

Enrichir automatiquement des dictionnaires électroniques de synonymes et de traduction : une application du modèle d'appariement multilingue des Atlas sémantiques

Sabine Ploux

▶ **To cite this version:**

Sabine Ploux. Enrichir automatiquement des dictionnaires électroniques de synonymes et de traduction : une application du modèle d'appariement multilingue des Atlas sémantiques. Revue de lexicologie, 2009, 25, pp.109-125. hal-00933543

HAL Id: hal-00933543
https://hal.science/hal-00933543

Submitted on 20 Jan 2014

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

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

Sabine PLOUX

Enrichir automatiquement des dictionnaires électroniques de synonymes et de traduction : une application du modèle d'appariement multilingue des Atlas sémantiques

Résumé

Dans cet article, nous proposons une application du modèle d'appariement sémantique entre langues, défini dans le cadre du modèle géométrique des Atlas sémantiques, à l'élaboration automatique de dictionnaires électroniques de synonymes et de traduction pour de nouvelles langues. Après quelques rappels sur la modélisation géométrique, nous présentons la méthode et nous l'appliquons à l'élaboration d'un dictionnaire de synonymes espagnols et d'un dictionnaire de traduction français-espagnol.

Mots clés : Appariement sémantique, dictionnaire, synonymie, traduction.

1 Introduction

1.1 Contexte sociétal

Les dictionnaires électroniques répondent à un besoin grandissant comme en

témoigne l'évolution des statistiques d'utilisation des ressources disponibles sur le Web. En outre, dans un contexte international de plus en plus présent pour les citoyens, les dictionnaires de traduction ont un impact particulier. Cet impact peut se mesurer à travers le mode d'utilisation de ces mêmes ressources électroniques. Ainsi, à titre d'exemple, le nombre de requêtes de type traduction (français-anglais, anglais-français) représente 70 % de l'ensemble des requêtes arrivant sur le site des Atlas sémantiques, AS dans la suite du texte, (http://dico.isc.cnrs.fr) contre 30 % pour les requêtes monolingues (français ou anglais). Enfin, si l'emploi de la langue anglaise reste dominant sur le Web, cependant, de 1998 à 2005, le pourcentage de pages en anglais est passé de 75 % à 45 % (données publiées par FUNREDES, http://www.funredes.org/ et l'Union latine, http://www.unilat.org/). Ainsi, l'encyclopédie libre Wikipédia contient plus de 6 millions d'articles dans 249 langues.

Dans ce contexte, il devient donc indispensable de proposer des dictionnaires de traduction pour un nombre étendu de langues. De plus, l'avènement d'Internet s'est accompagné d'une modification des modes de constitution des dictionnaires. La possibilité d'élaborer et d'interroger automatiquement des corpus de taille importante, le développement de modèles computationnels de représentation du sens lexical ont mené à une refonte de la conception des ressources lexicographiques. Dans le paragraphe qui suit, nous faisons un bref rappel de recherches computationnelles en lexicographie. Ce rappel nous conduira à positionner le modèle géométrique des AS et à montrer comment il permet d'aborder la question de l'appariement sémantique entre langues.

1.1 Contexte théorique

Il est connu que le découpage et le recouvrement lexical varie d'une langue à l'autre. Pour reprendre des exemples proposés par Chuquet et Paillard (1987), au mot *room* anglais correspond suivant les contextes les mots *pièce*, *chambre*, *bureau*,... du français. Et inversement au mot français *esprit* correspond en anglais plusieurs mots *mind*, *spirit*, *wit*,... Comment alors définir un modèle qui non seulement rende compte du sens lexical dans une langue mais aussi de l'appariement sémantique entre deux langues?

Certains auteurs, comme Abplanalp (1998), insistent sur l'impossibilité de traduire au niveau lexical et proposent comme alternative théorique un recours au niveau conceptuel. Les concepts dépendraient de capacités cognitives humaines générales et partagées. Le réseau lexical WordNet (Fellbaum : 1998, http://wordnet.princeton.edu/) a été conçu sous cette hypothèse. Dans WordNet, le sens de chaque mot est séparé en une liste de concepts lexicaux eux-mêmes représentés par un ensemble de synonymes (Synset). Par exemple le substantif *house* est représenté par 12 Synsets qui déterminent ses différents sens. Le Synset {*firm, house, business firm*} désigne *house* dans le sens de : “the members of a business organization that owns or operates one or more establishments”, le Synset {*family, household, house, home, menage*} désigne *house* dans le sens “a social unit living together”. WordNet est structuré à partir de liens comme l'hyper/hyponymie qui permettent à un utilisateur de naviguer dans l'ensemble du lexique. Cependant, l'hypothèse d'un niveau conceptuel indépendant des langues n'a pas permis à ce jour la réalisation d'un modèle interlingue comme cela était souhaité et

malgré des tentatives pour aboutir.

D'autres modèles lexicaux computationnels d'envergure comme le modèle vectoriel LSA (Landauer, McNamara, Dennis et Kintsch : 2007, http://lsa.colorado.edu/) représentent le sens des mots non plus à partir de concepts lexicaux prédéfinis mais en fonction de leurs contextes d'emploi. Ces contextes d'emploi sont repérés par compilation des relations de cooccurrences dans de larges corpus de texte. Les cooccurrences relevées permettent ensuite de construire un espace sémantique dans lequel le sens de chaque mot est représenté par les proximités qui le lient aux autres mots du lexique. Bien que sensiblement différents de WordNet tant du point de vue des hypothèses théoriques fondatrices que de leur mode de constitution respectif, les modèles vectoriels n'ont pas permis à ce jour de définir un mode d'appariement sémantique entre langues.

Dans le paragraphe qui suit nous présentons brièvement le modèle géométrique des AS. Nous montrons comment il répond à la question de l'appariement sémantique et comment il est utilisé pour l'enrichissement des dictionnaires de traduction et de synonymie.

2 Représentation du sens lexical dans le modèle géométrique des AS

Nous appelons géométrique un modèle qui associe à un mot non plus un vecteur mais un domaine[i] dans un espace multidimensionnel. Ce choix permet à la fois de représenter la structure interne du sens des mots par interprétation des différentes zones qui constituent le domaine et de rendre compte de la similarité sémantique entre

unités lexicales. La similarité sémantique entre deux unités est mesurée à la fois par le recouvrement entre les aires qui leur sont associées et par à la distance qui les sépare dans l'espace multidimensionnel. La figure 1 schématise les aires associées à plusieurs mots et les recouvrements possibles de sens dans un paradigme géométrique.

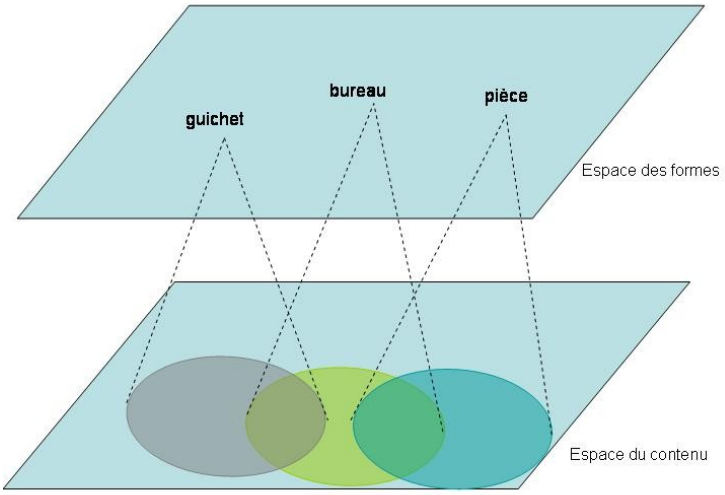


Fig. 1 – Figuration des recouvrements partiels associés à des mots sémantiquement proches.

Le fait de distinguer, comme le montre la figure 1 deux niveaux : celui des formes orthographiques et celui du contenu sémantique permet également de représenter les recouvrements partiels entre deux langues comme cela est figuré par le schéma suivant (figure 2).

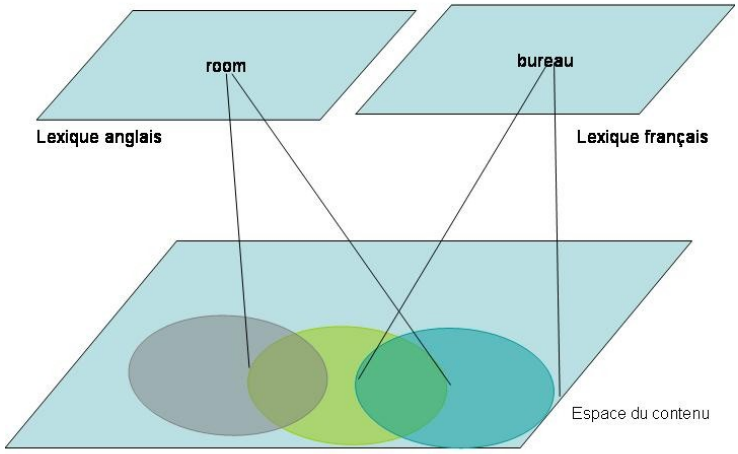


Fig. 2 – Figuration des projections de deux systèmes linguistiques sur un espace de contenu sémantique commun.

Cependant, nous n'avons pas accès directement au contenu sémantique. Nous ne pouvons nous en approcher qu'à travers les traces que ce contenu induit sur les réalisations langagières. La notion de clique est un artefact utile pour cela. Nous rappelons maintenant le principe de construction des cliques et leurs propriétés.

2.1 Cliques, mots et sens

Pour illustrer les notions présentées, nous utilisons ici l'exemple du mot français *insensible*. Cet exemple qui permet de faire la synthèse des différentes caractéristiques du modèle, nous servira aussi à décrire l'appariement sémantique entre langues et l'enrichissement des ressources synonymiques et de traduction vers une langue cible.

Le modèle des AS repose sur deux composantes : une base lexicale et le modèle géométrique à proprement parler.

2.1.1 Organisation de la base lexicale

WordNet définit les concepts lexicaux à partir de synonymes. Un lien de synonymie entre deux mots marque le partage d'un contenu sémantique. C'est aussi à partir d'une base de liens synonymiques et para-synonymiques que nous avons bâti un premier modèle de représentation de la sémantique des mots. Chaque ligne d'une base lexicale est de la forme :

Entrée: mot1, mot2, mot3,...

Par exemple l'entrée *insensible*, de la base lexicale française, comporte de nombreux termes : ... , *anesthésié, cruel*, *dur, détaché, endormi, endurci, engourdi, flegmatique, frigide, froid, féroce, glacial, glacé, impassible, imperceptible, imperméable, imperturbable, impitoyable, implacable, impénétrable, inabordable, inaccessible, inanimé, inapparent, indifférent, indiscernable, indolore, inerte, inexorable, inflexible, inhumain, insaisissable, insignifiant, invisible, invulnérable, léger, léthargique, mort, neutre, négligeable, paralysé, rebelle, rigide, réfractaire, sans coeur, sans entrailles, sec, sourd, stoïcien, stoïque, sévère, timide*,...

Certains de ces termes sont associés à une valeur morale comme *dur*, *inhumain*,... d'autres marquent une valeur physique : *anesthésié*,... ou encore une valeur perceptive qui ne qualifie plus une personne mais un phénomène externe : *inapparent,*....Comment distinguer et organiser ces différentes valeurs sémantiques à partir de la liste de ces synonymes?

2.1.2 La notion de clique

Si la représentation associée à la sémantique d'un mot est, comme nous l'avons supposé, un domaine dans un espace multidimensionnel, nous devons, pour construire ce domaine, représenter les unités qui le composent. Ces unités qui sont de granularité plus fine que celle du mot lui-même sont instanciées dans le modèle par des cliques. Les cliques sont des listes de mots tous synonymes les uns des autres. Du point de vue des aires, une clique est une intersection de plusieurs aires associées à un ensemble de synonymes tous synonymes entre eux (voir la figure 3).

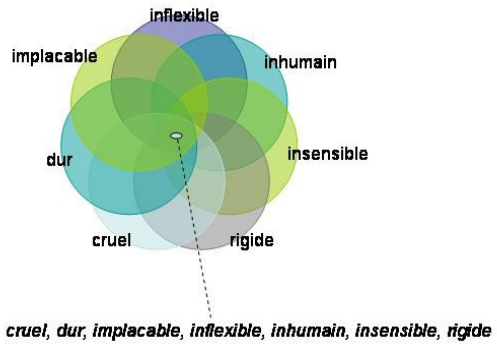


Fig. 3 – Une clique vue comme intersection des aires associées aux mots qu'elle comprend.

Voici des exemples de cliques qui figurent la valeur morale du mot *insensible* :

- *cruel, dur, impitoyable, implacable, inexorable, inflexible, inhumain, insensible*

- *cruel, dur, impitoyable, implacable, inexorable, inflexible, insensible, sévère*
- *cruel, dur, implacable, inflexible, inhumain, insensible, rigide*
- *cruel, dur, implacable, inflexible, insensible, rigide, sévère*
- *,...*

des exemples de cliques qui figurent la valeur physique :

- *anesthésié, insensible*
- *endormi, engourdi, inerte, insensible*
- *engourdi, froid, inerte, insensible*
- *engourdi, immobile, inerte, insensible, paralysé*
- *,...*

et enfin, des exemples pour la valeur perceptive :

- *imperceptible, inapparent, insensible, invisible*
- *imperceptible, indiscernable, insaisissable, insensible, invisible*
- *imperceptible, indiscernable, insensible, léger*
- *,...*

Notons qu'un terme peut appartenir à différentes cliques. Cette caractéristique est due à la non-transitivité de la relation de synonymie. La topologie sous-jacente à

l'ensemble des cliques peut se discerner sur une liste de cliques. Ici nous donnons un chemin dans lequel chaque clique partage au moins un terme avec la précédente, et qui fait passer d'une valeur morale à une valeur physique :

- *cruel, dur, impitoyable, implacable, inexorable, inflexible, insensible, sévère*
- *dur, froid, impitoyable, implacable, insensible, sévère*
- *dur, froid, inaccessible, indifférent, insensible*
- *dur, impassible, indifférent, insensible, stoïque*
- *calme, flegmatique, froid, impassible, imperturbable, insensible*
- *calme, froid, inanimé, insensible*
- *froid, inanimé, inerte, insensible*
- *inanimé, inerte, insensible, mort*

La construction de la forme associée au mot initial, ici *insensible*, permet de faire la synthèse de l'ensemble des liens de proximité.

2.1.3 Construction du domaine associé à un mot

Pour construire l'espace associé à un mot, on utilise une analyse factorielle des correspondances (Benzécri : 1980) sur la matrice des cliques et des termes (voir Ploux : 1997 et Ploux et Victorri : 1998 pour plus de détails sur la méthode employée). Cette analyse permet de calculer les coordonnées des cliques dans un espace multidimensionnel. Les cliques associées à une même valeur sont plus proches, même

si elles ne partagent aucun terme, que celles associées à des valeurs différentes. La projection des cliques sur le plan principal permet de visualiser les résultats. Les cliques sont représentées par des points, les termes sont représentés par l'enveloppe des points-cliques qu'ils contiennent. Enfin, un algorithme de classification permet de distinguer à partir du nuage de points formé par les cliques les différentes valeurs du mot. La figure 4 montre le plan principal associé au terme *insensible* sur lequel quatre valeurs ont été distinguées.

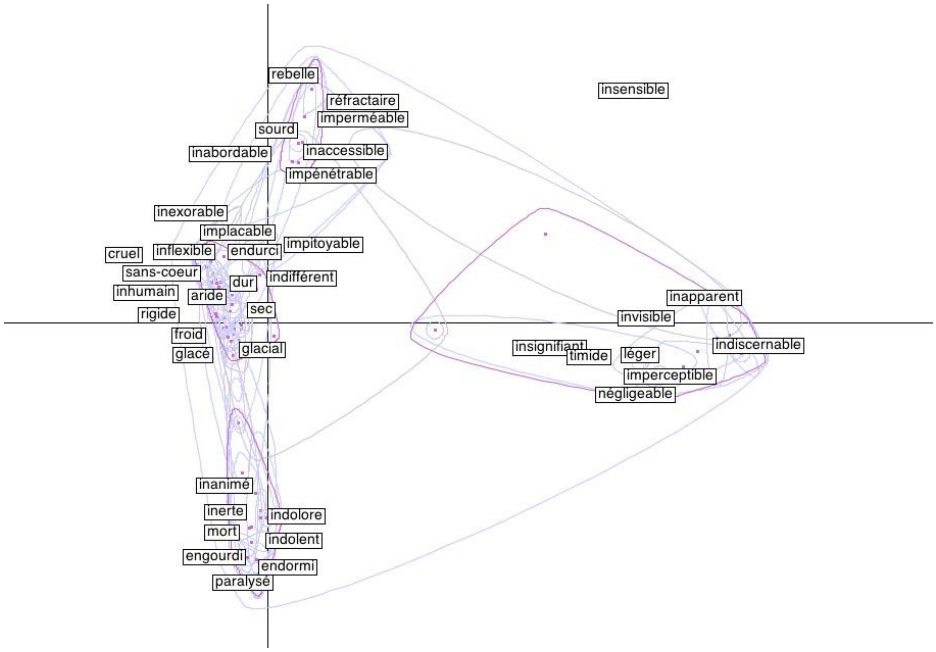


Fig. 4 – Plan principal associé au terme insensible.

Le nuage de points se divise en deux parties. La première, verticale, représente les valeurs d'*insensible* qui qualifient une personne. La seconde, nettement séparée de la première, qualifie un phénomène externe : *inapparent*, *invisible*,... Au centre, près de l'origine des axes, est positionnée la valeur la plus partagée, ici la valeur morale : *dur*, *inhumain*,... Proche de cette valeur se trouve la valeur physique : *engourdi*,… En somme, l'espace géométrique permet de construire une organisation du sens d'un *item*

lexical. L'espace est centré autour d'une valeur générique (quand elle existe), les valeurs proches du point de vue sémantique sont contiguës et les valeurs sémantiques quasi-homonymiques sont nettement distinguées.

3 Le modèle d'appariement sémantique entre deux langues

Nous avons élaboré un modèle d'appariement entre cartes sémantiques de deux langues. Comme nous l'avons précédemment rappelé, le découpage et le recouvrement lexical varient d'une langue à l'autre. Pour cette raison le mot n'est pas l'unité adaptée à la traduction même lexicale. Pour faire face à ce problème, nous avons choisi, dans un modèle de traduction du sens de relier non pas les mots mais les cliques, dont on suppose parce qu'elles représentent des grains plus fins, qu'elles sont moins sensibles aux découpages d'une langue donnée et qu'elles forment de meilleurs candidats pour un ajustement mutuel. Ce modèle dispose sur une même carte les valeurs sémantiques des deux langues et figure ainsi les proximités et les recouvrements. Trois bases lexicales sont nécessaires à l'application du modèle : la base de synonymes de la langue source (dans l'exemple que nous présentons, il s'agit du français), une base de traduction entre langue source et langue cible (il s'agit d'une base français-anglais), et une base de synonymes de la langue cible (l'anglais). Voici les différentes étapes, décrites pour le mot *insensible*, de la méthode d'appariement. Comme expliqué précédemment le système calcule l'espace sémantique du mot (voir figure 4). Ensuite, à partir des mots donnés dans la base de traduction (la base de traduction contient 5 équivalents anglais d'*insensible* : *imperceptible*, *insensible*, *insensitive*, *numb*, *unfeeling*), le modèle calcule l'ensemble des cliques de synonymes de l'anglais qui sont

liées sémantiquement aux cliques du français (voir Ploux et Ji : 2003 pour plus de détails sur la méthode). Voici quelques cliques du français suivies des cliques de l'anglais les plus proches calculées par le modèle :

- *cruel, dur, féroce, impitoyable, implacable, inexorable, inhumain, insensible*
  - *cruel, ferocious, fierce, ruthless, savage*
  - *cruel, inhuman, merciless, pitiless, ruthless, savage*
  - *bitter, cruel, fierce, ruthless, savage*
- *dur, indifférent, inhumain, insensible, sans-coeur*
  - *callous, hard, hardened*
  - *callous, cruel, hard, hard-hearted, heartless*
  - *difficult, hard, tough*
- *imperméable, insensible, rebelle, réfractaire, sourd*
  - *insensitive, unmoved*
  - *impassive, indifferent, phlegmatic, stoical*
  - *callous, impassive, insensible, unfeeling*
- *engourdi, froid, inerte, insensible*
  - *dull, inanimate, inert, lifeless*

- *dull, expressionless*
- *dead, inanimate, inert, lifeless*

➢ *engourdi, immobile, inerte, insensible, paralysé*

- *dead, numb, paralytic*
- *asleep, numb*
- *dull, dulled*

Les résultats sont ensuite projetés sur une carte qui met en évidence les appariements entre les valeurs sémantiques de la langue source et celles de la langue cible (figure 5).

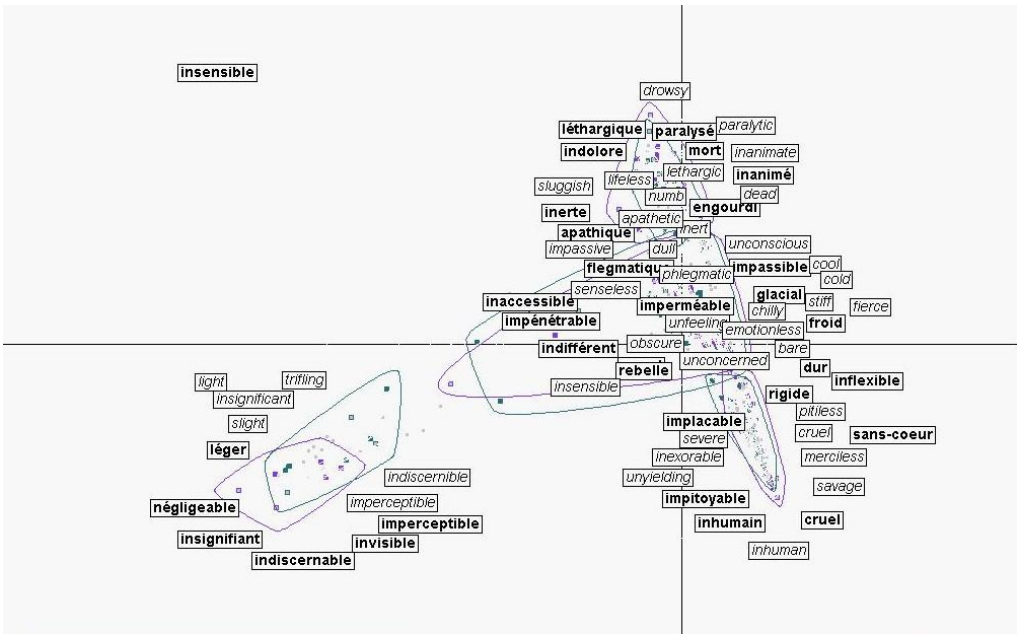


Fig. 5 – Résultat de l'appariement construit à partir de la vedette insensible du français.

Ainsi, le modèle d'appariement permet de projeter un espace sémantique lexical d'une langue source vers un espace sémantique d'une langue cible en respectant

l'appariement des valeurs que les mots peuvent prendre en contexte.

4 Comment utiliser le modèle d'appariement pour l'enrichissement de bases lexicales de synonymes et de traduction?

Nous avons vu au paragraphe précédent qu'à partir d'une base de traduction pauvre (il y a 5 équivalents vers l'anglais du mot français *insensible*), le modèle est capable de proposer à la fois de nouvelles traductions possibles mais aussi des termes candidats pour la synonymie dans la langue cible. C'est cette capacité qui a été exploitée pour l'enrichissement de dictionnaires de traduction et de synonymes pour une nouvelle langue. Le modèle d'enrichissement utilise des bases lexicales constituées (comme la base française ou la base anglaise dont nous disposons). Il doit également être amorcé par des bases lexicales de la langue cible qui peuvent être très réduites[ii]. La méthode, incrémentale, consiste ensuite à générer des espaces sémantiquement appariés vers la langue cible à partir des espaces des langues source. Ce processus permet de proposer à la fois de nouveaux termes candidats pour la traduction et de nouveaux termes candidats pour la synonymie dans la langue cible. Une étape de validation est alors nécessaire pour sélectionner parmi les termes candidats ceux qui sont intégrés aux dictionnaires. Dans le programme actuel cette étape de validation tient compte du nombre de cliques qui contiennent le terme candidat à valider et de la distance de ce terme aux synonymes (resp. traductions) déjà contenus dans les bases de synonymes (resp. traduction). Une fois la validation effectuée et les termes ajoutés aux bases lexicales, le processus est répété. Nous montrons ici, sur la base du mot *insensible*, comment cette méthode a été utilisée pour enrichir un dictionnaire de

synonymes espagnols et un dictionnaire de traduction français-espagnol.

Au départ, nous avons amorcé deux bases lexicales[iii] :

- Une base de traduction français-espagnol (le nombre moyen d'équivalents par mot est 1,4)
- Une base de synonymes espagnols (le nombre moyen de synonymes par mot est 1,3)

Par exemple, dans ces bases initiales, *insensible*$_{fr}$ n'a qu'une seule traduction : *insensible*$_{es}$ et à : *insensible*$_{es}$ n'est associé aucun synonyme espagnol. Voici étape après étape les mots proposés par le modèle d'enrichissement. Après chaque étape, les termes candidats qui respectent les contraintes de proximités sémantiques mentionnées plus haut, sont validés et intégrés aux bases respectives.

2.1 Résultats obtenus à l'étape 1

La méthode d'appariement entre français et espagnol appliquée au mot *insensibl*$_{fr}$ du français produit la carte suivante (figure 6) sur laquelle sont associés des synonymes français et des équivalents espagnols.

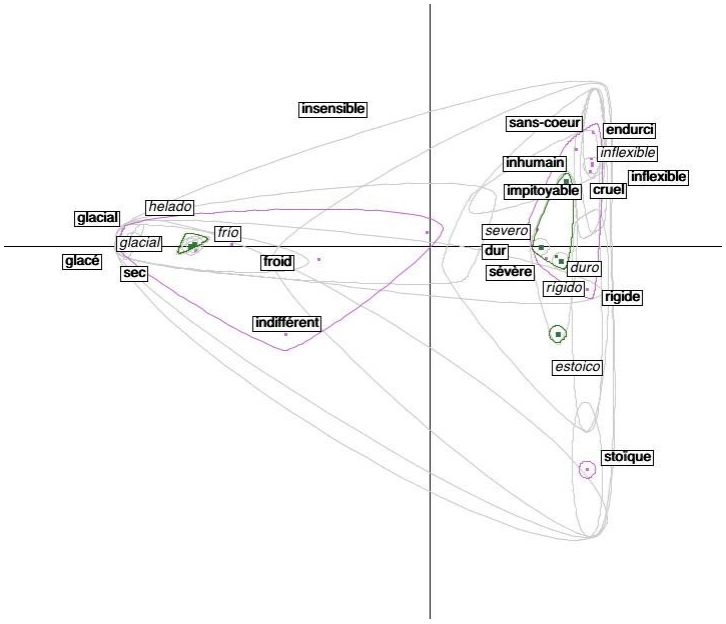


Fig. 6 – Carte d'appariement sémantique proposant des termes candidats à la traduction vers l'espagnol du mot insensible$_{fr}$ du français.

L'étape de validation sélectionne les synonymes suivants présentés sur la nouvelle carte du mot *insensible*$_{es}$ en espagnol (figure 7).

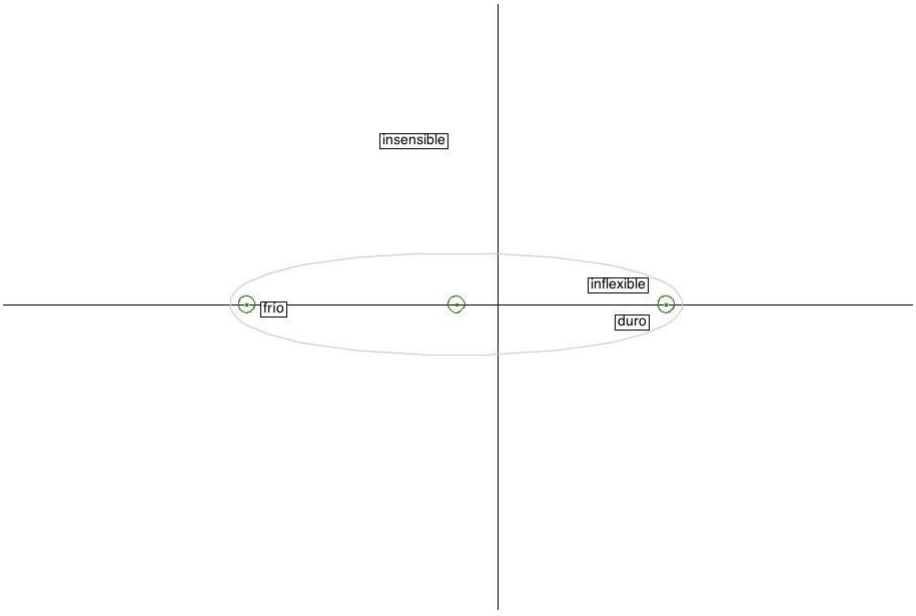


Fig. 7 – Synonymes espagnols du mot insensible$_{es}$ calculés et sélectionnés.

2.2 Résultats obtenus à l'étape 2

Une fois les nouveaux termes candidats validés et intégrés, on applique à nouveau le modèle d'appariement. On pourra remarquer que les nouvelles cartes produites sont plus riches à la fois en nombre de termes proposés et en valeurs sémantiques distinguées. Ainsi, la valeur sémantique associée à l'insensibilité émotionnelle représentée par les termes *inalcanzable, inaccessible*, *inasequible*, *impenetrable* de l'espagnol et qui était absente au premier passage (figure 6) apparaît sur la nouvelle carte (figure 8).

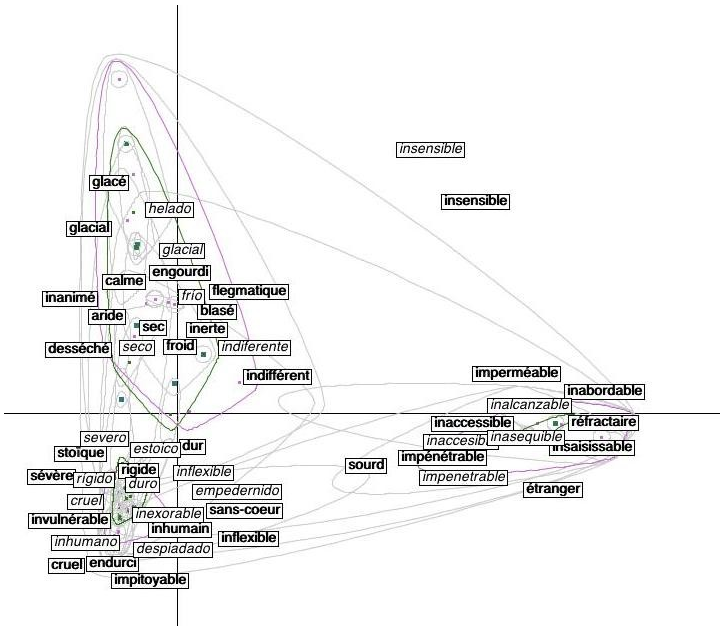


Fig. 8 – Termes candidats obtenus par application du modèle d'appariement pour le mot insensible$_{fr}$ construits à partir des dictionnaires augmentés à l'étape 1.

La validation permet de construire une nouvelle carte sémantique de termes synonymiques (figure 9).

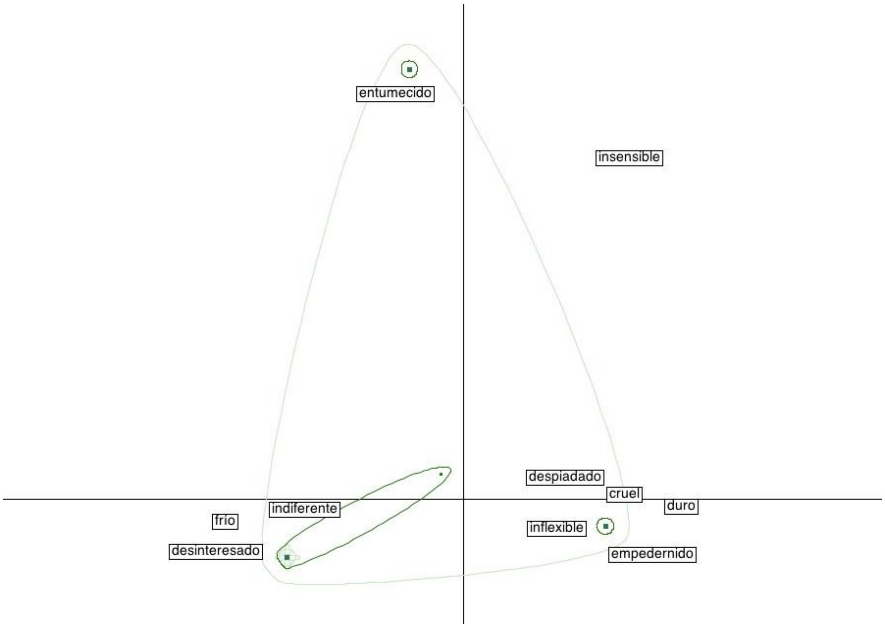


Fig. 9 – Synonymes espagnols du mot insensible$_{es}$ calculés et sélectionnés.

2.3 Résultats obtenus à l'étape 3 et à l'étape 4

Nous donnons ici les résultats obtenus à l'étape 3, pour l'appariement entre langues (figure 10):

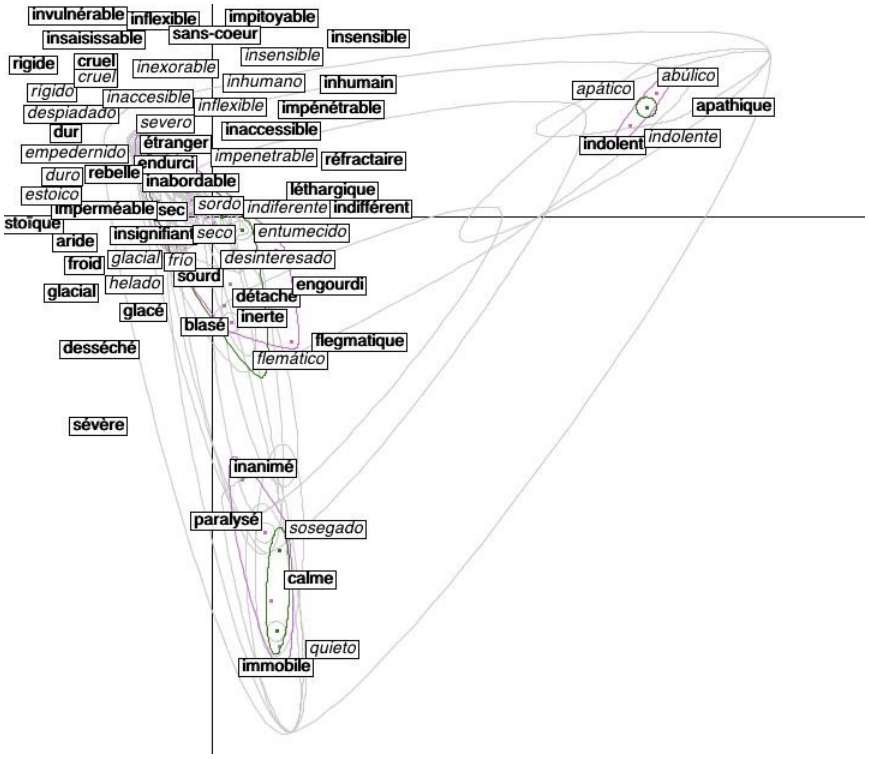


Fig. 10 – Termes candidats obtenus par application du modèle d'appariement pour le mot insensible$_{fr}$ construits à partir des dictionnaires augmentés à l'étape 2.

pour la synonymie en espagnol (figure 11).

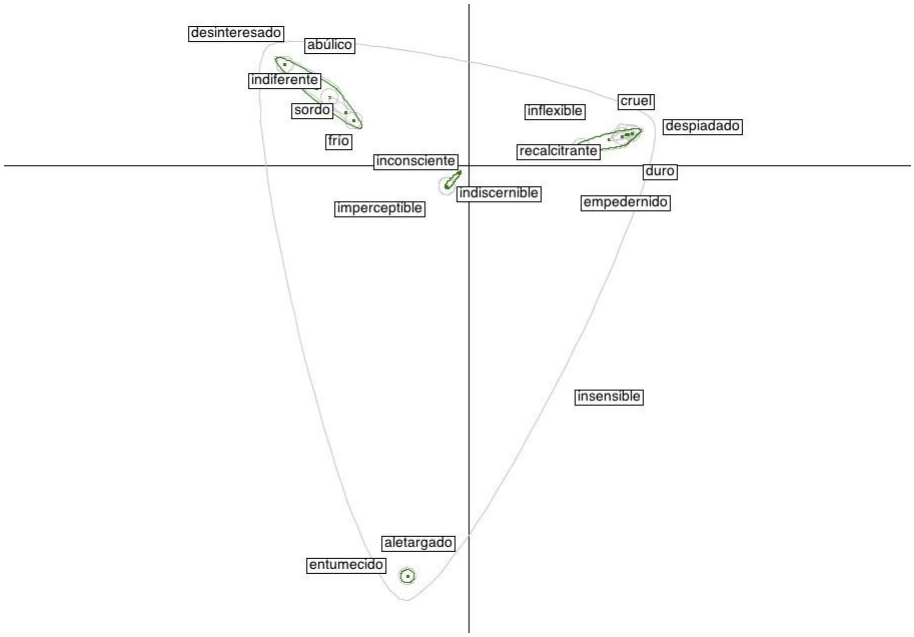


Fig. 11 – Synonymes espagnols du mot insensible$_{es}$ calculés et sélectionnés.

Sur la carte obtenue à l'étape 4 (figure 12), on voit apparaître la valeur sémantique du mot *insensible$_{fr}$* précédemment dénommée valeur perceptive. Elle contient les termes français *invisible*, *léger*, *insignifiant*, *imperceptible*,... et les termes espagnols *imperceptible*, *invisible*, *insignificante*,...

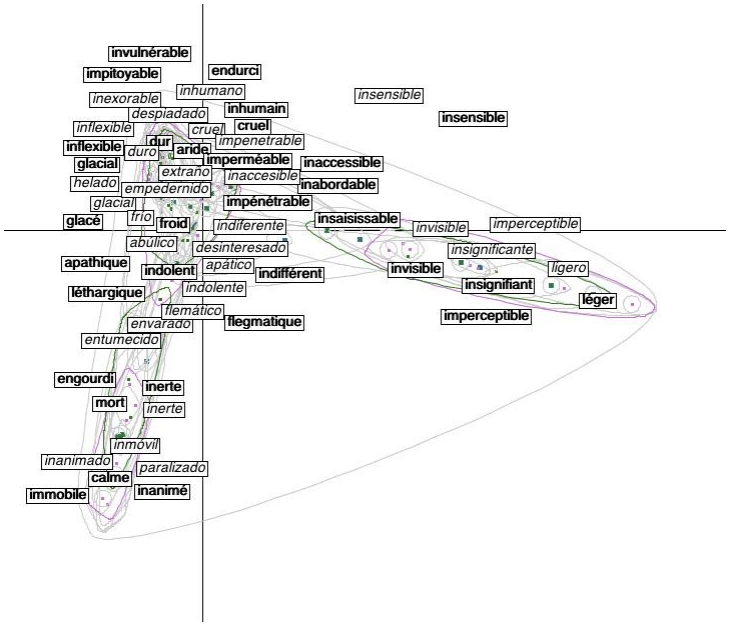


Fig. 12 – Termes candidats obtenus par application du modèle d'appariement pour le mot insensible$_{fr}$ construits à partir des dictionnaires augmentés à l'étape 3.

5 Conclusion

Nous avons montré comment le modèle d'appariement sémantique entre langues permet l'enrichissement de bases lexicales de traduction et de synonymie et nous l'avons appliqué à l'espagnol. Nous projetons d'utiliser les résultats permis par ce modèle pour la création de bases lexicales dans de nouvelles langues. En plus de l'utilité que ces nouvelles bases pourront représenter pour tous et pour la communauté scientifique, leur réalisation permet d'envisager la conception de nouveaux outils linguistiques comme la définition d'une mesure de la proximité sémantique entre langues. En effet, la typologie des langues a favorisé l'étude des propriétés phonologie et morphologique. Cependant, plus récemment, plusieurs auteurs ont cherché à élaborer une typologie sémantique universelle afin comparer la manière dont différentes

langues expriment des concepts similaires (voir Talmy : 2000). Cette notion d'universalité en sémantique fait débat, certains auteurs y opposent les variations entre langues des découpages sémantiques (voir Gumperz et Levinson : 1996, Bowerman : 1996). Dans ce domaine, l'usage de la synonymie permet de repérer la relation entre un sème et les visions conceptuelles retenues par les locuteurs de différentes langues. Cependant, les différentes études réalisées ont traité des groupe de mots très spécifiques (verbes de mouvement, de localisation, de position,...). De plus, ces différentes études ont été réalisées « à la main », sans usage de méthode permettant un traitement systématique. Dans ce contexte, nous proposons d'utiliser le modèle géométrique des AS pour bâtir un outil permettant une mesure objective des comparaisons des unités lexicales de deux langues sur la base de leur recouvrement sémantique. En particulier, cet outil de mesure sémantique pourrait permettre de tester l'hypothèse de relexification des langues créoles. La relexification des langues créoles est l'hypothèse suivante laquelle le lexique de ces langues est fortement semblable à celui de leur langue cible et plutôt éloigné de celui de leur langue source. Cette hypothèse de proximité lexicale entre créole et langue cible implique qu'il existe une application entre les deux lexiques avec fort recouvrement des espaces sémantiques lexicaux associés aux mots et à leur traduction. Le modèle géométrique devrait alors permettre de vérifier à grande échelle les prédictions de cette hypothèse en comparant de façon systématique les lexiques des différentes langues en interaction.

Sabine Ploux,

L2C2, Institut des sciences cognitives,

CNRS-Université Lyon I,

Lyon,

France.

Bibliographie

[1] ABPLANALP, L., *La pertinence et la traduction*. Travaux du centre de traduction littéraire, Lausanne. 1998.

[2] BENZÉCRI, J.-P., *L'analyse des données : l'analyse des correspondances*. Bordas, Paris. 1980.

[3] CHUQUET, H. ; & M. PAILLARD, *Approche linguistique des problèmes de traduction anglais-français*. Ophrys, Paris. 1987.

[4] FELLBAUM, C., editor, *Wordnet, An Electronic Lexical Database*. MIT Press, Cambridge, Massachusetts. 1998.

[5] GUMPERZ, J. ; & S. LEVINSON, editors, *Rethinking Linguistic Relativity*. Cambridge University Press, Cambridge. 1996.

[6] LANDAUER, T. K. ; D. S. MCNAMARA; S. DENNIS; & W. KINTSCH, *Handbook of Latent Semantic Analysis*. Lawrence Erlbaum Associates, Mahwah. 2007.

[7] PLOUX, S. 1997. « Modélisation et traitement informatique de la synonymie ». Linguisticae Investigationes, vol. 21, no. 1, pp. 1–28.

[8] PLOUX, S. ; & H. JI. 2003. « A model for matching semantic maps between languages

(French/English, English/French) ». Computational Linguistics, vol. 29, no. 2, pp. 155–178.

[9] PLOUX, S. ; & B. VICTORRI. 1998. « Construction d'espaces sémantiques à l'aide de dictionnaires informatisés des synonymes ». Traitement Automatique des Langues, vol. 39, no. 1, pp. 161–182.

[10] TALMY, L., *Toward a cognitive semantics*. MIT Press, Cambridge, Massachusetts. 2000.

i *Un domaine est tout d'abord un espace topologique, mais on suppose, de surcroît, que chacun de ses points possède un voisinage homéomorphe à un ouvert de* R^n *(c'est-à-dire qu'il existe une application bi-continue du voisinage vers un ouvert de* R^n*). Par exemple, les aires délimitées par un disque, un carré, un rectangle ou une ellipse sont des domaines.*

ii *On peut envisager d'initialiser ces bases à la main sur quelques termes ou bien de proposer des synonymes et des traductions à partir de similitudes de contexte.*

iii *Ces bases ont été réalisées par deux stagiaires lexicographes : M. Becquart et M-A Perez.*