



Repérage automatique de la néologie sémantique en corpus à travers des représentations cartographiques évolutives : vers une méthode de visualisation graphique dynamique de la diachronie de la néologie

Armelle Boussidan, Anne-Lyse Renon, Charlotte Franco, Sylvain Lupone,
Sabine Ploux

► To cite this version:

Armelle Boussidan, Anne-Lyse Renon, Charlotte Franco, Sylvain Lupone, Sabine Ploux. Repérage automatique de la néologie sémantique en corpus à travers des représentations cartographiques évolutives : vers une méthode de visualisation graphique dynamique de la diachronie de la néologie. Cahiers de Lexicologie, 2012, pp.1-18. hal-00934820

HAL Id: hal-00934820

<https://hal.science/hal-00934820>

Submitted on 22 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.

Armelle BOUSSIDAN, Anne-Lyse RENON, Charlotte FRANCO,

Sylvain LUPONE & Sabine PLOUX

Repérage automatique de la néologie sémantique en corpus à travers des représentations cartographiques évolutives

*Vers une méthode de visualisation graphique dynamique
de la diachronie des néologismes*

Résumé

La néologie sémantique présente la complexité d'un changement à la fois subtil et progressif dont la temporalité reste abstraite et difficile à cerner. Afin de pouvoir la détecter et l'analyser de façon automatique, nous proposons de nous appuyer sur des mécanismes de pivots du sens. Ces pivots sont repérables à l'aide d'une extension d'ACOM s'inscrivant dans le paradigme des AS (Atlas Sémantiques ®). Cette extension, en cours de développement, constituera un outil très précis qui présentera l'avantage d'être :

- objectif (pas de filtrage manuel ou semi-automatique subjectif de la part du linguiste)
- d'un grain extrêmement fin tant du point de vue temporel que sémantique
- dynamique et adaptable à d'autres corpus

En considérant que la néologie sémantique est une réorganisation structurale de la polysémie d'un terme dans le temps, nous adapterons l'analyse de la polysémie en synchronie que permet d'effectuer l'ACOM aux besoins de la diachronie.

En effet, dans les représentations produites par l'ACOM, les sens s'organisent en termes de distance sémantique et par rapport à l'origine des axes qui représente le cœur prototypique du sens. Ainsi les sens périphériques sémantiquement apparaissent également comme périphériques graphiquement. Il s'agira donc de repérer si un sens périphérique change de statut car cela reflète plusieurs mécanismes linguistiques pouvant être à l'œuvre dans les processus de néologie. La mise en évidence de ces processus sera effectuée en partant des données sans typologie préalable. Nous présenterons les avantages d'une méthode de regroupement de ces mécanismes sous un même observatoire graphique, notamment la réadaptabilité à différents corpus ou échelles temporelles. Nous proposons ainsi une approche exploratoire objective de détection et d'analyse de la néologie sémantique qui sera illustrée au travers d'une étude de cas basée sur le corpus *Le Monde 1997-2004*.

Abstract

Semantic change ¹ is challenging as is it a subtle and progressive type of change relying on an abstract temporality. In order to detect and analyse it we propose to take advantage of semantic pivot's mechanisms that we detect with the help of an extension ACOM created within the paradigm of the Semantic Atlases (SA). This extension, currently in the course of development, shall constitute a very fine grained tool offering the following advantages:

- objectivity (no manual filtering carried out by the linguist)
- to be fine grained on the semantic and temporal planes

- to be dynamic and adaptable to other corpora

Considering that semantic neology (taken here as semantic change at large) is a structural reorganization of a term's polysemy across time, the synchronic analysis of polysemy offered by ACOM can be adapted to the needs of diachrony.

In ACOM representations, meanings are organized in terms of semantic distance, and relatively to the origin of axes which represents the prototypical core of meaning. Therefore meanings which are semantically peripheral appear as graphically peripheral too. We shall then detect whether a peripheral meaning undergoes status change reflecting several linguistic mechanisms at work in the process of semantic change. To bring to light these processes we use data-driven detection with no pre-established typology. We shall present the advantages of a method that is merging several mechanisms under the same umbrella, offering a unified graphical observatory and allowing for adaptability to various corpora and time scales. We therefore offer an objective exploratory approach for the detection and analysis of semantic change, which is illustrated here with examples taken from the French press corpus *Le Monde* 1997-2004.

Introduction

La démarche présentée dans cet article est celle que poursuit notre équipe pour adapter l'ACOM (Automatic Contexonym Organizing Model) à la diachronie aussi bien du point de vue de la détection des changements sémantiques que du point de vue de leur analyse. Pour ce faire, un certain nombre d'indices se révèlent être utiles pour assister notre démarche. Ces indices, actuellement en cours de test, sont de nature à la fois mathématique, géométrique, graphique et linguistique. Dans cet article, la mise en œuvre de ces indices prend appui sur un exemple de changement sémantique de l'ordre de la connotation, celui du mot *mondialisation*, par contraste avec son synonyme *globalisation* qui présente une stabilité de connotation. Cette analyse est par la suite complétée d'une proposition de visualisation cartographique dynamique s'appuyant sur l'interpolation des cartes sémantiques temporelles générées via une applet java. Pour cela nous avons fait intervenir les outils du design graphique pour faire évoluer la méthode de visualisation actuelle vers une visualisation dynamique.

Le modèle

Dans le cadre de notre recherche en diachronie nous nous appuyons sur l'ACOM (Ji *et.al* 2003 ; Ji *et.al* 2008) qui s'inscrit dans le paradigme géométrique des Atlas Sémantiques ® (Ploux, 97 ; Ploux *et.al* 2010) pour détecter et analyser la néologie sémantique entendue ici dans sa définition la plus large.¹ Ce modèle computationnel propose dans son état initial une représentation du sens lexical statique extrêmement fine. Il s'agit d'un modèle géométrique de représentation du sens qui génère des cartes à partir d'une analyse factorielle des correspondances (Benzécri, 1980). La recherche en cours vise à la fois à adapter le modèle à une représentation dynamique en diachronie ainsi qu'à conduire une analyse du changement sémantique qui se fonde sur une approche computationnelle statistique, automatisée, géométrique et non typologique *a priori*.

Le modèle des Atlas Sémantiques (AS) est originellement construit sur un corpus de synonymes. L'ACOM est dérivé des AS et traite des relations de co-occurrence entre les mots (« contexonymes ») à partir de n'importe quel corpus, dans le même paradigme géométrique.

¹ Nous ne reviendrons pas ici sur les riches débats soulevés lors du colloque quant à la définition de la néologie sémantique. Bien que cela soit discutable, nous entendons par *néologie sémantique* « changement sémantique » et son équivalent anglo-saxon « semantic change » dans le cadre de cette étude.

Un algorithme de hiérarchie permet une division en clusters thématiques. Ces modèles produisent des cartes sémantiques qui sont navigables de façon interactive et peuvent être consultées librement à l'adresse <http://dico.isc.cnrs.fr/fr/index.html>.

L'originalité du modèle initial repose sur l'unité de clique, unité mathématique et infra-linguistique. Les cliques sont des sous-unités de sens composées par des listes de termes à l'intérieur desquelles chaque mot est lié à tous les autres par une relation : celle de synonymie dans les AS ou de co-occurrence dans l'ACOM. Les cliques sont organisées dans un paradigme continu. Le passage d'une clique à la suivante s'effectue par un terme commun ; ce terme peut être considéré comme un pivot sémantique autour duquel s'articule le passage d'une valeur sémantique à une autre. Cette notion est fondamentale en diachronie car nous cherchons à observer ces pivots non plus comme des points d'articulation figés mais comme de potentiels points d'articulation actifs dans le temps.

Le modèle attribue des coordonnées spatiales à chaque clique qui correspond sur la carte à un point. Ceci diffère de l'approche consistant à attribuer des coordonnées à chaque mot dans les autres modèles vectoriels de représentation lexicale tels que LSA (Landauer *et.al*, 2007), HAL (Burgess & Lund, 1997) ou d'autres Word Space Models (Sahlgren, 2006). Dans le paradigme des AS à l'inverse on construit un espace de représentation du sens en partant de la polysémie. Le modèle génère ainsi une visualisation multidimensionnelle qui permet trois niveaux de navigation conceptuelle : le niveau des cliques (représentées par des points), le niveau des mots (représentés par les enveloppes contenant les cliques) et le niveau des ensembles conceptuels ou clusters thématiques (représentés par de plus larges enveloppes).

L'organisation en cliques, qui est la particularité du modèle des AS permet une analyse très fine du passage d'une valeur sémantique à l'autre lors d'une analyse en synchronie. Cette spécificité sert les besoins de la diachronie, car un changement sémantique est le passage d'une valeur à l'autre dans le temps et ce passage peut être détecté au travers de changements de contextes d'emploi avec l'ACOM. La finesse de l'outil permet de tracer jusqu'aux premiers battements d'ailes du glissement sémantique d'un terme. Nous avons également par ce biais la possibilité d'observer des changements éphémères.

Le travail présenté ici repose sur l'interrogation du modèle à partir d'un exemple et non sur une détection de schéma dans les données. Néanmoins nous nous dirigeons vers la démarche inverse : partir des données brutes pour extraire semi-automatiquement des candidats au changement et élaguer les sorties pour l'analyse. Dans ce but nous testons des indices permettant de créer un modèle de détection.

Le corpus

Le corpus utilisé pour ce travail est constitué des archives complètes du journal « Le Monde » sur la période 1997-2004. Le corpus a été découpé en sections temporelles d'un mois, pour les besoins de la diachronie.

Indices

Pour détecter un changement sémantique nous nous appuyons sur les indices suivants :

- fréquence du terme étudié,
- classification des occurrences en fonction des sections d'apparition dans le corpus,
- réseaux de co-occurrence du terme dans le corpus et par phases,

- rangs hiérarchiques de co-occurrences,
- productivité néologique.

Nous ajoutons à cela une mesure que nous avons développée au regard des spécificités du modèle :

- la mesure de variabilité de la densité et de la cohésion.

La fréquence est un indice qui permet de voir la stabilité d'usage d'un terme. Lorsque la fréquence subit des variations importantes, le sens peut être affecté au travers des événements qui le mobilisent ou d'effets de modes linguistiques. Le corpus de presse, de par sa nature, est tout particulièrement lié aux événements d'actualité. Nous abordons la dimension de l'événement en post-traitement pour des questions d'objectivité linguistique.

Les sections que comporte le corpus au préalable permettent de compléter l'apport des outils de décryptage plus complexes en donnant des renseignements sur le type d'emploi d'un terme au sein du corpus. En effet, certains termes spécialisés appartiennent plus à une rubrique qu'à une autre, comme l'économie ou la technologie. L'apparition nouvelle d'un terme dans une rubrique procure donc un indice sémantique. Nous pourrions illustrer cela de façon fictive en imaginant les statistiques d'apparition du mot *souris* dans des rubriques animalières et technologiques dans un corpus généraliste couvrant les premières années du développement de l'informatique.

L'étude des réseaux de co-occurrence permet d'observer l'usage en contexte d'un terme. Lorsque le réseau est stable, le contexte d'emploi l'est également : le sens est solidement ancré dans un ou plusieurs contextes repérables et définis. Le réseau d'un terme est observable en faisant la moyenne des co-occurents sur l'ensemble du corpus et en les organisant de façon hiérarchique. Si le réseau associé à un terme se modifie, alors ce terme est susceptible de subir un changement sémantique, changement de connotation ou de domaine, étoffement, extension, restriction ou disparition, etc., de façon stable ou éphémère.

Pour observer ces modifications le réseau est soumis à un découpage temporel, en fonction du comportement fréquentiel observé au préalable. En recalculant les fréquences des termes co-occurents par tranches temporelles, on obtient des réseaux comparables qui permettent d'observer en diachronie la réorganisation du réseau d'un mot dans le corpus. À ce stade la dimension de l'impact événementiel peut être prise en compte dans le découpage. On peut donc intégrer la dimension de l'événement et conduire une nouvelle analyse sur les hiérarchies de co-occurrences obtenues.

Cependant, les réseaux que nous mettons à jour peuvent présenter un aspect similaire mais posséder une hiérarchie interne différente. Cette hiérarchie peut être étudiée grâce à l'analyse des rangs de co-occurrence.

Les rangs de co-occurrence permettent d'élucider la structure hiérarchique des réseaux de co-occurrence. En effet, si un terme co-occurent prend de l'importance fréquentielle, cela peut être sans conséquence pour les autres termes ou au détriment d'un certain nombre d'entre eux. Plus un co-occurent se rapproche du 1^{er} rang plus il a d'importance contextuelle. Dans le temps on peut observer si un terme co-occurent voit son rang changer de position de façon durable ou éphémère, ainsi que la réorganisation des autres co-occurents face à ce changement.

La productivité néologique permet de prendre en compte des données linguistiques qui illustrent la place d'un élément dans le système. En effet, si un terme devient productif, le

terme créé influe à son tour sur le sens du terme source. La production d'un néologisme formel fait également le pont avec la dimension de l'évènement, puisqu'elle implique un référent réel jusque là non nommé.

Enfin, la mesure de variabilité de la densité et de la cohésion que notre équipe « modèles mathématiques et informatiques pour le langage » a récemment développée est spécifique au paradigme des AS. Cette mesure se base sur la production de cliques effectuée par le modèle. Pour chaque terme en entrée le modèle génère un certain nombre de termes associés dans l'ACOM ainsi qu'un certain nombre de cliques liées à ces termes. Plus il y a de termes générés, plus le terme en entrée apparaît dans des contextes variés. Un foisonnement de termes associés peut être dû ou à l'étoffement des contextes en présence, ou à l'addition de nouveaux contextes. Plus il y a de cliques générées, plus ce terme a de sous-champs sémantiques continus. Un nombre de cliques élevé prenant le pas sur un nombre de cliques stables peut refléter un besoin de créer une continuité entre des contextes déjà présents ou entre de nouveaux contextes qui se rattachent aux précédents. Nous nous attachons surtout au rapport possible cliques-termes et à la variation de ce rapport dans le temps. En effet s'il y a beaucoup plus de cliques que de termes cela veut dire que le même terme peut être utilisé comme pivot sémantique à plusieurs reprises dans de nombreuses cliques pour créer une cohésion entre les valeurs. Si le rapport subit une grande variation, c'est que le chemin de continuité pour passer d'une valeur à l'autre est devenu plus long ce qui indique une multiplication des contextes. À l'inverse, si un mot n'apparaît que dans une seule clique c'est une unité de sens contextuellement isolée et pauvrement reliée au réseau existant de valeurs contextuelles. Si le nombre de termes nouveaux s'équilibre avec le nombre de cliques créées, le nombre de contextes en jeu est stable. Une différence soudainement élevée entre le nombre de cliques et le nombre de termes reflète ainsi un besoin de créer une nouvelle continuité dans la structuration conceptuelle entre les nouveaux termes co-occurents et entre les nouveaux contextes d'emploi.

Nous appliquons cette série d'indices lors d'une étude de cas : l'évolution de la connotation du terme *mondialisation* par contraste avec son synonyme *globalisation* qui présente une stabilité sémantique dans le temps.

1. Une étude de cas : évolution de la connotation du terme *mondialisation* par contraste avec *globalisation*

Ces termes ont été choisis car ils répondent à l'hypothèse que la coexistence de termes quasi-parfaitement synonymes suppose que l'un d'entre eux voire les deux subiront une modification (selon l'hypothèse saussurienne que les éléments de la langue n'existent que par leur relation d'opposition entre eux). Ces termes présentent un phénomène de concurrence synonymique. Seul *mondialisation* présente une néologie sémantique fine qui se double des effets de la création de néologismes formels par dérivation. Comparons tout d'abord les définitions de *mondialisation* et *globalisation* dans le *Petit Robert* 2010 :

- Mondialisation
n.f -1953 de mondial. Le fait de devenir mondial, de se répandre dans le monde entier. *La mondialisation d'un conflit*. SPECIALT Phénomène d'ouverture des économies nationales sur un marché mondial libéral, lié aux progrès des communications et des transports, à la libéralisation des échanges, entraînant une interdépendance croissante des pays. → **globalisation** (ANGLC.)
→ altermondialisme, antimondialisation

- Globalisation
n.f.-1968 de globaliser. 1. Action de globaliser, son résultat. 2. (de l'anglais *globalization*) ANGLC. Mondialisation. *La globalisation des marchés*.

On observe que les deux termes renvoient à une même notion bien que *mondialisation* ait une définition plus précise que *globalisation*, auquel il renvoie comme synonyme. *Globalisation* est par ailleurs un anglicisme, qui est entré dans le lexique français tout particulièrement au travers de son emploi dans le domaine de la finance ainsi que via des traductions de dépêches 'calquées'.

On peut noter que bien des locuteurs et des écrivains prennent la peine de préciser dans quel sens ils emploient ces termes. On entend et on lit souvent « mondialisation économique », « mondialisation financière » ou « mondialisation politique ». Pour certains encore la *globalisation* va plus loin que la *mondialisation* alors que pour d'autres la *globalisation* se limite à la sphère financière (www.techno-science.net). Cet état d'incertitude mérite d'être noté car il est un signe d'instabilité sémantique, et indique que la définition de ces termes n'est pas assujettie à un consensus établi.

1.1 Fréquence

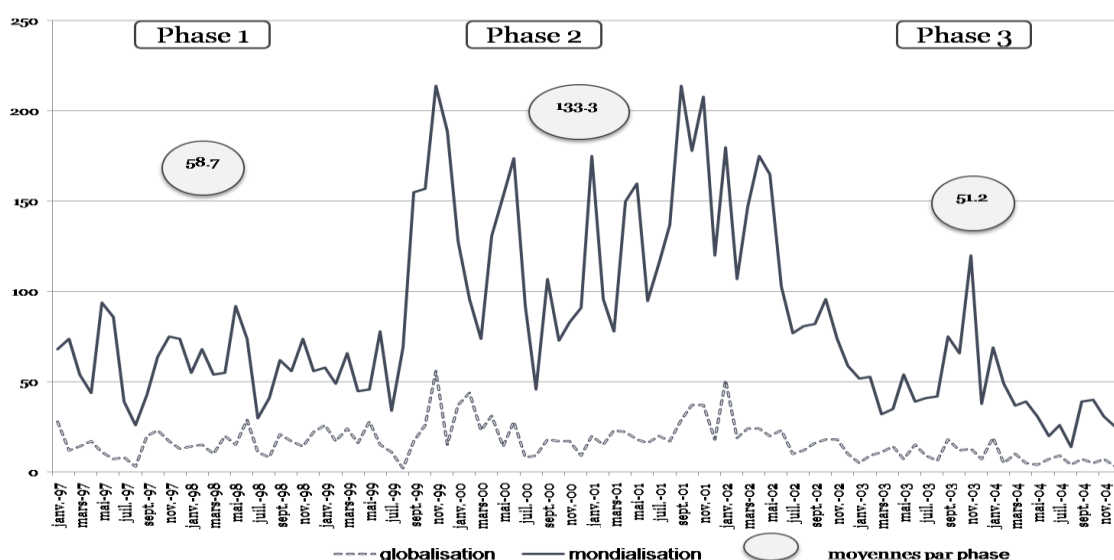


FIGURE 1 : Fréquence d'occurrence des termes *globalisation* et *mondialisation* sur l'ensemble du corpus *Le Monde* 1997-2004, et moyennes d'occurrence de *mondialisation* sur les 3 phases fréquentielles qui se dégagent. L'unité temporelle choisie ici est le mois.

On peut observer en premier lieu que la fréquence d'emploi de *globalisation* (1600 occ. au total) reste stable et que le terme est utilisé en moyenne 8,5 fois moins que son concurrent *mondialisation* (13637 occ. au total).

On peut découper le comportement de *mondialisation* en trois phases fréquentielles : une phase de janvier 1997 à juillet 1999 où la moyenne des occurrences est de 58,7 ; une seconde phase d'août 1999 à juin 2002 avec une moyenne de 133,3 et une troisième phase, de juillet 2002 à novembre 2004 avec une moyenne de 51,2 qui rejoint l'échelle de la première phase (fig.1). Nous conserverons ce découpage en phases dans la suite de l'analyse et essaierons de dégager ce qui se passe lors de la seconde phase.

1.2 Rubriques

Les termes *mondialisation* et *globalisation* entrent tous deux au sein de deux rubriques d'importance majeure dans le corpus *Le Monde* en mars 2000 : les rubriques « Éditorial et analyses » ainsi que les rubriques « Débats et décryptage » (fig 2). Avant cette date il n'y a aucune occurrence de ces termes dans ces rubriques. Cela indique que les termes arrivent à cette époque sur le devant de la scène médiatique. Cette entrée succède de quelques mois à l'envolée fréquentielle observée ci-dessus.

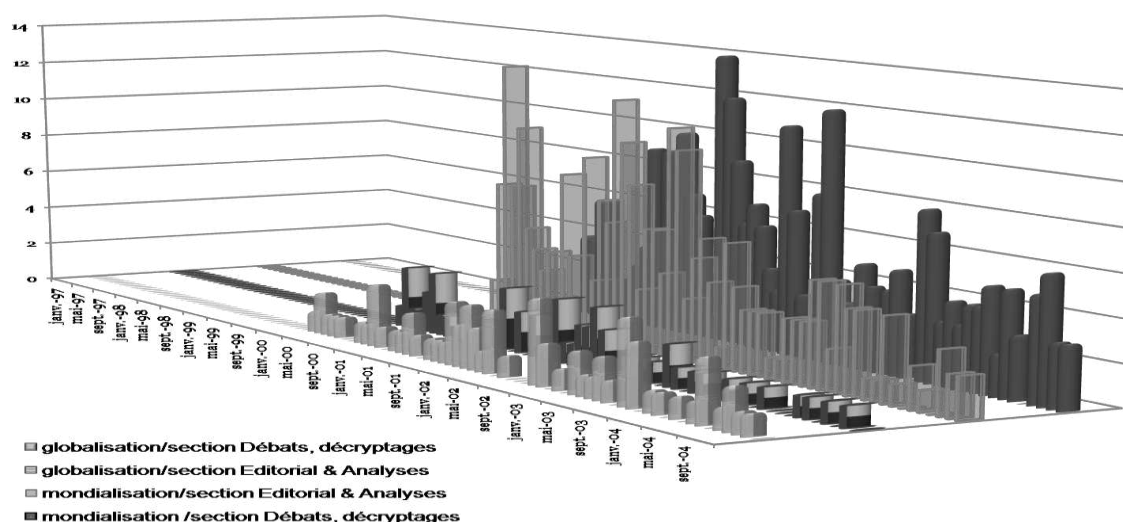


FIGURE 2 : nombre d'occurrences des termes *mondialisation* et *globalisation* dans les rubriques Débats, Décryptages et Editorial et analyses du journal *Le Monde* de 1997 à 2004

1.3 Réseaux de co-occurrence

Les termes co-occurents associés à *mondialisation* et *globalisation* sont les suivants (dans l'ordre hiérarchique) :

Globalisation :

1. économie, 2. monde, 3. pays, 4. économique, 5. politique, 6. mondialisation, 7. marchés, 8. mondiale, 9. face, 10. marché, 11. financière, 12. Europe, ..., 19. France...

Mondialisation :

1. contre, 2. pays, 3. Europe, 4. monde, 5. économie, 6. politique, 7. France, 8. économique, 9. face, 10. libérale, 11. effets, ..., 18. mondiale, ..., 23. marché...

Ces réseaux sont quasiment identiques et confirment un emploi synonymique très fort. Les seuls termes distinctifs sont *financière* pour *globalisation* et *contre* pour *mondialisation*. Ces listes reflètent l'emploi sur l'ensemble du corpus. Nous avons néanmoins mis en évidence trois phases fréquentielles pour *mondialisation* et allons donc nous intéresser aux variations de réseaux pour ce dernier.

1.4 Réseaux de co-occurrence découpés en phases

Les phases de fréquence mises en valeur plus haut sont un reflet de l'emploi mais également un reflet des événements qui mobilisent le terme *mondialisation*. Le redécoupage en phases (fig.3) donne le tableau suivant :

Janv. 97 – juil. 1999	Août 99 – juin 2002	Juil. 02 – déc. 04
économie	contre	Europe
pays	pays	contre
Europe	monde	Pays
France	Europe	monde
économique	économie	politique
face	économique	France
monde	politique	autre
marchés	libérale	libérale
contre	autre	gauche
politique	France	libérale

FIGURE 3 : tableau des 10 premiers mots pleins co-occurents de *mondialisation* dans chaque phase temporelle établie en fonction du comportement fréquentiel observé préalablement.

On observe que les termes mobilisés sont les mêmes lors des trois phases ce qui indique une stabilité sémantique, mais que néanmoins certains termes changent de position dans la hiérarchie d'une phase à l'autre.

Cette hiérarchie est observable grâce au rang de co-occurrence que l'on analyse manuellement ici : Le terme qui arrive en première position en première phase – *économie* – passe en 5^{ème} position en deuxième phase, puis en 66^{ème} position en troisième phase. On peut en déduire que l'on parle de moins en moins d'économie lorsque l'on parle de mondialisation ou inversement qu'il n'est plus nécessaire de mentionner l'économie car elle est peut être devenue partie intrinsèque de la définition de la mondialisation. Ainsi l'interprétation du changement de place hiérarchique d'un mot co-occurent est sujette à caution car elle peut avoir un sens et son contraire. Dans ce cas de figure, le comportement de *politique* donne un élément de réponse, car ce terme présente un comportement inverse : il passe de 10^{ème} position en première phase à 7^{ème} position en seconde phase à 5^{ème} position dans la dernière. On parle donc de plus en politique et de moins en moins d'économie dans la vicinity linguistique directe de *mondialisation*.

Le terme *contre* subit également une variation de position, passant de la 9^{ème} à la 1^{ère} position puis se stabilise en 2^{nde} position. Cette variation confère à ce mot le statut de « terme pivot » car les sens vont s'articuler autour de lui. Il prend une importance contextuelle capitale en phase 2 (fig.3).

Néanmoins l'observation manuelle des rangs est limitée pour des questions d'échelle et de quantité de données, c'est pourquoi nous avons créé un outil permettant de les repérer automatiquement.

1.5 Rangs de co-occurrence

Plus le rang d'un terme co-occurent se rapproche de 1, plus il a une place primordiale dans l'environnement contextuel et connotationnel du terme étudié. Les rangs pris en compte dans cette étude comptabilisent les mots-fonctions qui sont très fréquents. Cette méthode permet de faire resurgir le comportement de termes qui apparaissent loin dans la hiérarchie des listes de co-occurents et ne peuvent donc être traités manuellement ainsi qu'une approche détaillée du comportement des rangs mois par mois (fig.4).

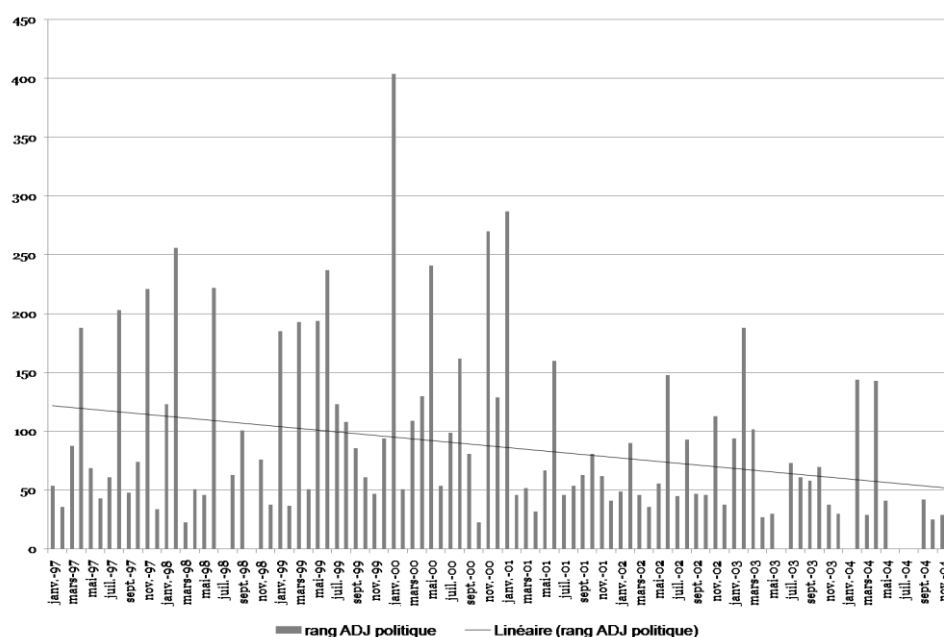


FIGURE 4 : Rangs de co-occurrence de *politique* avec *mondialisation*, et régression linéaire sur les rangs.

1.6 Croisement des données réseaux et rangs

Pour un mot donné, on peut combiner les informations de sa co-occurrence et les informations relatives à son rang pour avoir une vue d'ensemble, comme l'illustre la (fig. 5) pour le mot *politique*.

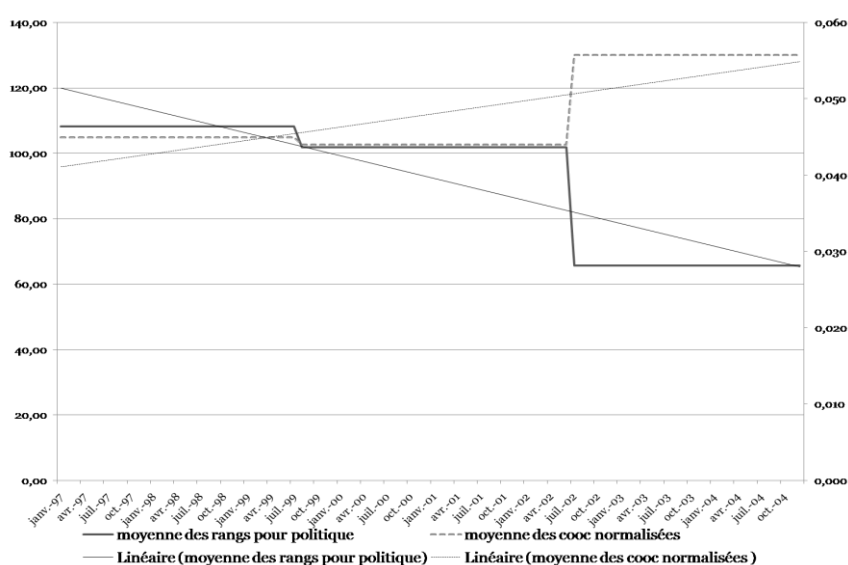


FIGURE 5 : Moyennes par phases des rangs et des co-occurrences pour *politique* et régressions linéaires sur ces moyennes.

On observe sur cette figure qu'en troisième phase il y a plus de termes co-occurents et que les rangs baissent (donc prennent plus d'importance sémantique) simultanément. Le terme gagne donc doublement en impact connotationnel.

1.7 Termes pivots

À travers l'analyse des réseaux de co-occurrence et des rangs, nous avons mis en évidence un certain nombre de termes pivots parmi les co-occurents de *mondialisation*. *Contre* voit sa fréquence de co-occurrence augmenter sur les 3 phases et voit son rang prendre une place capitale en phase 2. *Politique* voit sa fréquence de co-occurrence augmenter sur les 3 phases et son rang décroître alors qu'*économique* voit sa co-occurrence décroître catégoriquement.

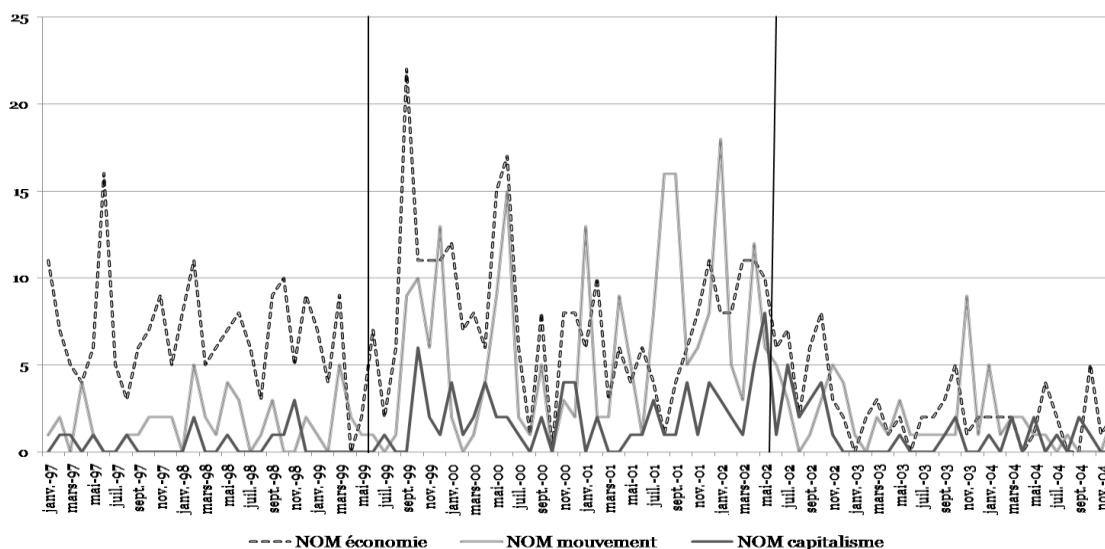


FIGURE 6 : Fréquences de co-occurrence des mots *économie*, *mouvement* et *capitalisme* avec *mondialisation*.

D'autres termes ont des comportements encore plus extrêmes comme *défi* qui n'apparaît plus du tout à partir de janvier 2003. Sur la (fig. 6) on voit que les co-occurrences de *mouvement*, *économie*, et *capitalisme* sont élevées en phase 2 par contraste avec les phases 1 et 3.

C'est autour de ces termes-pivots que s'articule la néologie sémantique de *mondialisation*, mais le mouvement de ces termes renvoie avant tout à une série d'événements dont traite le corpus.

1.8 Evènements

La phase 2, d'août 1999 à juin 2002, lors de laquelle *mondialisation* voit sa fréquence augmenter à plus du double, et pendant laquelle *contre* prend la première position parmi ses co-occurents, est une période marquée par des événements nationaux et internationaux : manifestations à Seattle contre le sommet de l'OMC en novembre-décembre 1999, actions militantes de José Bové en août 1999 suivie de son procès, premier Forum Social mondial de Porto Alegre en janvier 2001, événements du 11 septembre 2001, G8 de Gênes et manifestations en juillet 2001, réunion de l'OMC à Doha et manifestations en novembre 2001 ainsi que second Forum Social mondial de Porto Alegre en février 2002... Tous ces événements marquent l'histoire d'un courant de pensée qui devient un *mouvement politique* se positionnant *contre* une certaine forme de *capitalisme* : l'*antimondialisme* qui se renomme ensuite *altermondialisme*.

1.9 Productivité néologique

La mondialisation est perçue comme un défi de société. C'est un phénomène *contre* lequel certains se mobilisent, ce sont les défenseurs de l'*antimondialisation*, terme qui entre dans le *Petit Robert* en 1997 et qui donne naissance à *antimondialiste*, *altermondialisme*, *altermondialisation* et *altermondialiste* selon le schéma suivant :

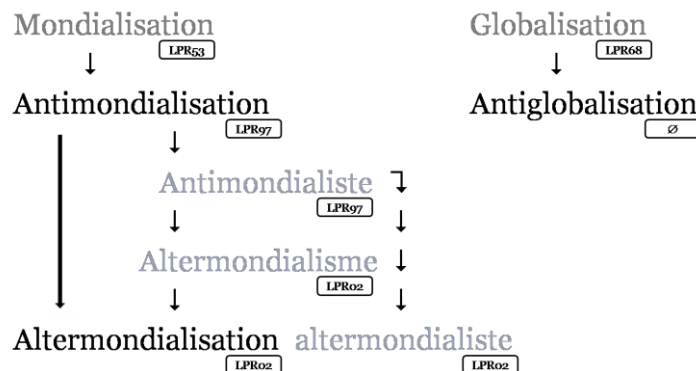


FIGURE 7 : Chemins de créations de néologismes formels à partir de *mondialisation* et *globalisation*, avec leur date d'entrée dans Le *Petit Robert* en 1997 ou 2002.

De la même manière *globalisation* donne naissance à *antiglobalisation*, terme que nous trouvons dans le corpus mais qui n'entre pas dans le dictionnaire. À l'inverse le dictionnaire intègre cinq néologismes dérivationnels à partir de *mondialisation*, deux construits sur le préfixe *anti-* en 1997 et trois sur *alter-* en 2002. On peut se demander s'il y a une différence de sens entre ces constructions. Pour cela nous effectuons un retour à leur définition :

- Antimondialisation. n.f et adj.inv.- 1997. De **anti-** et **mondialisation**. Mouvement de protestation qui s'oppose à la mondialisation, qui redoute ses conséquences économiques, sociales, écologiques. → aussi altermondialisation, altermondialisme. – adj.inv. *Les militants antimondialisation*. → **antimondialiste**.
- Altermondialisation. n.f. – 2002. De **antimondialisation**, d'après **altermondialisme**. Courant d'opinion qui propose un type de développement économique opposé au modèle libéral (mondialisation) plus soucieux du développement de l'homme et de la protection de l'environnement. → **altermondialisme**.

Il s'agit là de définitions très similaires mais un glissement s'opère entre un « mouvement de protestation qui s'oppose » à la mondialisation et un « courant d'opinion qui propose » un modèle alternatif. On lit en filigrane que le mouvement antimondialiste a dû se redéfinir au-delà d'un simple mouvement d'opposition en tant que mouvement politique qui propose une vision.

Dans notre corpus ces néologismes formels apparaissent au sein de la phase 2 décrite plus haut pour *mondialisation*, qui agit comme une sorte de « catalyseur ». Étonnamment les termes les plus usités aujourd'hui en *alter-* présentent assez peu d'occurrences alors que les

formes en *anti-* en présentent plus. *Antiglobalisation* qui n'entre pas dans le dictionnaire n'en entre pas moins dans le corpus (fig.8).

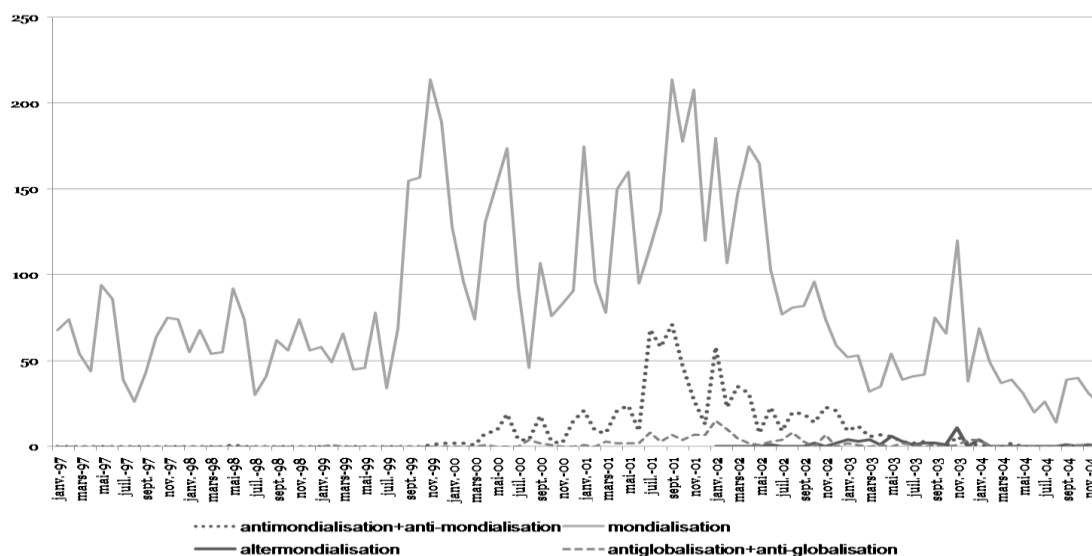


FIGURE 8 : Occurrences de *mondialisation*, d'*anti-mondialisation* et *antimondialisation* additionnées, *anti-globalisation* et *antiglobalisation* additionnées ainsi qu'*altermondialisation*.

On peut supposer que la création de ces néologismes préfixés en *anti-* et *alter-* remet en jeu la définition du terme source. En effet ces termes prennent pour base le mot *mondialisation* en lui attribuant une connotation bien définie : c'est une connotation économique et politique qui renvoie à un modèle de société qui se dégage bien plus qu'une définition vaste basée sur le sens général de mondialisation entendue comme « le fait de devenir mondial » ou reposant sur l'idée de « village global » telle que l'on peut la trouver chez (McLuhan, 1964, 2003).

1.10 Mesure de variabilité de la densité et de la cohésion

Nous avons détecté des variations de fréquence et des variations structurales de co-occurrence qui nous indiquent qu'un changement a lieu. Le modèle offre un autre angle de vue sur ces changements grâce au système de cliques. La (fig. 9) montre une stabilité du rapport cliques-termes qui reste compris entre 0,25 et 1 pour *globalisation* et une variabilité conséquente de ce rapport pour *mondialisation* en phase 2 avec une variation entre 0,5 et 2,71.

Le foisonnement de termes associés à *mondialisation* sur cette période indique une augmentation de la densité, soit due à l'étoffement des contextes en présence soit à l'addition de nouveaux contextes. La génération de nombreuses cliques contenant majoritairement de nouveaux termes associés indique qu'il s'agit ici d'un apport de nouveaux contextes. Le rapport élevé constaté à certaines dates révèle des phases de structuration de la cohésion. La densité des liens entre les termes associés augmente. On en déduit que *mondialisation* est employé dans un ou plusieurs nouveaux contextes pendant cette phase.

Par exemple, le nouveau contexte d'emploi émergeant qu'est l'*anti-* et l'*alter-* mondialisation se manifeste au travers de l'introduction d'un nouvel ensemble de termes associés fin 1999. Il nécessite une création de nouveaux pivots entre cliques pour être intégré de façon continue dans la suite des cliques existantes.

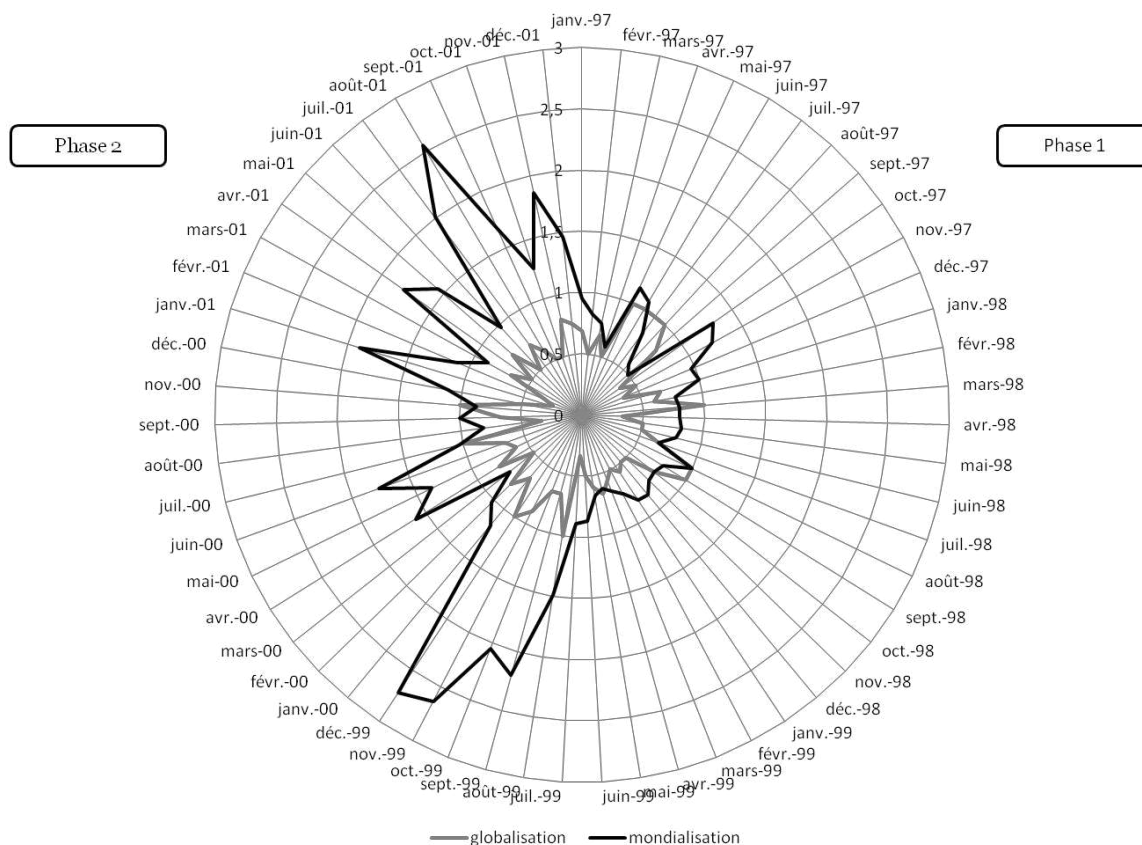


FIGURE 9 : représentation de la variabilité de la densité et de la cohésion exprimée par le rapport entre le nombre de cliques et le nombre de termes pour *mondialisation* et *globalisation*.

2. Cartographie dynamique

2.1 Méthode et enjeux

La visualisation des cartes s'effectue par le biais d'une applet java, sur des sous-corpus divisés en périodes temporelles. Cette applet appelle un programme C auquel on passe en paramètres quatre coefficients de diffusion de l'espace (Ji *et.al*, 2003) servant à calculer le seuil d'intégration du nombre de contexonymes (mots co-occurents) et de fils de contexonymes. Ce programme génère en amont la base des contexonymes ainsi que les coordonnées de chaque clique pour une période de temps donnée.

Les éléments novateurs de notre démarche par rapport au modèle existant reposent sur la possibilité de choisir la fenêtre temporelle souhaitée en fonction de l'objet de la recherche et sur la dynamique créée entre les étapes. On obtient des séries de cartes montrant l'évolution de contextes associés dans le temps. À partir de ces cartes, l'utilisation d'un dispositif dynamique utilisant l'interpolation permet la visualisation du passage entre les étapes. En passant de la relation topologique qu'offre la juxtaposition de deux images statiques au déploiement entre les enveloppes dans un paradigme dynamique nous interrogeons les propriétés structurales des formes générées et leur relation combinatoire. Nous proposons ici une réflexion graphique sur les cartes qui s'appuie sur ces propriétés structurales pour la représentation de l'évolution du sens. En utilisant les propriétés de surface et de contour des

enveloppes, l'interpolation permet de visualiser de manière dynamique les glissements sémantiques au travers de l'évolution graphique. Elle permet également d'interroger le statut de l'espace-temps présent dans le procédé de l'animation. Nous créons ainsi un espace dynamique dont la forme donne un accès qualitatif et quantitatif aux contenus informationnels. L'interface et l'animation facilitent la lisibilité de données complexes et permet d'accéder à une vision plus globale des phénomènes.

Cette réflexion s'inscrit dans un cadre plus général qui englobe la modélisation, la simulation et plus généralement la représentation spatiale de l'information, liées à l'expansion du numérique et à la nature des traitements informatisés de l'information. En effet les questions d'interface et de navigation dans des environnements virtuels sont devenues des préoccupations majeures dont nous souhaitons intégrer les enjeux dans notre recherche.

2.2 Cartographie dynamique pour « mondialisation »

La version animée de cette visualisation est disponible à l'adresse suivante : <http://dico.isc.cnrs.fr/fr/diachro.html>

La cartographie (fig. 10) décompose en 5 temps la représentation de l'évolution du sens du mot *mondialisation* dans un corpus du journal *Le Monde* entre 1997 et 2001. Nous avons délibérément sélectionné quelques mots mis en jeu : *défi*, *menace*, *progrès* et *alternatif* car ils sont représentatifs de l'évolution des contextes d'emploi de *mondialisation*. On peut observer ici le passage d'un sens relativement générique à des sens plus spécifiques, liés entre autres à la création des néologismes formels sur cette période. On voit également l'évolution de l'enveloppe globale du mot et de la densité à l'intérieur de celle-ci. Les clusters thématiques changent également au cours du temps, regroupant parfois des notions qui s'opposent par la suite.

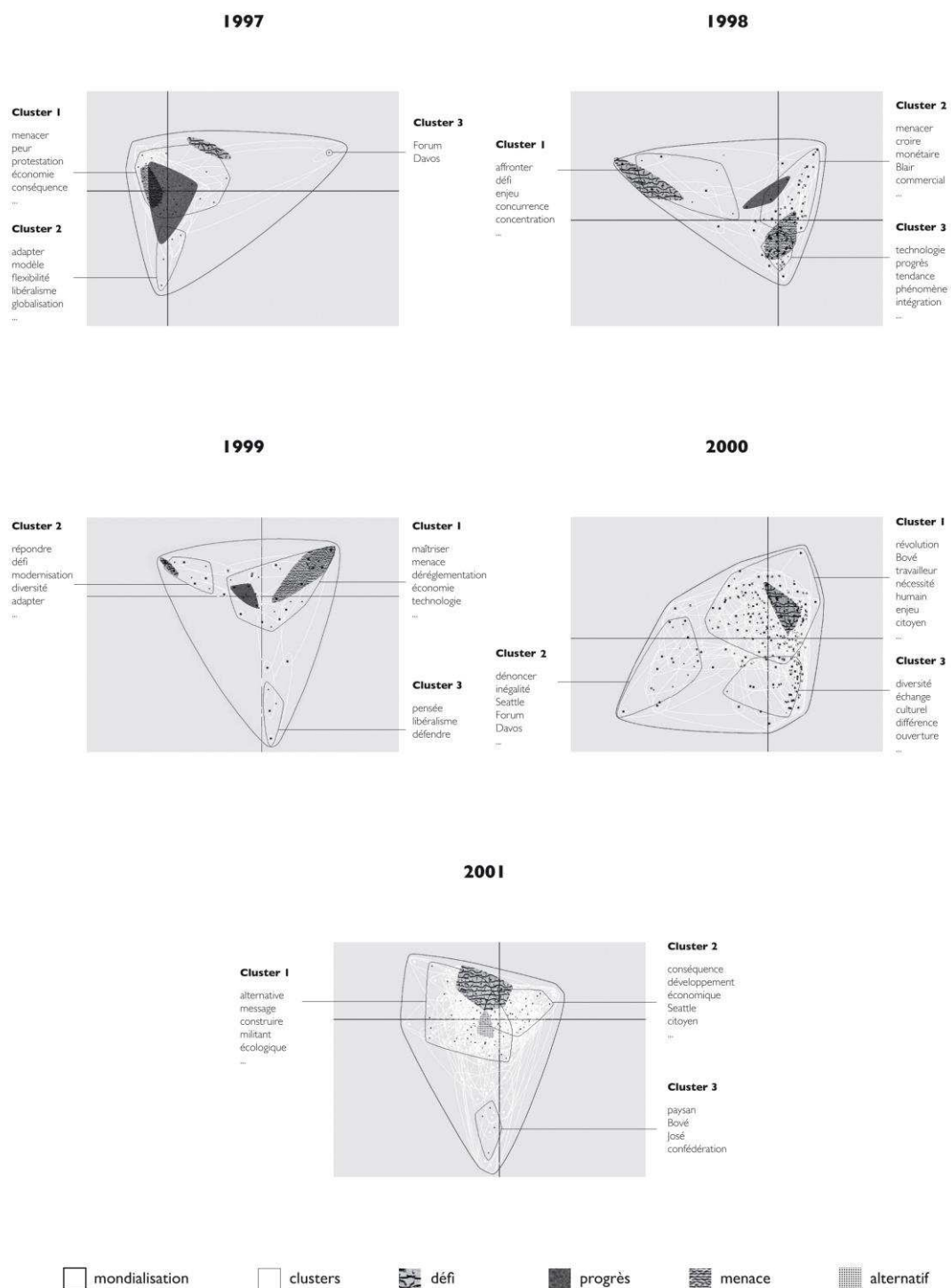


FIGURE 10 : Cartographie dynamique pour *mondialisation* de 1997 à 2001.

En 1997, on observe trois clusters thématiques² :

1. Menacer, peur, protestation économie, conséquence...
2. Adapter, modèle, flexibilité, libéralisme, globalisation...
3. Forum, Davos

Les enveloppes de *progrès*, *menace* et *défi* se trouvent dans le même cluster. Les enveloppes de *progrès* et *menace* sont superposées bien que *progrès* soit à cheval sur deux clusters. L'enveloppe de *progrès* est très centrale et de taille conséquente.

Cette carte montre comment les contextes d'emploi de *mondialisation* s'articulent autour d'une perception de cette dernière comme un *défi*, une *menace* ou un *progrès*. L'idée d'adapter la société à un modèle économique se mélange avec des termes reflétant les différentes prises de position apportant des réponses à cette question. Le troisième cluster qui affiche le Forum économique de Davos est également inclus dans les enveloppes de *planète* et *économie* (en gris). Sur cette carte, le réseau contextuel de *mondialisation* laisse transparaître un état des lieux dans lequel la mondialisation est perçue comme un défi qui peut faire peur ou être le nécessaire déclencheur du progrès. Différents modèles économiques et politiques sont discutés dans le corpus et se positionnent autour de cette réflexion.

Entre 1997 et 1998, les 3 enveloppes de *défi*, *menace* et *progrès* se séparent pour se placer dans trois clusters bien que *menace* et *progrès* soient tous deux à cheval sur deux clusters (2 et 3) et proches du centre, alors que *défi* se positionne dans le cluster majeur (1). En 1998, les trois clusters se définissent ainsi :

1. Affronter, défi, enjeu, concurrence, concentration...
2. Menacer, croire, monétaire, Blair, commercial...
3. Technologie, progrès, tendance, phénomène, intégration...

On obtient ainsi un triangle dans lequel face à *défi* qui connote l'ensemble du cluster principal, *menace* et *progrès* se positionnent comme deux contextes différents, en opposition. *Menace* est associé aux contextes relatifs à la prise de position de Tony Blair alors que *progrès* est associé à la technologie.

Entre 1998 et 1999, la taille de l'enveloppe de *menace* augmente alors que celles de *progrès* et *défi* diminuent beaucoup. *Menace* et *progrès* se placent en opposition de part et d'autre de l'axe vertical bien qu'ils se positionnent tous deux à l'intérieur du cluster principal en grande partie occupé par *menace*.

En 1999, les trois clusters se structurent comme suit :

1. maîtriser, menace, asiatique, déréglementation, économie, technologie...
2. répondre, défi, modernisation, diversité, adapter...
3. pensée, libéralisme, défendre

Le premier cluster mélange les idées de progrès technologique et de menace de l'économie asiatique. Le second laisse transparaître l'idée que le défi est de s'adapter à la modernité. Le troisième s'ancre autour de la défense du libéralisme. Sur cette carte on voit donc que les contextes d'emploi se structurent autour de l'idée d'affronter un défi, posé par le progrès technologique de l'Asie et des États-Unis (en gris), et soutenu par une forme de libéralisme.

Entre 1999 et 2000 les termes *menace* et *progrès* disparaissent de la carte alors que *défi* se place de façon très centrale et voit sa taille augmenter considérablement.

² Les mots listés par cluster ont été sélectionnés en tant qu'échantillons représentatifs de tous les mots contenus dans le cluster, qu'il serait trop long de lister ici.

2000 est la carte du changement, sur laquelle la densité explose alors que l'enveloppe générale change radicalement de forme. C'est également à ce moment là qu'*antimondialisation* entre dans le corpus. On voit 3 clusters imposants composés ainsi :

1. Révolution, Bové, travailleur, nécessité, humain, enjeu, citoyen...
2. Dénoncer, inégalité, Seattle, Forum, Davos...
3. Diversité, échange, culturel, différence, ouverture...

Les deux plus gros clusters regroupent du vocabulaire lié à l'altermondialisme et à des événements organisés par ses partisans ou par les pouvoirs en place contre lesquels ils se mobilisent. Le troisième cluster possède un contenu à forte teneur idéologique, qui véhicule des valeurs comme l'ouverture ou la diversité.

Entre 2000 et 2001 *défi* garde sa taille conséquente et reste central. Nous avons sélectionné le terme *alternatif* qui apparaît en 2001, très central et à cheval sur les 2 clusters majeurs.

En 2001 la carte se structure ainsi :

1. Alternative, message, construire, militant, écologique...
2. Conséquence, développement, économique, Seattle, citoyen...
3. Paysan, Bové, José, confédération

Elle est entièrement articulée autour de l'altermondialisme avec un cluster principal ancré dans le contexte du message altermondialiste, un second cluster regroupant les réflexions politiques et économiques autour de la mondialisation incluant peur, opposition, ou pensée, et un troisième cluster complètement restreint à l'altermondialisme et un de ses acteurs.

Ainsi, entre 1997 et 2001 on peut observer une évolution nette : on passe d'une carte qui comporte à peine quelques occurrences liées à l'antimondialisation en 1997 à une carte en 2001 complètement centrée autour d'une réponse altermondialiste française d'opposition et de proposition d'alternative, qui mobilise des acteurs définis. Les contextes d'emploi de l'*anti-* et l'*alter-mondialisation* teintent très fortement ceux de *mondialisation* jusqu'à redéfinir la carte de contextes. *Mondialisation* se spécialise dans un emploi plus politique, économique et social, se rattachant au libéralisme, alors que *globalisation* ne subit pas de variation majeure de contexte ou de fréquence.

Conclusion et perspectives

Cette étude met en jeu un outillage informatique qui permet d'aborder la sémantique lexicale de façon à la fois très détaillée et sur de vastes ensembles de données. Comme nous l'avons mentionné plus haut la démarche poursuivie ici part de l'exemple et nous souhaiterions partir d'un repérage semi-automatique dans le futur. En effet tous les indices présentés ici constituent l'outillage que nous envisageons de structurer pour repérer le changement sémantique en corpus au lieu d'interroger le modèle sur des termes spécifiques. Ces outils pourront également être mis en œuvre et validés sur d'autres types de corpus (en terme de genre et dans d'autres langues) pour aboutir à une théorie plus générale et adaptable du changement sémantique en TAL. L'étude de cas que nous avons présentée a pour but de confronter ces indices à un exemple concret de repérage et d'analyse du changement sémantique en corpus sur des périodes courtes. Nous avons montré comment la visualisation peut permettre d'appréhender des données sémantiques complexes : ici la modélisation d'un cas de néologie sémantique très fin de l'ordre du changement de connotation. Nous envisageons également une réflexion plus poussée sur l'utilisation des outils graphiques employés dans la visualisation de données linguistiques.

**Armelle Boussidan, Anne-Lyse Renon, Charlotte Franco,
Sylvain Lupone & Sabine Ploux**

L2C2, Institut des Sciences Cognitives-CNRS, Université de Lyon, Bron, France
Armelle.boussidan@isc.cnrs.fr; arenon@isc.cnrs.fr; charlotte.franco@isc.cnrs.fr,
sylvain.lupone@isc.cnrs.fr; sploux@isc.cnrs.fr

Remerciements

Ce projet est soutenu par la Région Rhône-Alpes (Projet Cible 2009 « Modélisation sémantique dynamique en TAL et pour le web »). Nous remercions également Anne Cheylus, ingénieur au L2C2, auteur de certains outils.

Références bibliographiques

- Benzécri, J. P. (1980). *L'analyse des données : L'analyse des correspondances*. Bordas, Paris.
- Burgess, C., & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes*, 12, p. 177-210.
- Ji, H., Ploux, S. & Wehrli, E. 2003. Lexical knowledge representation with contextonyms. *Proceedings of the 9th Machine Translation Summit*, p. 194-201.
- Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.
- McLuhan, Marshall. (1964, 2003) *Understanding Media*. Gingko Press.
- Ploux, S. 1997. Modélisation et traitement informatique de la synonymie. *Linguisticae Investigationes*, 21(1): p. 1-28.
- Ploux, S., Boussidan, A. & Ji, H. (2010). The Semantic Atlas: an Interactive Model of Lexical Representation. *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, May 19-21, Valletta, Malta.
- Sahlgren, M. (2006). The Word Space Model. Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces. PhD Thesis. Stockholm University, Sweden.

Sites internet

<http://dico.isc.cnrs.fr/fr/index.html>

<http://dico.isc.cnrs.fr/fr/diachro.html>

<http://www.techno-science.net/?onglet=glossaire&definition=5465>