

#### 2.4.2. Validation semi-automatique des traductions candidates

Un filtre est utilisé : seules les traductions candidates ayant une fréquence sur le Web supérieure ou égale à 500 sont conservées, ce qui nous permet d'éliminer d'emblée les traductions peu usitées sur le Web. Les listes des traductions candidates conservées sont comparées automatiquement avec la liste du terme complexe source : chaque mot de la liste anglaise est recherché parmi les traductions de la liste française et nous comptabilisons le nombre de mots en commun. Pour la comparaison, nous utilisons le coefficient de Jacquard, qui mesure le degré de similitude entre deux ensembles :  $|\text{inter}(X,Y)| / |\text{union}(X,Y)|$ . Ces scores sont multipliés par mille afin d'être rendus plus lisibles. Un filtre fixé de façon expérimentale est appliqué aux coefficients de Jacquard : nous ne conservons que les traductions ayant un score supérieur ou égal à 40 et nous ne ramenons que la traduction candidate atteignant le plus haut score, sauf lorsque la différence de score est inférieure à un. Enfin, les couples de traduction sont testés par le biais de nouvelles requêtes. Le Web est riche d'un grand nombre de documents bilingues. Si la traduction est correcte, le couple doit apparaître au moins une fois dans un même document : de nombreux documents du Web sont des textes parallèles<sup>12</sup> (manuels, catalogues, etc.). Il peut s'agir aussi de traductions ponctuelles dans le texte d'un document monolingue :

*Further support was guaranteed [...], the Saudi Fund, France's Central Fund for Economic Cooperation (Caisse Centrale de Coopération Economique--CCCE).*

Nous engendrons automatiquement une requête pour chaque couple restant, du type de "caisse centrale" "central fund" et ne conservons que les couples dont la fréquence est supérieure ou égale à 1.

### 3. Premiers résultats

Le tableau 2 montre la proportion de traductions après chaque étape.

Termes complexes français	Traductions candidates générées	Filtre automatique			Validation manuelle
		Filtre seuil fréquence	Filtre indice de similarité	Filtre "couples"	Traductions valides
132	5585	1357	87	57	57

Tableau 2. Résultats de la validation des traductions

<sup>12</sup> Il ne s'agit pas d'un « corpus » au sens propre : les pages Web peuvent être courtes et ne sont pas nécessairement alignées.

Notre but étant la production automatique de ressources bilingues, nous privilégions volontairement la précision au détriment du rappel, de façon à obtenir une ressource de bonne qualité avec un effort manuel minimal. Notre expérience a retourné 57 traductions correctes sur les 132 termes complexes proposés (rappel). La précision, elle, est excellente, puisque aucune des traductions n'était erronée, ce qui permet d'ajouter les expressions à notre dictionnaire sans autre intervention. Nous présentons un exemple de traductions :

*appareil électrique* > *electrical appliance/device*  
*appareil électroménager* > *domestic appliance*  
*appareil électronique* > *electronic device*  
*appareil ménager* > *domestic appliance*  
*appareil de mesure* > *measurement apparatus*  
*appareil mobile* > *mobile device*  
*appareil numérique* > *digital camera*  
*appareil respiratoire* > *respiratory apparatus*

Les mondes lexicaux obtenus sont en eux-mêmes intéressants, et peuvent probablement être exploités comme ressource bilingue. Voici deux extraits de mondes lexicaux pour *appareil numérique* et sa traduction *digital camera* :

*appareil numérique* : *canon, photographie, nikon, informatique, produit s, accessoires, digital, mémoire, kodak, pc, olympus, flash, zoom, cartes, argentique, prises, gamme, reflex, prise, matériel*

*digital camera* : *photography, film, computer, kodak, technology, olympus, canon, right, zoom, sony, memory, resolution, lens, imaging, fuji, ratings, pc, brands, product, battery*

## 4. Analyse des difficultés et perspectives

### 4.1. Amélioration de la précision

La précision étant excellente, nous privilégions des réflexions afin d'augmenter le rappel. Il n'est pas impossible que sur un plus grand nombre de termes, une faible proportion d'erreurs apparaisse. Si l'on devait améliorer la précision malgré tout, une perspective serait d'imposer une fréquence plus élevée aux couples français/anglais, dans la phase de filtrage finale.

### 4.2. La « non-compositionnalité »

Une difficulté est l'absence de la traduction de l'un des éléments du terme complexe dans le dictionnaire. *Grosse caisse* devrait être traduit par *kick drum* ou *bass drum*. Or, *drum* (« tambour ») ne se trouve pas dans la liste des traductions candidates de *caisse*. Le principe de compositionnalité est le fait que « le sens d'un tout est une fonction du sens de ses parties et de la façon dont elles se combinent » (Bouillon, 1998). Pour *grosse caisse*, la compréhension du sens global ne peut être déduit de chaque constituant (Morin *et al.*, 2004). Une perspective serait de prospecter le nom cible parmi le monde lexical du terme complexe français : *tambour* fait partie du monde lexical de *grosse caisse*. Si l'on engendrait des requêtes avec les traductions du monde lexical français (ici *drum*), il serait parfois possible de sélectionner les traductions adéquates. Toutefois, le nom-cible ne se situe pas toujours en tête de liste, ce qui demande de produire un nombre conséquent de requêtes.

### 4.3. Noms « généraux »

Une difficulté concerne les noms cibles très généraux, n'engendrant pas un monde lexical précis, comme l'exemple de *case* (*caisse*). La traduction candidate *retirement case* (*cas de retraite*) provoque un monde lexical en thématique avec la retraite (*benefit, age, pension, financial, etc.*), ce qui bruite nos comparaisons. Une solution serait d'éliminer de tels mots.

### 4.4. Amélioration du filtre des fréquences sur le Web

Une autre amélioration serait le calcul du rapport entre la fréquence sur le Web du terme complexe français et celui des traductions candidates. Par exemple, « *caisse de retraite* » apparaît 157 000 fois. « *Retirement case* » apparaît 2 850 fois, tandis que « *retirement fund* » apparaît 1 240 000 fois. On pourrait exclure les traductions ayant une fréquence inférieure au terme français (le rapport entre le français et l'anglais est d'environ 1/20 sur *Yahoo*).

### 4.5. Comparaison et extraction des mondes lexicaux

Certains mots au sein des listes françaises et anglaises sont différents morphologiquement, mais proches de façon sémantique, comme l'exemple de *pédagogique* et *éducation*. Si l'on pouvait inclure des liens lexicaux du type *éducation* > *pédagogique*, nos comparaisons pourraient être plus fines. Une autre perspective concerne l'amélioration de l'extraction des mondes lexicaux. Nous pourrions extraire des unités lexicales complexes au sein des paragraphes. Par exemple, la traduction d'*aspirateur* est *vacuum cleaner*. *Vacuum* et *cleaner* sont recensés séparément, mais il faudra les regrouper par une analyse syntaxique des résumés anglais.

## 5. Conclusion

Nous avons présenté une méthode de traduction automatique de termes complexes basée sur la comparaison de mondes lexicaux sur le Web entre le terme complexe source et toutes ses traductions candidates. D'autres critères sont pris en compte tels que la fréquence sur le Web des traductions candidates, ainsi que la présence des couples français/anglais au sein d'un même document. Nos tests sur un échantillon de noms français fortement polysémiques, atteignant une précision de 100 %, montrent qu'une telle méthode permet d'automatiser la construction de ressources bilingues avec une fiabilité importante, en vue d'applications telles que la recherche d'informations interlingue ou la traduction automatique. Ces mondes lexicaux peuvent à la fois constituer des ressources en tant qu'aide à la traduction, mais aussi être exploités pour une organisation de la connaissance bilingue de type ontologique. Nos perspectives d'évolution concernent l'amélioration de l'extraction des mondes lexicaux, grâce à une analyse des dépendances syntaxiques, en vue d'augmenter le rappel.

## Références

- BOUILLON P. (1998). *Traitement automatique des langues naturelles*. Duculot, Paris-Bruxelles.
- CAO Y., LI H. (2002). « Base noun phrase translation using Web data and the EM algorithm ». In *Proceedings of International Conference on Computational Linguistics (COLING)* : 127-133.
- CHUQUET H., PAILLARD M. (1987). *Approche linguistique des problèmes de traduction : anglais-français*. Ophrys, Gap.
- DAGAN I., ITAI A., SCHWALL U. (1991). « Two languages are more informative than one ». In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL)* : 130-137.



- DAILLE B. (1994). Approche mixte pour l'extraction automatique de terminologie : statistiques lexicales et filtres linguistiques. Thèse de Doctorat en Informatique Fondamentale, Université Paris 7.
- FUNG P., MCKEOWN K. (1997). « Finding terminology translations from non-parallel corpora ». In *Proceedings of Annual Workshop on Very Large Corpora* : 192-202.
- FUNG P., YEE L. Y. (1998). « An IR approach for translating new words from nonparallel, comparable texts ». In *Proceedings of International Conference on Computational Linguistics (COLING)* : 414-420.
- GREFENSTETTE G. (1999). « The WWW as a resource for example-based MT tasks ». In *Proceedings of Translating and the Computer Conference*.
- GREIMAS A. (1986). Sémantique structurale : recherche de méthode. PUF, Paris .
- KIKUI G. (1998). « Term-list translation using monolingual co-occurrence vectors ». In *Proceedings of International Conference on Computational Linguistics (COLING)* : 670-674.
- LAFOURCADE M., RODRIGO F., SCHWAB D. (2004). « Low Cost Automated Conceptual Vector Generation from Mono and Bilingual Resources ». In *Actes de PAPILLON-2004*.
- LEON S., MILLON C. (2005). « Acquisition semi-automatique de relations lexicales bilingues (français-anglais) à partir du Web ». In *Actes de RECITAL 2005* : 595-604.
- MORIN E., DUFOUR-KOWALSKI S., DAILLE B. (2004). « Extraction de terminologies bilingues à partir de corpus ». In *Actes de TALN 2004*.
- PICHON R., SÉBILLOT P. (1999). « Différencier les sens des mots à l'aide du thème et du contexte de leurs occurrences : une expérience ». In *Actes de TALN 99* : 279-288.
- RAPP R. (1999). « Automatic Identification of Word Translations from Unrelated English and German Corpora ». In *Proceedings of Association for Computational Linguistics (ACL)* : 519-525.
- SCHÜTZE H. (1998). « Automatic word sense discrimination ». In *Computational Linguistics* (24) 1 : 97-124.
- TANGUY L. (1999). « Isotopies sémantiques pour la vérification de traduction ». *Actes de TALN 99* : 289-298.
- VÉRONIS J. (2000). Parallel Text Processing : Alignment and use of translation corpora. Kluwer Academic Publishers, Dordrecht.
- VÉRONIS J. (2003). « Hyperlex : cartographie lexicale pour la recherche d'informations ». In *Actes de TALN 2003* : 265-274.
- YAROWSKY D. (1993). « One Sense per Collocation ». In *Proceedings of ARPA Human Language Technology Workshop* : 266-271.