HAL
open science

De l'annotation sémique globale à l'interprétation locale : environnement et image sémiques d' " économie réelle " dans un corpus sur la crise financière

Coralie Reutenauer, Mathieu Valette, Evelyne Jacquey

▶ **To cite this version:**

Coralie Reutenauer, Mathieu Valette, Evelyne Jacquey. De l'annotation sémique globale à l'interprétation locale : environnement et image sémiques d' " économie réelle " dans un corpus sur la crise financière. Colloque ARCo'09 Interprétation et problématiques du sens, Dec 2009, Rouen, France. pp.29-39. hal-00441166

HAL Id: hal-00441166

https://hal.science/hal-00441166v1

Submitted on 15 Dec 2009

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.

De l'annotation sémique globale à l'interprétation locale : environnement et image sémiques d' « *économie réelle* » dans un corpus sur la crise financière

Coralie REUTENAUER, ATILF (CNRS – Nancy Université), coralie.reutenauer@atilf.fr ; Mathieu VALETTE, ATILF (CNRS – Nancy Université), mathieu.valette@atilf.fr ; Evelyne JACQUEY, ATILF (CNRS – Nancy Université), evelyne.jacquey@atilf.fr

RESUME. Le travail présenté ici se situe dans le cadre général de l'élaboration d'une méthodologie et d'une instrumentation pour l'analyse des phénomènes interprétatifs. S'inspirant des travaux en sémantique interprétative de F. Rastier, cette recherche entend rendre compte de différents mécanismes cognitifs en jeu dans la compréhension d'un texte (identification du thème, cohésion textuelle). L'étude présentée ici porte sur la structuration du contenu sémantique d'un concept lexicalisé à partir du contexte de production. Au moyen d'une annotation en traits sémantiques et de méthodes statistiques, nous étudions la thématisation et la construction sémantique du concept d'« *économie réelle* » pendant la crise financière de 2008-2009 dans un corpus de presse.

Mots clés : Sémantique interprétative, modélisation de l'interprétation, représentation sémique, annotation sémique

ABSTRACT. The framework of the paper is the construction of methods and tools for analyzing and modeling interpretative phenomena. Based on the works of F. Rastier related to Interpretative Semantics, the research aims to show the different cognitive mechanisms used in text understanding (topic identification, textual cohesion). The study here is about the build of semantic content of a lexicalized concept from speech context. By using semantic feature tagging and statistical methods, we will study the construction of concept *"économie réelle"* (*"real economy"*) during the 2008-2009 financial crisis in a newspaper corpus.

Keywords: Interpretative semantics, interpretation modeling, seme representation, seme tagging,

I INTRODUCTION

Les approches mentalistes dominent globalement le paysage linguistique et académique en ce qui concerne la description et la modélisation de

l'interprétation dans une perspective cognitive. On songe, entre autres, aux travaux effectués dans le sillage des théorisations initiales de Langacker, Lakoff ou Talmy, mais aussi, à celles de Pottier, Culioli et de leurs héritiers[1]. Ces approches linguistiques de la cognition relèvent d'une position épistémologique que l'on peut à bon droit affilier à l'ontologie aristotélicienne : elles visent à restituer à partir d'un matériau donné (les textes, les énoncés) assimilable à l'*ergon*, les mécanismes cognitifs ou linguistiques d'amont, la *dunamis,* qui sont à l'origine de l'énonciation et de l'interprétation. Sans chercher à opposer les vertus d'un paradigme à celles d'un autre, nous nous intéressons ici aux possibilités descriptives offertes par le modèle interprétatif de (Rastier 1991, 2003). L'approche interprétative de cet auteur est concentrée sur les déterminations contextuelles du sens dans les textes et non sur la restitution d'un univers mental qu'il juge hypothétique. Il ne s'agit donc plus de décrire l'élaboration du sens de la cognition à l'énoncé, mais de mettre en place une « praxéologie » de l'activité linguistique recentrée sur la relation des signes entre eux, c'est-à-dire en contexte.

C'est dans cette perspective épistémologique que nous souhaitons ici faire état de l'élaboration d'une méthodologie et d'une instrumentation pour l'analyse et la simulation des phénomènes interprétatifs[2]. L'étude présentée se focalise sur la structuration du contenu sémantique d'un concept lexicalisé à partir du contexte de production tel qu'il est rendu disponible grâce à l'annotation sémantique d'un corpus de textes. Au moyen de méthodes statistiques, nous étudions la thématisation et la construction sémantique du concept d' *« économie réelle* » pendant la crise financière de 2008-2009 dans un corpus de presse.

II PROBLEMATIQUE

Alors que la pratique des ontologies et de la terminologie peut faire l'économie de la contextualisation des concepts, il en va autrement pour la linguistique pour peu qu'elle ait une vision différentielle du signe linguistique. Si on se permet une lecture rapide et forcément très parcellaire, on pourrait dire que dans la perspective de la sémantique interprétative de (Rastier *op.cit.*), interpréter c'est repérer dans le texte des structures sémantiques stables préalablement identifiées (dans d'autres textes) et évaluées à l'aune des nouvelles actualisations effectuées[3]. Les structures sémantiques sont notamment (i) les *fonds sémantiques*, c'est-à-dire des faisceaux de traits sémantiques (ou *sèmes*) récurrents qui donnent sa cohésion au texte et (ii) les formes sémantiques, c'est-à-dire des groupements stabilisés de sèmes hétérogènes. Notre thèse est que certaines de ces formes sémantiques peuvent être considérées comme des

[1] Pour une discussion, on lira Fuchs, éd. 2004, Fuchs 2008.

[2] Cette recherche intéresse notamment l'ingénierie des connaissances, la recherche d'informations et la veille lexicale. On lira pour un approfondissement (Valette 2008), (Valette, Estacio-Moreno *et al.* 2006), Grzesitchak *et al.* 2007).

[3] On prie le lecteur de bien vouloir excuser l'extrême rapidité de cette définition.

concepts non lexicalisés, ou, plus précisément, comme des signifiés sans signifiant stable alloué. Autrement dit, il n'y a pas co-avènement du signifié et du signifiant.

C'est donc la relation entre les formes sémantiques diffuses et le concept en cours de lexicalisation ("*économie réelle*") que nous entendons étudier maintenant. On espère ainsi rendre compte des phénomènes d'allocation de sens à un concept par son contexte d'actualisation.

III SUPPORT ET OUTILS

III.1 Description du corpus

Le corpus utilisé relève du discours journalistique. Il est constitué de 1587 articles de presse, tirés de deux quotidiens nationaux aux lignes éditoriales contrastées, *Le Figaro* et *l'Humanité*. Les articles sélectionnés, entre septembre 2008 et février 2009, ont pour sujet la crise économique et financière. Le corpus se présente sous forme de deux versions parallèles : la version lexicale, d'un million d'occurrences de formes, et une version « sémique » de 23 millions d'occurrences de ce que nous qualifierons prudemment, en l'absence de validation linguistique systématique, de « candidats-sèmes », par analogie aux *candidats-termes* de la terminologie. L'image sémique du corpus est obtenue en substituant à chaque forme lexicale un sémème théorique issu des définitions lexicographiques du *TLFi* (Dendien *et al*. 2003). Sont considérés comme candidats-sèmes les lemmes des noms, verbes, adjectifs, adverbes[4] présents dans ces définitions.

Les informations principales sur la répartition des formes et des candidats-sèmes sont récapitulées ci-dessous.

	Total	*le Figaro*	*l'Humanité*
Nombre d'articles	1 587	928	659
Dates	30/08/2008 - 11/02/2009		
Formes			
Nombre d'occurrences	920 551	533 117	387 434
Nombre de formes	35 147	26 433	23 203
Candidats-sèmes			
Nombre d'occurrences	23 198 346	13 329 284	9 869 062
Nombre de candidats-sèmes	29 661	25 741	24 434

Figure 1. Présentation du corpus

4 Lire (Grzesitchak *et al.* 2007) pour une présentation du programme d'annotation sémique et (Valette 2008) pour une discussion sur la constitution d'une ressource sémique à partir d'un dictionnaire.

Afin d'exploiter le corpus, le fichier de sortie est structuré par journal, par article et par paragraphe. Cette structure permet de bâtir des partitions d'étude suivant deux critères : d'une part, la présence d'un mot-pôle sur le plan lexical, donc de ses candidats-sèmes sur le plan sémique ; d'autre part, la source journalistique (*l'Humanité* ou *le Figaro*). Ce sont les partitions ainsi constituées qui guident les analyses, et non l'ordre linéaire ni la structure syntaxique des segments textuels analysés.

III.2 Méthode mathématique choisie

Les expériences reposent sur un outil de statistique textuelle répandu : le calcul des spécificités, tel que calculé par Lexico3 (Salem *et al*. 2003), provenant du modèle hypergéométrique (Rouchaleau 2008) et utilise des comparaisons entre partie et tout (le tout étant généralement l'ensemble du corpus, la partie, l'ensemble des contextes contenant le mot-pôle).

Soit T la taille du corpus, t la taille de la partie, s le nombre d'occurrences de l'unité considérée dans l'ensemble du corpus et k le nombre d'occurrences de l'unité dans la partie. La probabilité d'observer k occurrences de l'unité dans le sous-corpus est :

$$p(X=k)=\frac{\binom{s}{k}\binom{T-s}{t-k}}{\binom{T}{t}}$$

Si k est supérieur à la valeur modale (resp. inférieur), l'exposant de $p(X \geq k)$ (resp. $p(X \leq k)$) en notation scientifique sera la spécificité en valeur absolue, positive si supérieure à la valeur modale, négative sinon. Pour plus de précisions sur le calcul de spécificités, on se réfèrera à (Lafon, 1984, 54-77).

IV INTERPRETATION LOCALISEE : CARACTERISATION DU SENS D'UN MOT-POLE

Nous avons cherché à faire apparaître l'influence de la ligne éditoriale de chacun des journaux sur l'environnement d'un mot-pôle en optant pour une double caractérisation sémantique du mot-pôle : d'une part à partir d'un faisceau d'unités de sens qui émergent du voisinage ; d'autre part à partir d'une image sémique du mot-pôle préconstituée qui va être modulée par le voisinage.

Le mot-pôle étudié est *« économie réelle »*. Il est présent 176 fois dans 168 paragraphes : 87 issus du *Figaro* et 81 de l'*Humanité*. A la lecture des paragraphes, la crise économique apparaît comme une pathologie contagieuse ou comme une catastrophe naturelle se propageant de la sphère financière, considérée comme virtuelle, à la sphère industrielle, correspondant à l'économie dite

réelle. Ces observations du lecteur ont servi par la suite à guider et à valider les analyses.

IV.1 Etude des voisinages

Nous tentons ici de caractériser le mot-pôle à travers les candidats-sèmes les plus spécifiques[5] du voisinage sur le corpus total (figure 2). La catégorie grammaticale des candidats-sèmes est précisée après le caractère "#". Les candidats-sèmes domaniaux (c'est-à-dire issu des indicateurs de domaines du dictionnaire) sont introduits par la notation "D=".

Sème	Spécificité	Sème	Spécificité	Sème	Spécificité
Budget#subst	21	appréciable#adj	10	effondrement#subst	9
particulier#subst	16	capitaliste#adj	10	enthousiasme#subst	9
ressource#subst	16	collision#subst	10	financier#adj	9
régir#v	15	contagion#subst	10	galaxie#subst	9
Argent#subst	14	décisif#adj	10	intense#adj	9
particulier#adv	14	dysfonctionnement#subst	10	noeud#subst	9
répercussion#subst	14	économie#subst	10	pathologique#adj	9
Théâtre#subst	13	époux#subst	10	phénomène#subst	9
bien#subst	12	profond#subst	10	progressif#adj	9
chômage#subst	12	subit#adj	10	retentissement#subst	9
déterminant#adj	11	boursier#adj	9	rupture#subst	9
diminution#subst	11	craindre#v	9	sous-production#subst	9
néfaste#adj	11	D=dramaturgie	9	surproduction#subst	9
ralentissement#subst	11	développement#subst	9		
roi#subst	11	économique#adj	9		

Figure 2. Candidats-sèmes les plus spécifiques des paragraphes contenant *« économie réelle »*

De cette liste émerge nettement une isotopie économico-financière (/budget/, /argent/, /capitaliste/, /boursier/), ainsi qu'une isotopie cataclysmique

[5] On utilise ici l'adjectif *spécifique* au sens statistique du terme et non en référence à la terminologie de la sémantique interprétative (sème spécifique *vs* sème générique).

(/collision/, /subit/, /effondrement/). La propagation de la crise à l'économie réelle, très marquée à la lecture, apparaît également de façon sensible à travers son impact (/répercussion/), et à travers un caractère pathologique (/contagion/, /pathologique/).

Les tendances repérées, concordantes avec celles de la lecture, se constituent en forme sémantique (groupement stable de sèmes caractéristique du voisinage d' *« économie réelle »*, cf. *supra)*.

Pour compléter cette étude, nous avons choisi de définir des classes sémantiques au degré de généralité variable conformes aux tendances mentionnées. Ces classes ne forment pas une partition : elles se superposent parfois et ne couvrent pas nécessairement toutes les subtilités portées par l'« économie réelle ». Pour chaque classe, la liste des candidats-sèmes de spécificité supérieure à 3 a été parcourue. A titre d'exemple, voici le tableau de la classe //maladie-pathologie//. On constate que la classe est non seulement fournie quantitativement, mais encore est solide qualitativement, avec par exemple la présence des sèmes /maladie/, /épidémie/ ou encore /remédier/.

Candidat-sème	Spécificité	Candidat-sème	Spécificité
contagion#subst	10	injection#subst	3
dysfonctionnement#subst	10	tremblement#subst	3
pathologique#adj	9	bistouri#subst	3
troubler#v	8	défaillir#v	3
trouble#subst	8	soigner#v	3
psychologique#adj	8	tiraillement#subst	3
crise#subst	7	crisper#v	3
mal#subst	7	éternuer#v	3
physiologique#adj	6	psychose#subst	3
infection#subst	5	transpiration#subst	3
épidémie#subst	5	crispation#subst	3
maladie#subst	5		
remédier#v	4		
saignée#subst	4		
contagieux#adj	4		
perturbation#subst	4		

Figure 3 : Candidats-sèmes spécifiques de la classe //maladie-pathologie//

IV.2 Image sémique du mot-pôle : structuration par le cotexte

Dans la seconde approche, le sémème d' "économie réelle", réunion des sémèmes d' "économie" et de "réel" est neutralisé afin d'observer son écho à l'intérieur des paragraphes. Seules les spécificités des candidats de ce sémème apportées par les paragraphes sont observées de façon à déterminer quels sont les candidats qui sont activés et quels sont ceux qui sont inhibés, autrement dit de façon à étudier le complexe sémique (Rastier 1987) d'*« économie réelle »*.

Sans entrer dans le détail du complexe sémique, on observe en particulier que certains candidats-sèmes parmi les plus spécifiques renvoient à une économie centrée sur la production ou la consommation de biens, ou encore à une finance axée sur du matériel et non une finance spéculative ou spécialisée. Près de la moitié des candidats de spécificité supérieure à 3 sont rattachés à cet ensemble : /ressource/, /argent/, /bien/, /économie/, /revenu/, /économie/, /consommation/ et /dépense/[6]. Autrement dit, les sèmes du sémème d' *« économie réelle »* sont propagés dans les passages le contenant.

IV.3 Influence des lignes éditoriales des journaux

A partir des résultats précédents, nous avons cherché à mesurer la sensibilité des informations au contexte éditorial des deux quotidiens de notre corpus. Pour ce faire, l'ensemble des passages centrés sur le mot-pôle a été partitionné en deux sous-ensembles, un correspondant à l'*Humanité*, l'autre au *Figaro*.

Pour l'étude du voisinage, la méthodologie employée est la suivante :

- le corpus de référence est l'ensemble des passages centrés sur le mot-pôle, subdivisé en deux sous-corpus : les passages du *Figaro* et les passages de l'*Humanité ;*

- les spécificités sont calculées sur chaque sous-corpus par rapport au corpus de référence, avec un seuil de fréquence minimale de 3 et un seuil de spécificité de 2 ;

- seuls sont conservés les candidats-sèmes déjà répertoriés dans les classes mentionnées en III.1.

Sur l'ensemble des résultats, que nous ne détaillerons pas ici, la proportion par classe sémantique de candidats-sèmes spécifiques à l'un ou l'autre des deux journaux reste faible, environ 15% sur les candidats présents dans l'ensemble des classes, pour un seuil de spécificité de 2. Cependant, cette faible proportion fait apparaître des contributions variables selon les axes sémantiques et le journal considérés. Pour la classe sémantique de la maladie, par exemple, le *Figaro* semble plus contribuer que l'*Humanité* à faire émerger des sèmes. Pour la classe //argent et économie//, un double apport se dessine, provenant de chacun des deux journaux. En particulier, les sèmes de *l'Humanité* se rapportent au gain, au bénéfice, et proviennent probablement d'une condamnation des profits, tendance idéologique sensible à la lecture.

---

[6] Notons également la présence de faux sèmes, ou *métasèmes* qui sont en fait des éléments de définition non filtrés pour cette expérience. La valeur de leur spécificité n'est pas interprétable, par exemple relatif#adj, ensemble#subst ou élément#subst, sur lesquels nous reviendrons ultérieurement.

Nous avons également étudié l'influence du journal sur le sémème du mot-pôle. Le corpus de référence est cette fois le corpus dans sa totalité. Les spécificités ont été calculées sur les passages centrés sur le mot-pôle et propres à un journal. Seuls les candidats-sèmes du mot-pôle ont été observés.

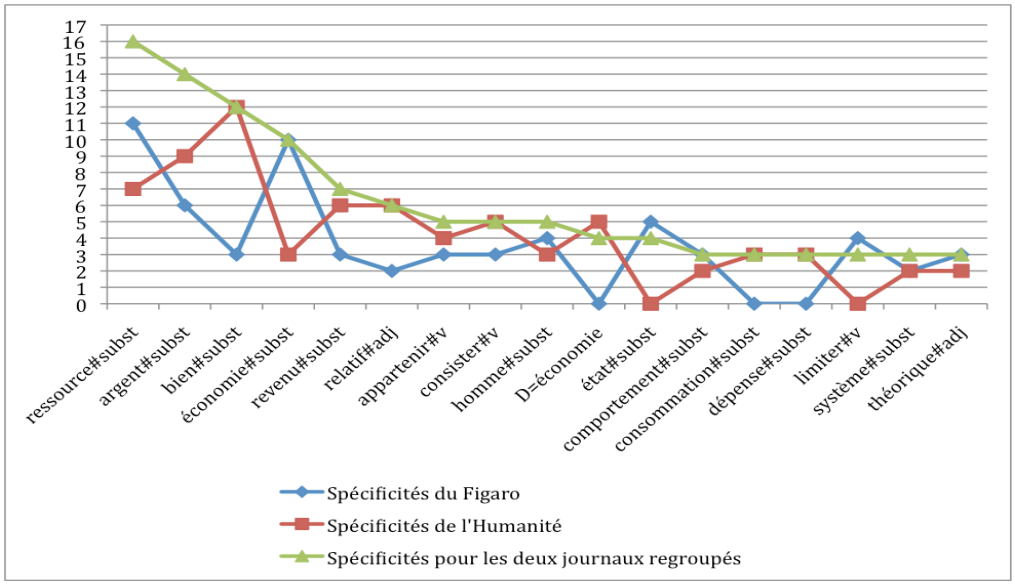

Figure 4 : Spécificités en fonction des candidats-sèmes d' *« économie réelle »*, pour les candidats de spécificité supérieure à 3 sur l'ensemble des passages : étude de l'apport de chaque journal à la représentation globale

L'apport des deux journaux à l'activation des candidats est clairement différencié : l'*Humanité* active par exemple les sèmes /bien/ (substantif), /revenu/, /consommation/ et /dépense/, tandis que /ressource/ et /économie/ sont actualisés par le *Figaro*. On en déduit une vision plus globalisante, plus macroéconomique dans le *Figaro* et une vision plus localisante, plus en lien avec les consommateurs dans l'*Humanité*. Cette analyse est corroborée par la lecture humaine.

V VALIDATION ET APPORTS DU MODELE

Les analyses précédentes reposent uniquement sur la représentation sémique du corpus. Cette représentation est-elle pertinente par rapport à une approche lexicale classique et permet-elle d'expliciter les phénomènes en jeu dans l'interprétation humaine ? Nous avons étudié les points suivants : la convergence des résultats sur les plans sémique et lexical pour des axes sémantiques majeurs ; les plus et moins-values de notre enrichissement sémantique. Notre

démarche s'appuie sur l'observation des résultats sur le plan lexical et leur confrontation aux résultats obtenus de façon similaire sur le plan sémique.

V.1 Validité du modèle : convergence avec le plan lexical

Pour valider le modèle sémique, nous avons choisi deux approches. La première approche est non locale (elle n'est pas centrée sur un mot-pôle) donc globale, à l'échelle du sous-corpus d'un journal dans sa totalité. Elle porte sur les caractéristiques idéologiques dominantes d'un journal donné. Cette approche globale a été adoptée car elle s'avère plus facile à contrôler et à valider humainement. La seconde approche est en revanche locale, centrée sur le mot-pôle, avec réplication de l'approche sémique de la première partie sur le plan lexical. Seuls les comportements du voisinage ont été étudiés ; les études d'image sémique du mot-pôle sont, quant à elles, propres au plan sémique et difficilement transposables au plan lexical.

La première approche s'appuie sur la comparaison des spécificités calculées sur l'ensemble d'un journal par rapport à la totalité du corpus (réunion des deux journaux). Lors de la constitution du corpus, le parcours des articles a permis de dégager des axes très caractéristiques de chaque source journalistique. Ces axes (ou classes) ont servi à agencer les unités lexicales ou sémantiques. La démarche entreprise est la suivante : pour chaque classe définie, des formes et des sèmes pertinents pour l'axe concerné ont été sélectionnés. Autre critère de sélection : le rapprochement facile des formes et des sèmes, qui s'est fréquemment concrétisé par une analogie morphologique. L'objectif est d'observer la convergence des formes et des sèmes sur des axes sémantiques majeurs. Notons que la sélection effectuée n'est pas exhaustive, il n'est donc pas question d'étudier l'expansion des unités du plan lexical vers le plan sémique, ni quantitativement (étude de la proportion de sèmes par rapport au nombre de formes participant à une isotopie par exemple), ni qualitativement (introduction ou perte de nuances lors du passage des formes aux sèmes).

Le tableau ci-dessous présente un court extrait des résultats structurés en catégories. A chaque catégorie (lignes unicolonnes) est associée une liste de candidats-sèmes (quatrième colonne) ou formes (première colonne) affectés de leur spécificité. La spécificité de la forme (resp. candidat-sème) calculée dans le sous-corpus *le Figaro* est précisée en colonne 2 (resp. 5) ; celle de *l'Humanité* est précisée en colonne 3 (resp. 6).

Formes	*Figaro*	*Humanité*	Candidats-sèmes	*Figaro*	*Humanité*
Syndicats					
syndicat	-6	6	syndic#subst	<-50	>50
syndicats	-30	30	syndicat#subst	-34	34

syndical	-11	11	syndicalisme#subst	-22	22
syndicale	-12	12	syndical#adj	-13	13
syndicales	-12	12	intersyndical#adj	-3	3
syndicalisme	-4	4			
syndicaliste	-12	12	militant#subst	-4	4
syndicalistes	-9	9	militer#v	-38	38
syndicaux	-6	6	militant#adj	-25	25
intersyndicale	-3	3			
Thibault	-9	9	thibault#nam	-9	9
délégué	-10	10	délégué#subst	-6	6
délégués	-3	3			
Acteurs socio-économiques et catégories socio-professionnelles					
agriculteurs	-3	3	D=agriculture	-2	2
paysans	-3	3	agriculteur#subst	-6	6
ouvriers	-5	5	ouvrier#adj	-48	48
ouvrière	-4	4	ouvrier#subst	12	-12
travailleur	-4	4	travailleur#subst	<-50	>50
travailleurs	-23	23	travailleur#adj	-8	8
salarié	-7	7	salarié#subst	<-50	>50
salariés	<-50	>50	salarier#v	<-50	>50
			salarié#adj	-10	10
patron	7	-7	patronat#subst	-9	9
patrons	-3	3			
patronat	-10	10	patronal#adj	-42	42
patronale	-6	6			
patronales	-5	5	dirigeant#subst	-12	12
dirigeant	-	-	dirigeant#adj	-20	20
actionnaire	-2	2	actionnaire#subst	-14	14
actionnaires	-9	9			
consommateurs	6	-6	consommateur#subst	5	-5
investisseur	2	-2	consommateur#adj	22	-22

investisseurs	16	-16	investisseur#subst	17	-17
banquier	3	-3	banquier#subst	21	-21
épargnants	2	-2	épargnant#adj	10	-10

Figure 5 : comparaison des plans sémique et lexical

Les résultats obtenus indiquent d'une part une adéquation entre les observations humaines et les spécificités, d'autre part une convergence entre plan sémique et lexical. Cette convergence est manifeste dans l'extrait présenté : les acteurs de la finance (banquiers, épargnants, investisseurs) sont, dans l'ensemble, plus spécifiques du *Figaro,* tandis que les acteurs économiques des différentes classes sociales (au sens marxiste) sont dans l'ensemble plus spécifiques de *l'Humanité*. On remarque également que les spécificités marquées du plan lexical semblent se renforcer sur le plan sémique, comme si l'analyse sémique, que nous considérons par hypothèse comme une modélisation de l'interprétation, accusait des phénomènes plus ténus autrement.

V.2 Apports et limites du modèle interprétatif

La confrontation des plans sémique et lexical permet de valider le modèle sémique et fait émerger ses apports, à savoir un enrichissement quantitatif et qualitatif du plan lexical (cf. *infra* et section IV.1). Ainsi, concernant la classe // maladie-pathologie //, le nombre de représentants est multiplié et les nuances latentes au niveau lexical sont explicitées par exemple avec la présence des sèmes /pathologique/ et /maladie/.

Plan lexical		Comparaison avec le plan sémique			
Forme	Spécificité	Candidat-sème	Spécificité		
crise	9	contagion#subst	10	infection#subst	5
contagion	6	dysfonctionnement#subst	10	épidémie#subst	5
affectée	4	pathologique#adj	9	maladie#subst	5
injectés	3	troubler#v	8	remédier#v	4
affecter	3	trouble#subst	8	saignée#subst	4
aggravée	3	psychologique#adj	8	contagieux#adj	4
		physiologique#adj	6	perturbation#subst	4

Figure 6 : Axe maladie-pathologie, comparaison formes-sèmes

Cependant l'enrichissement sémantique est parasité par de mauvais candidats de types variés : des « métasèmes » provenant des définitions (consister#v ou relatif#adj, voir note 5), des candidats trop polysémiques (état#subst : au sens politique ou d'une situation) ou des faux-amis (complexe#adj dans la définition de *« réel »* (opposition des nombres réels aux nombres complexes)). Un filtrage est actuellement à l'étude, à partir de critères domaniaux (par exemple, pour l'élimination du domaine mathématique dans le cas de *« réel »*) et par l'identification préalable et la neutralisation des métasèmes.

VI CONCLUSION

Nous avons cherché à faire émerger le contenu sémantique d'une unité lexicale complexe à partir de ses occurrences en contexte, partant de l'hypothèse sous-jacente que les contextes permettent de caractériser l'unité lexicale étudiée. Du fait de l'enrichissement du corpus, donc de l'augmentation des informations à traiter, un critère de sélection est nécessaire : fondée sur une *présomption d'isotopie*, la spécificité d'un sème intègre à la fois sa récurrence et son degré de représentation par rapport au reste du corpus.

Différentes stratégies ont été mises en place pour filtrer l'information. La première est une recherche des sèmes pertinents parmi les plus spécifiques des contextes d' "*économie réelle*". Ces sèmes participent à des isotopies qu'il est possible de développer, mais leur réunion récurrente constitue une forme sémantique caractéristique du mot-pôle, ouvrant sur un sémème potentiel, ou un enrichissement de sémème. La deuxième stratégie part d'un sémème affecté *a priori* à "*économie réelle*" par réunion des sémèmes issus des définitions lexicographiques d' "*économie*" et de "*réel*". Seuls les candidats-sèmes de ce sémème sont observés. Parmi eux, les candidats les plus spécifiques des contextes contenant "*économie réelle*" sont considérés comme activés, les autres inhibés. Pour le contenu sémantique affecté *a priori* à une unité lexicale, cette seconde expérience permet de n'en garder que les éléments significatifs.

Ces deux approches offrent des perspectives complémentaires : à l'activation (sélection à partir d'un ensemble connu de candidats-sèmes) fait pendant l'enrichissement sémantique (affectation de sèmes extraits du voisinage à un sémème inconnu ou considéré comme incomplet). Une question reste toutefois en suspens : comment combiner les deux approches pour obtenir une représentation adéquate du sémème du mot-pôle en contexte ? Par ailleurs, dans le cas de l'activation des unités constitutives du sémème, les candidats-sèmes sont hiérarchisés de façon unidimensionnelle, selon une échelle de spécificité, mais ne sont pas organisés entre eux, en fonction de leur similarité de comportement par exemple. Ces considérations invitent ainsi à développer une représentation sémique plus complexe, multidimensionnelle, capable de donner une image plus structurée et épurée des sèmes caractéristiques d'un mot-pôle, autrement dit une *cartographie sémique* dont nous pensons qu'elle est de nature à rendre compte de l'émergence du sens en contexte.

VII REFERENCES

Dendien, J., Pierrel, J.-M. (2003) « Le trésor de la langue française informatisé. Un exemple d'informatisation d'un dictionnaire de langue de référence » *TA*L, 44-2, 11-37.

Fuchs, Catherine (2004) ed. La linguistique cognitive, Paris, Maison des Sciences de l'Homme / Ophrys.

Fuchs, Catherine (2008) « Linguistique française et cognition », *CMLF'08,* Institut de Linguistique Française (linguistiquefrancaise.org), Paris, 2008, 61-72.

Grzesitchak, Mick, Jacquey, Evelyne, Valette, Mathieu (2007) « Systèmes complexes et analyse textuelle : Traits sémantiques et recherche d'isotopies », *ARCo'07*, 227-235.

Lafon, P. (1984) *Dépouillements et statistiques en lexicométrie*, Genève-Paris, éd. Slatkine – Champion.

Rastier, François, Cavazza, Marc, Abeillé, Anne (1994), *Sémantique pour l'analyse. De la linguistique à l'informatique*, Paris, Masson.

Rastier, François (1991 [2001]) *Sémantique et recherches cognitives*, Paris, PUF.

Rastier, François (2003) « Parcours de production et d'interprétation. Pour une conception unifiée dans une sémiotique de l'action », in *Parcours énonciatifs et parcours interprétatifs*. *Théories et applications*, A. Ouattara, éd., Gap/Paris, Ophrys, 221-242.

Rouchaleau Y., 2008 , « *Traitement numérique du signal* », Les Presses de l'Ecole des Mines, p.18

Salem A. Lamalle C. Martinez W., Fleury S., Fracchiolla B., Kuncova A., Maisondieu A. (2003) « *Lexico3 – Outils de statistique textuelle. Manuel d'utilisation*. », Syled-CLA2T, Université de la Sorbonne nouvelle – Paris 3 : http://www.cavi.univ-paris3.fr/Ilpga/ilpga/tal/lexicoWWW.

Schmid G. (1994) « TreeTagger – a language indépendant part-of-speech tagger »

Valette, Mathieu (2008) « A quoi servent les lexiques sémantiques ? Discussion et proposition », *Cahiers du CENTAL*, 5, P.U. de Louvain, 43-58.

Valette, Mathieu, Estacio-Moreno, Alexander, Petitjean, Etienne, Jacquey, Evelyne (2006) « Éléments pour la génération de classes sémantiques à partir de définitions lexicographiques. Pour une approche sémique du sens », *Cahiers du CENTAL,* 2.1 (*TALN'06*), P.U. de Louvain, pp. 357-366.