



Néologismes de sens : contribution à leur caractérisation dans un corpus autour du thème de la crise financière.

Coralie Reutenauer, Evelyne Jacquey, Sandrine Ollinger

► To cite this version:

Coralie Reutenauer, Evelyne Jacquey, Sandrine Ollinger. Néologismes de sens : contribution à leur caractérisation dans un corpus autour du thème de la crise financière.. II Congrès International de Néologie des Langues Romanes (Cineo2011), Dec 2011, São Paulo, Brésil. halshs-00872138

HAL Id: halshs-00872138

<https://shs.hal.science/halshs-00872138>

Submitted on 11 Oct 2013

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.

Néologismes de sens : contribution à leur caractérisation dans un corpus autour du thème de la crise financière

Coralie Reutenauer, Evelyne Jacquey, Sandrine Ollinger

UMR ATILF-CNRS-Lorraine Université

Coralie.Retenuer@atilf.fr, <mailto:Evelyne.Jacquey@atilf.fr>,
Sandrine.Ollinger@atilf.fr

RÉSUMÉ

De nombreux domaines comme la recherche d'information, la linguistique ou le traitement automatique des langues recourent à des ressources lexicales. Dans le contexte de la nécessaire mise à jour de telles ressources, cet article examine une méthodologie de détection et de qualification des néologismes de sens dans la lignée des travaux de (Bauer ; Renouf, 2000) et (Picton, 2009) notamment. Fondée sur l'exploitation d'un corpus d'étude constitué autour du thème de la crise financière de fin 2008, l'étude présentée s'intéresse plus précisément aux noms *bouclier*, *tempête*, *tsunami* et aux adjectifs *pourri*, *toxique* et *financier*. La méthodologie proposée repose sur une exploration des réseaux de cooccurents d'ordre 2 obtenus à partir des séquences nom-adjectif contenant les cibles lexicales. Elle hiérarchise l'information apportée par les cooccurents et l'articule avec le ou les sens délivré(s) par une ressource lexicale sous forme de traits sémantiques. Les cooccurents sont sélectionnés à l'aide du logiciel Termostat (Drouin ; Paquier ; Ménard, 2006), sur un critère de spécificité dans le corpus financier par rapport à un corpus de référence. Les traits sémantiques sont puisés automatiquement dans un dictionnaire de langue, le Trésor de la Langue Française informatisé. Les cooccurents et les traits sémantiques sont comparés à partir de classes manuelles interprétées comme des facettes sémantiques. L'approche, partiellement automatisée et complétée par une analyse manuelle des résultats, cherche à faire émerger des éléments associés au nouveau sens des lexies retenues ainsi que les liens avec les sens préexistants.

1. Introduction

La mise à jour de ressources lexicales repose sur la capacité à cerner de

nouveaux usages linguistiques. L'assimilation de ces usages peut notamment se concrétiser à travers l'acquisition de néologismes en cours d'implantation. L'élaboration d'outils d'acquisition est susceptible de répondre à deux objectifs : d'une part, détecter l'existence d'un nouveau sens, d'autre part qualifier le nouveau sens. Le développement d'outils connaît des stades d'avancement variables selon le type de néologisme et l'objectif : la détection est relativement avancée, surtout pour les néologismes de forme comme en témoigne l'existence de plateformes telles que *Telanaute* (Issac ; Ouenniche, 2010), *Wortwarte* (Lemnitzer ; Ule, 2005) ou encore *Termostat*, initialement conçue pour la détection de termes, qui a par la suite été utilisée pour la détection de différents types de néologismes (Drouin ; Paquier ; Ménard, 2006).

Au niveau de la qualification du nouveau sens, certains traitements ont été amorcés, en particulier pour les néologismes de forme, comme l'analyseur *DeriF* qui est capable d'affecter un sens aux néologismes constructionnels par décomposition morphologique (Namer, 2009). Mais globalement, la qualification des néologismes de sens reste à un stade exploratoire. Leur analyse exige de s'appuyer sur les informations apportées par le contexte, faiblement exploitées dans les procédures automatiques actuelles. Les indices contextuels à exploiter ont été mis en évidence à travers diverses études, en particulier (Bauer ; Renouf, 2000) (Picton, 2009). Ils se situent sur deux axes d'observation transversaux, syntagmatique et paradigmatique. Sur l'axe syntagmatique, les cooccurents sont des indices qui présentent un fort potentiel qualitatif. Sur l'axe paradigmatique, on compte les unités participant du foisonnement néologique (Dury, 2008 citant Guilbert 1965), tels que *cellulaire*, *mobile*, *portable*. Plus généralement, le paradigme d'unités partageant les mêmes contextes d'emploi que la cible néologique est susceptible d'avoir un apport qualitatif important. Ce paradigme peut se concevoir comme un sous-ensemble de cooccurents de cooccurents, c'est-à-dire de cooccurents d'ordre 2. Notre étude se focalisera sur cet indice pour représenter l'information contextuelle.

Par ailleurs, la qualification de néologismes sémantiques suppose de se positionner par rapport aux sens préexistants. Bien que l'utilisation d'un dictionnaire ou même d'un ensemble de dictionnaires reste sujette à critique (Sablayrolles, 2008), (Valette, 2009) propose une méthodologie pour extraire d'un dictionnaire un ensemble de traits sémantiques associé à toute unité lexicale, afin d'obtenir des représentations dans l'esprit des sémèmes de (Rastier, 1987). Cette proposition a été rendue opérante à travers la mise en place d'un outil d'extraction de traits sémantiques (Semy, Grzesitchak ; Jacquey ; Baider, 2008).

L'objectif du présent travail est d'étudier l'apport de l'axe paradigmatique à la qualification du nouveau sens, autrement dit les cooccurents de cooccurents, pour savoir s'il y a nouveauté, ce qu'est le nouveau sens et comment il s'articule au(x) sens préexistant(s). Nous proposons une méthodologie pour extraire un profil sémantique de différentes cibles lexicales par interaction entre un dictionnaire et un corpus : les emplois discursifs de différentes cibles lexicales sont observés à travers le paradigme de cooccurents d'ordre 2, puis articulés à une représentation des sens existants. Nous présentons en section 2 les supports de l'étude, à savoir le corpus et les cibles lexicales ; en section 3, la méthodologie ; en section 4, les résultats et leur analyse.

2. Corpus et cibles lexicales

2.1 Un corpus d'étude sur le thème de la crise financière

Le corpus d'analyse est constitué de 1587 articles de presse, correspondant à près d'un million d'occurrences de mots et environ 35 000 lexèmes. Les articles sont issus de deux quotidiens français, *L'Humanité* et *Le Figaro*. Ils s'étendent sur la période de septembre 2008 à février 2009 et portent sur le thème de la crise financière.

Ce corpus a été choisi car il a déjà fait l'objet d'études antérieures (Reutenauer ; Valette ; Jacquey, 2009) qui ont témoigné de l'existence de phénomènes néologiques, par interaction entre la thématique de la finance et des métaphores biologiques et météorologiques de la crise comme *maladie* ou *catastrophe naturelle*.

2.2 Cibles lexicales

Trois types de cibles ont servi de support aux expériences :

- **trois cibles néologiques** : *toxique*, employé pour qualifier des produits et instruments financiers en contexte de crise (*créances toxiques, actifs toxiques, etc.*) ; *tsunami*, pour qualifier des événements violents et destructeurs, en l'occurrence la crise financière dans le corpus d'étude ; *bouclier*, comme dispositif fiscal qui impose un plafonnement des sommes dues légalement (*bouclier fiscal, bouclier social*).
- **deux cibles à statut intermédiaire** : sans relever à proprement parler de la néologie sémantique, elles s'y rattachent par certains aspects : *tempête*, qui connaît des emplois métaphoriques en contexte de crise ; *pourri*, dont les emplois sont proches de ceux de *toxique* et dont le statut est ambigu, car il peut s'agir soit d'un emploi métaphorique qui n'a pas de caractère de nouveauté ou de nouveaux emplois propres au domaine financier.

- **une cible témoin** : il s'agit d'une cible lexicale non néologique, à savoir l'adjectif *financier*.

La présélection des cibles répertoriées comme néologiques a été guidée par deux critères, d'une part un sentiment de nouveauté, d'autre part l'existence d'indices conscients de néologie tels que les guillemets (Bauer ; Renouf, 2000).

3. Méthodologie

L'objectif de la méthodologie définie et mise en oeuvre dans les travaux présentés ici est de parvenir à qualifier un sens nouveau, pour une cible donnée, en comparant une représentation du sens en discours et une représentation du ou des sens existant(s), c'est-à-dire issue d'un dictionnaire de référence, en l'occurrence le TLFi. La représentation du sens en discours exploite les cooccurents d'ordre 2 (cooccurents de cooccurents) de chaque cible lexicale. La représentation du ou des sens existant(s) s'appuie sur une extraction du contenu sémantique associé à la cible. Ce contenu sémantique est extrait sous forme de traits sémantiques. Ces deux ensembles descriptifs (cooccurents d'ordre 2 pour le sens en discours et traits sémantiques pour le sens existant) ne sont pas immédiatement comparables. Pour pouvoir les comparer, ces deux ensembles ont été ramenés à des facettes sémantiques jouant le rôle de dénominateurs communs, par le biais d'un classement manuel inductif. Autrement dit, ce classement a été réalisé à partir des éléments constitutifs des ensembles que l'on a progressivement cherché à généraliser et non à partir d'une liste d'étiquettes de classes préalablement définies que l'on aurait ensuite confrontée aux éléments constitutifs (cooccurents et traits sémantiques) des ensembles.

3.1. Représentation du ou des sens existant(s)

Le sens codé associé à chaque cible lexicale est obtenu à partir d'une ressource de référence, le *Trésor de la Langue Française informatisé* (Dendien ; Pierrel, 2003). Il est généré à l'aide de la plateforme Semy (Grzesitchak ; Jacquey ; Baider, 2008). À chaque cible lexicale est associée un ensemble de définitions, dont sont extraits les noms, verbes, adjectifs, adverbes et étiquettes de domaines. Les éléments ainsi extraits sont considérés comme les traits sémantiques de la cible lexicale et constituent la représentation de son contenu sémantique tel qu'il est décrit dans le dictionnaire de référence utilisé.

Adjectifs	/toxique/, /minéral/, /animal/, /végétal/, /physique/, /médisant/, /dépréciateur/
Adverbes	/négativement/, /ouvertement/
Noms	/intoxication/, /destruction/, /organisme/, /individu/, /psychisme/, /poison/, /toxine/, /propos/, /empoisonnement/, /quantité/, /présence/, /action/, /organe/, /tissu/, /origine/, /cause/, /pollution/
Verbes	/provoquer/, /vivre/, /agir/, /être/, /contenir/, /distiller/, /devenir/, /devoir/, /produire/

Tableau 1. Traits sémantiques de la cible lexicale *toxique*

3.2. Sens en discours

Le sens en discours associé à chaque cible lexicale est obtenu à partir du corpus d'analyse. À chaque cible lexicale est associé un paradigme de cooccurents d'ordre 2 obtenu automatiquement à partir des sorties de la plateforme Termostat (Drouin, 2003). Termostat extrait les unités lexicales simples ou complexes spécifiques à un corpus d'analyse relativement à un corpus de référence intégré¹. Le calcul de spécificités utilisé par Termostat suit la proposition de (Lebart ; Salem, 1994)².

Notre corpus d'analyse, décrit en 2.1, est génériquement analogue au corpus de référence, mais s'en distingue thématiquement et chronologiquement. L'écart chronologique est pertinent pour étudier la néologie, de façon analogue à (Drouin ; Paquier ; Ménard, 2006). L'écart thématique participe de l'émergence de tropes, sources de néologismes, par interaction entre le domaine de la finance et d'autres domaines éloignés.

La soumission du corpus d'analyse à la plateforme Termostat retourne 3 523 unités, dont le seuil minimal de spécificité est de 3 et le seuil de fréquence de 21. Parmi les unités retournées, ce sont les unités complexes de type « Nom Adjectif » qui sont exploitées pour constituer les paradigmes de cooccurents d'ordre 2 des 6 cibles lexicales étudiées.

Dans un premier temps, les structures « Nom Adjectif » contenant les cibles lexicales permettent d'associer des paradigmes de cooccurents nominaux (resp. adjectivaux) d'ordre 1 aux cibles adjectivales (resp. nominales).

¹ Le corpus de référence intégré dans Termostat est constitué d'articles variés du journal *Le Monde* de l'année 2002, constitué d'environ 28 millions d'occurrences.

² Elles correspondent à la valeur-test, indicateur statistique qui se rattache à l'écart-type d'une distribution observée par rapport à la distribution théorique, approximation normale d'une distribution hypergéométrique.

Cible lexicale	Structures NA spécifiques contenant la cible	Cooccurrents d'ordre 1
<i>bouclier</i>	2	<i>fiscal, social</i>
<i>financier</i>	38	<i>système, marché, capitalisme, secteur...</i>
<i>pourri</i>	3	<i>titre, créance, crédit</i>
<i>tempête</i>	2	<i>financier, économique</i>
<i>toxique</i>	4	<i>titre, crédit, emprunt, produit</i>
<i>tsunami</i>	1	<i>financier</i>

Tableau 2. Cooccurrents d'ordre 1 des cibles lexicales

Dans un second temps, les structures complexes de type « Nom Adjectif » contenant les cooccurrents d'ordre 1 permettent d'associer des paradigmes de cooccurrents adjectivaux (resp. nominaux) d'ordre 2 aux cibles adjectivales (resp. nominales). Le même cooccurrent d'ordre 2 d'une cible lexicale donnée peut être associé à plusieurs de ses cooccurrents d'ordre 1, sous différentes formes fléchies.

Cible lexicale	Structures NA contenant les cooccurrents d'ordre 1	Cooccurrents d'ordre 2	
<i>bouclier</i>	39	38	<i>paquet, besoin, protection</i>
<i>financier</i>	140	72	<i>financier, bancaire, mondial</i>
<i>pourri</i>	26	19	<i>financier, douteux, hypothécaire</i>
<i>tempête</i>	125	112	<i>crise, système, relance</i>
<i>toxique</i>	31	24	<i>financier, hypothécaire, titrisé</i>
<i>tsunami</i>	79	79	<i>crise, système, marché</i>

Tableau 3. Cooccurrents d'ordre 2 des cibles lexicales

3.3. Classification sémantique manuelle des cooccurrents et des traits sémantiques

Afin de confronter sens existant(s) et sens en discours, non comparables en l'état, leurs représentations sont transposées à un niveau d'information plus abstrait via un classement sémantique manuel. Ce classement manuel est d'abord établi sur l'ensemble des cooccurrents d'ordre 2 des 6 cibles lexicales étudiées à l'issue duquel les cooccurrents de chaque cible sont qualifiés en fonction des classes sémantiques. Autrement dit, le sens en discours est alors représenté par la liste des classes

sémantiques associées aux cooccurrents de cette cible. De plus, cette représentation en classes peut être raffinée par un indice d'activation de chaque classe, calculé en fonction du nombre de cooccurrents par classe.

Une fois la représentation en classes obtenue pour les cooccurrents (le sens en discours), les traits sémantiques de chaque cible (sens existants) sont répartis dans les classes sémantiques de cooccurrents ou affectés à de nouvelles classes.

La comparaison des ensembles de classes sémantiques représentées pour chaque cible en discours (cooccurrents) et dans le dictionnaire (traits sémantiques) rend alors possible la qualification des néologismes de sens.

Construction des classes sémantiques manuelles de cooccurrents

L'ensemble des cooccurrents est réuni en deux listes catégoriellement définies : les cooccurrents nominaux (cooccurrents d'ordre 2 des cibles lexicales nominales étudiées, à savoir *bouclier*, *tempête*, *tsunami*) et les cooccurrents adjectivaux (cooccurrents d'ordre 2 des cibles lexicales adjectivales étudiées, à savoir *financier*, *pourri*, *toxique*). Deux annotateurs ont réparti en aveugle (sans communication intermédiaire entre eux) les cooccurrents dans des classes construites au fur et à mesure des regroupements opérés, autrement dit la méthode suivie est inductive : sans étiquette préalable, les éléments qui semblent partager un certain noyau sémantique ou qui semblent se rapprocher sémantiquement sont regroupés. Pour cela, les annotateurs ont pu faire appel à leur connaissance de la langue, appuyée ou non sur la consultation de ressources externes, ou bien à leur connaissance du corpus d'analyse. Par la suite, les résultats sont comparés et les désaccords réglés par consensus. Lorsque sa nature sémantique le nécessitait ou que le consensus n'a pas été possible, le déploiement d'un même cooccurrent sur plusieurs classes a été pratiqué.

Au cours de cette première étape de classement, les principales difficultés rencontrées ont été liées au degré de polysémie de certains cooccurrents (*opération* est présent dans 3 regroupements [math-mesure], [argent-finance] et [politique]) et à la prise en compte ou non de la connaissance du corpus. Avec, dans un premier temps, des classes pouvant sembler très différentes entre les deux annotateurs, nous nous sommes aperçues que leur contenu en revanche faisait consensus. Dans la grande majorité des cas, il a donc été assez aisé d'obtenir un consensus sur l'étiquette de classe à choisir. Cette première étape aboutit à une répartition en 9 classes d'adjectifs et 12 classes de noms, qui se recourent partiellement.

Classes	Cooccurrents adjectivaux	Cooccurrents nominaux
activité-secteur	<i>aéronautique, immobilier...</i>	<i>architecture, filiales...</i>
argent-finance	<i>monétaire, spéculatif...</i>	<i>crédit, marché...</i>
espace	<i>américain, international...</i>	<i>édifice, front...</i>
évaluatif	<i>dur, progressiste...</i>	
évaluatif-négatif	<i>douteuses, excessif...</i>	<i>crise, marasme...</i>
évaluatif-positif	<i>sûr, vertueux...</i>	<i>cadeau, stabilité...</i>
instrument		<i>baromètre, machine...</i>
math-mesure		<i>cycle, minima...</i>
phénomène naturel		<i>bourrasque, séisme...</i>
politique	<i>démocratique, capitaliste...</i>	<i>gouvernement, mondialisation...</i>
protection		<i>aide, soutien...</i>
temporalité	<i>frais, inédit...</i>	<i>prévisions, quotidien...</i>
non classé	<i>émergent, issus...</i>	<i>ces, jeu...</i>

Tableau 4. Classes sémantiques obtenues à partir des cooccurrents d'ordre 2

Confrontation avec les sens existants (traits sémantiques)

Contrairement aux cooccurrents, les traits sémantiques ne se répartissent pas en noms et adjectifs. Il s'agit des noms, verbes, adjectifs, adverbes et étiquettes de domaines des définitions d'une cible lexicale donnée. La suppression des traits non nominaux ou adjectivaux a été jugée non-pertinente, car trop coûteuse en termes d'information sémantique. En effet, l'information présente dans les définitions est concentrée, sans redondance et le contexte définitionnel lève les ambiguïtés qui apparaîtraient hors contexte. Ainsi, le trait /distiller/, associé à la cible lexicale *toxique*, ne sera pas entendu comme un processus de décomposition chimique, mais comme un acte de parole, car son contexte est le suivant :

« *Qui distille plus ou moins ouvertement des propos médisants ou dépréciateurs.* »

Lorsque les classes constituées pour les cooccurrents ne prenaient pas en compte certains aspects des sens préexistants (par exemple la facette sémantique apportée par *distiller* ci-dessus), de nouvelles classes ont été ajoutées, 7 au total : [danger-attaque] (/agression/, /intoxication/), [causalité] (/origine/, /cause/), [fabrication-transformation] (/constituer/, /ramollir/), [instabilité] (/agitation/, /trouble/), [langage-propos] (/distiller/, /mot/), [manière] (/façon/, /particulier/) et [vivant] (/insecte/, /pourriture/).

Cet ajout de classes a bien sûr conduit à réviser la classification des cooccurrents. Ainsi, des cooccurrents non classés lors de la première étape

le deviennent, comme l'adjectif *issus* dans la classe [causalité], et davantage de contenu sémantique est pris en compte : *séisme* intègre, en plus de [phénomène naturel], la nouvelle classe [danger-attaque].

Au final, les traits sémantiques (sens existants) et les cooccurents (sens en discours) de chaque cible lexicale sont répartis au travers d'un même ensemble de 19 classes plus une classe [non classé]. Le nombre de traits ou de cooccurents associés à chaque classe pour chaque cible lexicale permet de déterminer un taux d'activation de cette classe pour cette cible. Par ailleurs, le taux d'activation d'une classe peut être raffiné par le fait que chaque cooccurent ou trait étant qualifié par une fréquence au sein de l'ensemble des traits sémantiques d'une cible.

Traits classe	cardinal	Cooc2 classe	cardinal
protection	8	politique	10
évaluatif	6	argent - finance	8
fabrication – transformation	6	math - mesure	7
espace	5	Évaluatif – négatif	6
instruments	4	activité - secteur	5
Math – mesure	4	espace	4
évaluatif - positif	3	Évaluatif – positif	4
vivant	3	instabilité	4
non classé	2	protection	4
activité - secteur	1	temporalité	3
danger - attaque	1	causalité	2
langage - propos	1	Danger – attaque	1
politique	1	instrument	1
temporalité	1	Langage – propos	1
		non classé	1
		phénomènes naturels	1

Tableau 5. Répartition en classe pour la cible lexicale **bouclier**

4. Exploitation/analyses

La distribution des traits et des cooccurents dans les classes est étudiée en deux temps. L'objectif est de décider du caractère néologique de la cible et de qualifier le néologisme sémantique. Deux critères sont utilisés : les facettes sémantiques (classes) partagées par les traits et les cooccurents puis une pondération des facettes sémantiques en fonction du nombre de traits ou cooccurents qui les composent.

4.1. Analyse des classes sur critère de présence/absence

Dans cette première partie de l'analyse, les classes sont étudiées selon un critère de présence/absence des traits (resp. cooccurents) dans les différentes classes : on considère que la présence d'au moins un trait sémantique (resp. cooccurent) dans une classe signifie que cette classe est activée au niveau du sens codé (resp. sens en discours). A partir du critère de présence/absence, la variation sémantique peut être qualifiée en termes

d'activation, d'inhibition et d'enrichissement, selon l'interprétation suivante :

- la facette sémantique évoquée par la classe est **activée** si la classe contient au moins un trait sémantique et au moins un cooccurrent ;
- la facette est **inhibée** si la classe ne contient que des traits sémantiques et aucun cooccurrent ;
- la facette est **nouvelle**, autrement dit il y a enrichissement si la classe ne contient que des cooccurrents et aucun trait sémantique.

Les résultats obtenus sur *toxique* (Tableau 6) font émerger quatre classes d'enrichissement liées au thème du corpus ([argent-finance] et [politique]) ainsi qu'à d'autres notions ([évaluatif-positif] et [fabrication-transformation]). Six classes sont activées, associées aux idées de propagation ([causalité], [espace]) au sens biologique ([vivant]) ou au caractère négatif ([évaluatif-négatif]). Quatre classes sont inhibées, dont deux évoquant des sens plus abstraits ([langage-propos], [math-mesure]).

Pour *financier*, l'enrichissement compte sept classes ([évaluatif-négatif], [évaluatif-positif], [causalité], [temporalité], [instabilité], [fabrication-transformation] et [vivant]), l'activation compte six classes ([activité-secteur], [politique], [argent-finance], [évaluatif], [espace], [manière]) et l'inhibition une classe ([math-mesure]).

Classes	Présence des traits	Présence des cooccurrents	Mécanisme sémantique
argent-finance	oui	non	Enrichissement
politique	oui	non	
évaluatif-positif	oui	non	
fabrication-transformation	oui	non	
activité-secteur	oui	oui	Activation
évaluatif	oui	oui	
espace	oui	oui	
causalité	oui	oui	
évaluatif-négatif	oui	oui	
vivant	oui	oui	
manière	non	oui	Inhibition
math-mesure	non	oui	
langage-propos	non	oui	
non-classés	non	oui	
danger-attaque	non	oui	

Tableau 6. Tri des classes de toxique selon le critère présence-absence

Dans les deux exemples mentionnés, la distribution des classes selon la tripartition enrichissement-activation-inhibition fournit une première

approximation de ce que pourrait être le néologisme sémantique associé aux emplois en corpus. Cependant, les résultats restent relativement grossiers, car les classes associées à chaque mécanisme sont indifférenciées. De plus, la quantité de classes associées à chaque mécanisme inviterait à voir un caractère néologique plus marqué pour *financier* que pour *toxique*, ce qui ne rejoint pas les observations manuelles. Le critère de présence/absence est donc insuffisant pour statuer sur le caractère néologique. Afin de préciser les résultats, les analyses sont renouvelées en fonction d'un second critère dépendant de pondérations.

4.2. Analyse des classes à partir de pondérations

Nous faisons l'hypothèse que toutes les classes n'ont pas la même importance selon qu'elles comptent beaucoup ou peu d'éléments (traits sémantiques ou cooccurents). Le calcul est le suivant : soit M le nombre maximal de traits (resp. cooccurents) par classe, k la partie entière de $M/3$. Les classes sont réparties selon les intervalles $[0 ; k[$ (« faible »), $[k ; 2k[$ (« moyen ») et $[2k ; M]$ (« fort »). Ce critère permet de hiérarchiser les classes en fonction du nombre de traits ou cooccurents présents dans chaque classe pour une cible lexicale donnée. Les configurations possibles sont donc multipliées et la description plus nuancée. Par exemple, les classes qui participent à un enrichissement de façon marquée se distinguent de celles qui y participent faiblement. De même, si les traits sont nombreux dans une classe, mais les cooccurents peu nombreux, l'activation sera faible, proche d'une inhibition.

Classes	Présence des traits	Présence des cooccurents	Mécanisme sémantique
argent-finance	/	Fort (10)	Enrichissement
Politique	/	Moyen (4)	
évaluatif-positif	/	Faible (1)	
fabrication-transformation	/	Faible (1)	
activité-secteur	Faible (1)	Moyen (5)	Activation
évaluatif	Faible (1)	Moyen (3)	
espace	Faible (1)	Faible (2)	
causalité	Faible (3)	Faible (2)	
évaluatif-négatif	Faible (3)	Faible (2)	
vivant	Fort (13)	Faible (2)	
manière	Faible (1)	/	
math-mesure	Faible (1)	/	Inhibition
langage-propos	Moyen (4)	/	
non-classés	Moyen (5)	/	
danger-attaque	Moyen (6)	/	

Tableau 7. Tri des classes de toxique en fonction de pondérations

Les résultats sont observés sous deux angles :

- avec une focalisation sur les classes caractérisées par une présence forte ou moyenne de traits, pour savoir ce qu'elles deviennent au niveau des cooccurents. Si les classes fortes ou moyennes au niveau des traits comptent peu ou pas de cooccurents, une configuration nouvelle se dégage, car le noyau sémantique est inhibé, ou du moins éclipsé par des facettes sémantiques mineures ou nouvelles.
- avec une focalisation sur les classes caractérisées par une présence forte ou moyenne de cooccurents, pour savoir dans quelle mesure elles s'exprimaient au niveau du sens codé. Si les classes dominantes au niveau des cooccurents sont des classes caractérisées par une présence moyenne ou forte de traits, on considère qu'il n'y a pas de nouveau sens. S'il s'agit essentiellement de classes d'enrichissement, on considère qu'il y a nouveauté.

Ainsi, pour *toxique*, les classes les plus fortes au niveau des traits, à savoir [vivant], [danger-attaque] et [langage-propos], sont soit inhibées, soit faiblement activées au niveau des cooccurents, ce qui témoigne d'une néosémie par inhibition du sens biologique ainsi que du sens langagier de *toxique*. Les classes principales au niveau des cooccurents sont soit des classes absentes au niveau des traits, à savoir [argent-économie] (seule classe de la catégorie « fort ») et [politique], soit des classes faiblement exprimées par les traits, à savoir [activité-secteur] et [évaluatif]. L'enrichissement serait donc principalement dû au caractère économique et le lien avec le sens codé se ferait notamment à travers l'idée d'évaluation.

Pour *financier*, les résultats sont assez nettement différents. Les classes avec présence forte ou moyenne de traits, [argent-finance] et [évaluatif], se retrouvent en tête de liste des classes exprimées par les cooccurents. Parmi les 6 classes de la catégorie « fort » des cooccurents, trois proviennent de la catégorie « faible » des traits, à savoir [espace], [politique] et [activité-secteur], et une seule est nouvelle, [évaluatif-positif]. Les classes [temporalité] et [évaluatif-négatif] de la catégorie « moyen » sont elles aussi nouvelles. Cette configuration témoigne de l'expression de sens existants. Toutefois, il y a émergence de connotations nouvelles, notamment les connotations positives et négatives. De plus, [espace] émerge en tête de liste, ce qui évoque l'idée de mondialisation de la finance ou encore de finance internationale.

Sans entrer dans le détail de l'analyse pour les autres exemples, nous synthétisons les résultats dans le tableau 8.

Cible	Principaux résultats de l'analyse	Hypothèse sur le sens en corpus
<i>financier</i>	Les facettes dominantes sont essentiellement des facettes importantes du sens codé.	Pas de néologie
<i>toxique</i>	Les facettes importantes du sens codé sont inhibées. Les facettes principalement exprimées par les cooccurents sont nouvelles ou mineures dans le sens codé.	Néologie sémantique
<i>tsunami</i>	Les nouvelles facettes sont nombreuses. Parmi les facettes principales, 4 sont nouvelles et 1 est activée ([espace]). D'autres facettes sont activées, mais de façon moindre, telle que [danger-attaque] et [phénomènes naturels].	Néologie sémantique
<i>pourri</i>	Une nouvelle facette émerge ([argent-finance]). Une facette forte au niveau des traits est activée ([évaluatif-négatif]). Il y a une inhibition massive des autres facettes du sens codé.	Néologie sémantique probable
<i>bouclier</i>	Les cooccurents font ressortir deux nouvelles facettes ([argent-finance], [évaluatif-négatif]). Plusieurs facettes du sens codé ressortent de façon moindre ([protection], [évaluatif-positif]).	Reconfiguration du sens ou nouveaux emplois
<i>tempête</i>	Plusieurs facettes du sens codé sont activées, notamment des facettes mineures. Seules deux nouvelles facettes apparaissent et presque aucune classe n'est inhibée.	Reconfiguration du sens et activation d'un sens différent des sens principaux

Tableau 8. Synthèse des résultats pour les différentes cibles

5. Conclusion

Nous avons proposé une méthodologie pour qualifier un néologisme sémantique par confrontation entre sens codé et sens en discours. La démarche a consisté à construire un paradigme de cooccurents d'ordre 2 puis à utiliser des classes interprétées comme des facettes sémantiques afin d'articuler le sens codé et le sens en discours. L'utilisation de poids établis en fonction du cardinal de chaque classe a permis d'obtenir une représentation nuancée des résultats, qui a contribué à statuer sur le caractère néologique des emplois, servi à identifier les facettes proprement nouvelles et les facettes activées et a mis en évidence des configurations variables selon les cibles lexicales.

La part manuelle dans les traitements est conséquente, notamment pour

la constitution des classes, ce qui soulève le problème de la reproductibilité des expériences. Pour évoluer vers une systématisation ou une automatisation de la classification, le recours à des ressources extérieures, l'exploitation de réseaux lexicaux ou encore la mise en oeuvre de techniques de classification automatique ouvrent des perspectives prometteuses. Par ailleurs, afin d'affiner le modèle sous-jacent, il conviendrait d'intégrer d'autres informations pour affecter des poids aux classes, notamment la structuration des définitions au niveau des traits sémantiques et le degré de saillance des cooccurents, c'est-à-dire leur spécificité en corpus. Enfin, la néologie sémantique est un processus. Pour en avoir une approche complète, il conviendrait de prendre en compte le paramètre temps. Ceci permettrait de distinguer des emplois métaphoriques ponctuels d'emplois néologiques et d'évoluer de représentations statiques des emplois vers des représentations dynamiques de l'évolution de sens.

Références

Livres

GUILBERT, Louis. *La formation du vocabulaire de l'aviation*, Paris, Librairie Larousse, 1965.

LEBART, Ludovic ; SALEM, André. *Statistique Textuelle*, Dunod, 1994, 344 p.

MULLER, Charles. *Principes et méthodes de statistique lexicale*, Larousse, 1977, réimpression Champion-Slatkine, 1992, 211p.

NAMER, Fiammetta. *Morphologie, Lexique et Traitement automatique des langues : l'analyseur DériF*, Lavoisier, 2009, 448p.

PICTON, Aurélie. *Diachronie en langue de spécialité. Définition d'une méthode linguistique outillée pour repérer l'évolution des connaissances en corpus. Un exemple appliqué au domaine spatial*, Doctorat de l'université de Toulouse, 2009.

Sablayrolles, Jean-François. Néologie et dictionnaire(s) comme corpus d'exclusion, *Néologie et terminologie dans les dictionnaires*, Honoré Champion, Paris, 2008, p.19-36.

VALETTE, Mathieu. *Approche textuelle du lexique*, mémoire pour l'Habilitation à Diriger des Recherches, Institut National des Langues et Civilisations Orientales, Paris, 2009.

Revue

BAUER, Laurie ; RENOUF, Antoinette. Contextual clues to word-meaning. *International Journal of Corpus Linguistics*, v. 5-2, p. 231-259, 2000.

DENDIEN, Jacques ; PIERREL, Jean-Marie. Le Trésor de la langue française informatisé. Un exemple d'informatisation d'un dictionnaire de langue de référence. *TAL*, v. 44-2, p.11-37, 2003.

DURY, Pascaline. « Les noms du pétrole : une approche diachronique de la métonymie onomastique », *Lexis, E-Journal in English Lexicology*, <http://SCRECHERCHE.UNIV-LYON3.FR/LEXIS/>, 2008.

Conférences

DROUIN, Patrick. Acquisition de termes simples fondée sur les pivots lexicaux spécialisés. *Conférence TIA-2003*, Strasbourg, 31 mars et 1er avril 2003.

DROUIN, Patrick ; PAQUIN, Annie ; MÉNARD Nathan. Extraction semi-automatique des néologismes dans la terminologie du terrorisme, *JADT 2006*.

GRZESITCHAK, Mick ; JACQUEY, Evelyne ; BAIDER, Fabienne. Annotation sémantique : profilage textuel et lexical. *Actes de la conférence Lexicographie et Informatique : Bilan et Perspectives*. Nancy, France, 2008.

ISSAC, Fabrice ; OUENNICHE, Soundous. Pour une veille néologique à partir du web : l'outil telanaute, in *Actes del i Congrès internacional de neologia de les llengües romàniques (CINEO 2008)*, Barcelone, Espagne, 7-10 mai 2008, p.1165-1173, 2010.

LEMNITZER, Lothar ; ULE, Tylman. *Die Wortwarte - auf der Suche nach den Neuwörtern von morgen*. In : URL : <http://www.wortwarte.de/> (<http://www.wortwarte.de/Projekt/index.html>). Consulté le 11/07/2011.

NAMER, Fiammetta ; BOUILLON, Pierrette ; JACQUEY, Evelyne ; RUIMY, Nilda. Morphology Enhancement of a French SIMPLE Lexicon. *Proceedings of the 5th International Conference on Generative Approaches to the Lexicon*, Italie, 2009.

REUTENAUER, Coralie ; VALETTE, Mathieu ; JACQUEY, Evelyne. De l'annotation sémantique globale d'un texte à l'interprétation locale d'un mot. In *Cognitica, ARCo'09*, Rouen, 9-11 décembre 2009.