

Métaphores et dynamique sémique : expériences sur corpus

Vincent Perlerin, Stéphane Ferrari, Pierre Beust
GREYC-CNRS UMR 6072 – 14000 Caen

I. Introduction

Les recherches en Traitement Automatique des Langues (TAL) présentées dans cet article, ont pour objet la caractérisation des métaphores à l'aide d'un modèle de la dynamique des sèmes. Notre démarche expérimentale se fonde sur une observation d'une métaphore conceptuelle conventionnelle sur un corpus électronique à l'aide d'outils informatiques. Après avoir posé le cadre et les objectifs de notre étude, nous exposerons les principes du modèle de représentation sémique que nous utilisons. Nous détaillerons ensuite la méthodologie et les outils d'analyse que nous avons élaborés pour cette étude. Finalement, nous présenterons des résultats d'expériences et leurs conclusions avant d'envisager les perspectives de ces travaux.

II. Cadre et objectifs

II.1 Étude sur corpus d'un fait de langue : la métaphore conceptuelle

Les travaux présentés ici, visent à l'élaboration d'outils automatiques pour la détection et l'aide à l'interprétation des métaphores. En TAL, différentes approches se côtoient, qui apportent déjà des solutions à ce problème. Sans en faire une présentation exhaustive, nous en retiendrons les principales orientations. Certaines recherches visent à expliciter la relation existant entre la source et la cible d'une métaphore, nécessitant de grandes bases de connaissances sémantiques, fortement structurées, pour leur mise en œuvre (Fass 1997). Selon les approches, la relation source-cible peut alors être vue comme essentiellement fondée sur l'analogie (Gentner 1983; Falkenhainer et al. 1989), ou plutôt porteuse de nouveauté, créatrice (Indurkha 1992 ; Gineste et al. 1997). D'autres travaux n'ont plus pour objectif de rendre explicite la relation source-cible, mais plutôt de vérifier que l'interprétation des métaphores est similaire à celle des autres énoncés. Ainsi, Kintsch (2000) montre que la signification d'une métaphore peut être représentée par un vecteur multidimensionnel, comme tout autre signification en ASL¹. Notre approche est en partie motivée par deux aspects spécifiques que les approches précédentes ne couvrent pas simultanément : l'utilisation de ressources simples à développer, d'une part, et la production pour un utilisateur final d'une aide à l'interprétation qui lui soit rapidement accessible, d'autre part. Les travaux qui cherchent à expliciter la relation source-cible de la métaphore ont actuellement pour inconvénient de nécessiter des ressources complexes, selon nous difficilement réalisables par un utilisateur non spécialiste. Ceux entrant dans la lignée de l'ASL présentent quant à eux l'inconvénient de ne pas fournir d'interprétation explicite de par la nature des représentations construites.

Nous proposons d'utiliser un modèle de représentation des significations des mots permettant à un utilisateur de représenter sa propre connaissance sur le lexique d'un domaine. Sa projection sur un texte sélectionné produit de nouvelles représentations, constituant pour l'utilisateur une aide à l'interprétation personnalisée. Ce modèle, *Anadia*, est présenté dans la sous-section suivante. Afin de valider son apport à l'interprétation des métaphores, nous nous intéressons plus particulièrement aux métaphores conventionnelles conceptuelles telles qu'introduites par Lakoff et Johnson (1980). Elles se caractérisent par l'existence d'emplois métaphoriques récurrents qui font intervenir à la fois un même domaine source (aspect conventionnel) et un même domaine cible (aspect conceptuel). Nous proposons d'en étudier une en particulier, afin de dégager des principes que nous espérons pouvoir généraliser par la suite. Pour mener nos observations, nous reprenons un corpus déjà utilisé pour une étude comparable (Ferrari 1997). Ce corpus est thématiquement homogène. Il est constitué d'environ 600 articles du journal *Le Monde sur Cd-rom*, tous relatifs au domaine de la bourse, et couvrant la période 1987-1989. Il contient notamment de nombreux emplois d'une métaphore conceptuelle décrivant les phénomènes boursiers en termes météorologiques. Le corpus est d'autant plus intéressant pour notre étude qu'il contient aussi des emplois non métaphoriques de termes en rapport avec la météorologie.

Avant de présenter plus en détail notre méthodologie d'analyse sur corpus et les outils que nous avons développés pour la mettre en œuvre, nous rappelons brièvement les principes sous-jacents au modèle de représentation lexicale que nous utilisons.

¹ ASL, Analyse Sémantique Latente, est une méthode de représentation où le sens d'un mot est défini par une position dans un espace à plusieurs dimensions. (Landauer et Dumais 1997).

II.2 Le modèle Anadia

Dans la lignée de travaux de Saussure (Saussure 1915), du courant structuraliste et plus récemment de la Sémantique Interprétative (SI) de Rastier (Rastier 1987), nous proposons un modèle de représentation des significations des lexies² spécifiquement dédié à des applications informatiques. Ce modèle, intitulé *Anadia*, a déjà fait l'objet de plusieurs implémentations³ et de plusieurs expérimentations informatiques (Beust 1998) (Nicolle et al. 2001).

Anadia est un modèle de représentation lexicale componentielle et différentielle. Componentielle en tant qu'il décrit des significations sous forme de *sémèmes* constitués de traits sémantiques (traits appelés *sèmes*) et différentielle en tant qu'il établit qu'un signifié est une *valeur*, au sens de Saussure, qui est déterminée par un jeu d'oppositions avec d'autres signifiés sémantiquement proches. Par exemple, la lexie « dépression » peut avoir deux significations : l'une indiquant un état mental, l'autre un phénomène météorologique. Chacune d'elles correspond à un *sémème*, par exemple le *sémème dépression* représentant un phénomène météorologique pourra être décrit par les sèmes suivants : /géographiquement situé/, /basse pression/, /mauvais temps/ ... Dans une perspective différentielle, les sèmes qui prennent part à un *sémème* ont pour but d'établir une distinction avec un autre *sémème* d'un même champ lexical, ainsi le sème /basse pression/ peut différencier le *sémème dépression* du *sémème anticyclone*.

Dans le modèle Anadia (à la différence du point de vue classique en linguistique), le sème est une relation de nature oppositionnelle dans laquelle une lexie s'inscrit au même titre que celles qui présentent des significations proches (Beust 1998). Par exemple, plutôt que de considérer que la lexie *avocat* porte le sème /défenseur/ qui est a priori sans relation avec le sème /qui requiert une peine/ porté dans la lexie *procureur*, nous considérons que les deux lexies portent un même sème oppositionnel actualisé de deux façons différentes. Ce sème est une opposition entre *défendre* et *requérir* qui structure des rôles possibles (ce qu'on appelle un domaine d'interprétation) dans le domaine juridique. Ce sème peut être noté [Rôle : 'défendre' vs. 'requérir'].

Pour structurer les significations d'un domaine lexical avec Anadia, il convient d'abord d'isoler les sèmes pertinents pour différencier entre elles les lexies du domaine. Ensuite, on construit différentes tables à l'aide de ces sèmes. Une table est une structure qui sert à identifier, à différencier et/ou à regrouper des lexies selon la combinatoire de certains sèmes. Un ensemble de tables décrivant le champ lexical d'un même domaine est appelé *dispositif*. Par exemple, les tables suivantes concernent certaines lexies du domaine de la bourse, elles sont extraites d'un dispositif ayant pour sujet ce domaine :

Agents, activités	Action	Rapport à l'activité
actionnaire, bénéficiaire, boursicoteur, investisseur, porteur, petit porteur, souscripteur	intervient	rôle
analyste	étudie, analyse	rôle
agent de change, opérateur	intervient	profession
économiste, COB	étudie, analyse	profession

Lieux	Situation
palais Brongniart, place financière, place bousière, rue Vivienne Wall Street, City de Londres	situé
entreprise, front office, back office, middle office, corbeille, salle de marché, banque, bureau	non situé

Figure 1. Extrait d'un dispositif Anadia en rapport avec le domaine boursier. Cet extrait présente deux tables, l'une, intitulée « Agents, Activités », est construite par la combinatoire des sèmes [Action : 'intervient' vs. 'étudie, analyse'] et [Rapport à l'activité : 'rôle' vs. 'profession'] et l'autre intitulée « Lieux » est construite à l'aide du sème [Situation du lieu : 'situé' vs. 'non situé'].

Il est important de noter qu'une table, et plus généralement un dispositif, représente toujours le point de vue (qui peut éventuellement être critiqué) de son auteur pour une tâche donnée. Par exemple, dans la Figure 1, les lexies *Wall Street* et *City de Londres* n'ont pas été distinguées (elles figurent dans la même ligne de table) mais

² Lexie est une notion plus générale que le mot : « pomme de terre » est une lexie faite de 3 mots.

³ <http://www.info.unicaen.fr/~perlerin/recherche/anadia.htm>

elles pourraient l'être par l'usage d'autres sèmes. Anadia est un modèle anthropocentré au sens de Thivitis (1998 : 47), où la construction de représentations est avant tout guidée par l'utilisateur et par la tâche pour laquelle il les construit. Les dispositifs construits n'ont pas pour objectif premier d'être universels ou de recueillir un consensus. Ils sont justifiés par leur auteur qui a pour charge de les compléter ou de les réviser.

Suivant les principes de la SI, les représentations lexicales componentielles sont utilisées pour rendre compte de la dynamique des sèmes dans l'interprétation de productions langagières. Le concept clef de cette dynamique indiquant une propriété spécifique des langues naturelles est celui d'isotopie : récurrence d'un même sème dans une unité syntagmatique. La recherche d'isotopies dans des corpus guidée par les dispositifs d'un utilisateur constitue une analyse interprétative personnalisée du corpus. Elle permet le repérage de différentes formes de répartition thématique (alternances des isotopies dites génériques (Rastier 1987)) et pour chaque dispositif elle pointe les différentes occurrences des principaux sèmes évoqués traduisant ainsi certains fonds sémantiques tels que les décrit Rastier (ibid.). La démarche générale et les outils mettant en œuvre ce type d'analyse sont le sujet de la prochaine partie.

III. Méthodes et outils pour une analyse de corpus

III.1 Démarche cyclique

L'utilisation du modèle Anadia pour l'étude d'un fait de langue sur corpus s'effectue par une succession d'étapes s'insérant dans une démarche cyclique (Figure 2). La première tâche à réaliser est la caractérisation du lexique du ou des domaines considérés. Un ensemble de lexies représentatives du corpus peut être obtenu à l'aide d'un logiciel d'extraction de termes⁴. Dans nos études, nous utilisons le logiciel *Memlabor* (Perlerin 2002). Celui-ci ne nécessite qu'une infime quantité de données extérieures au corpus et s'appuie essentiellement sur la loi statistique de Zipf (Zipf 1949). Il s'agit d'un logiciel proposant une aide à l'extraction de termes : une extraction supervisée.

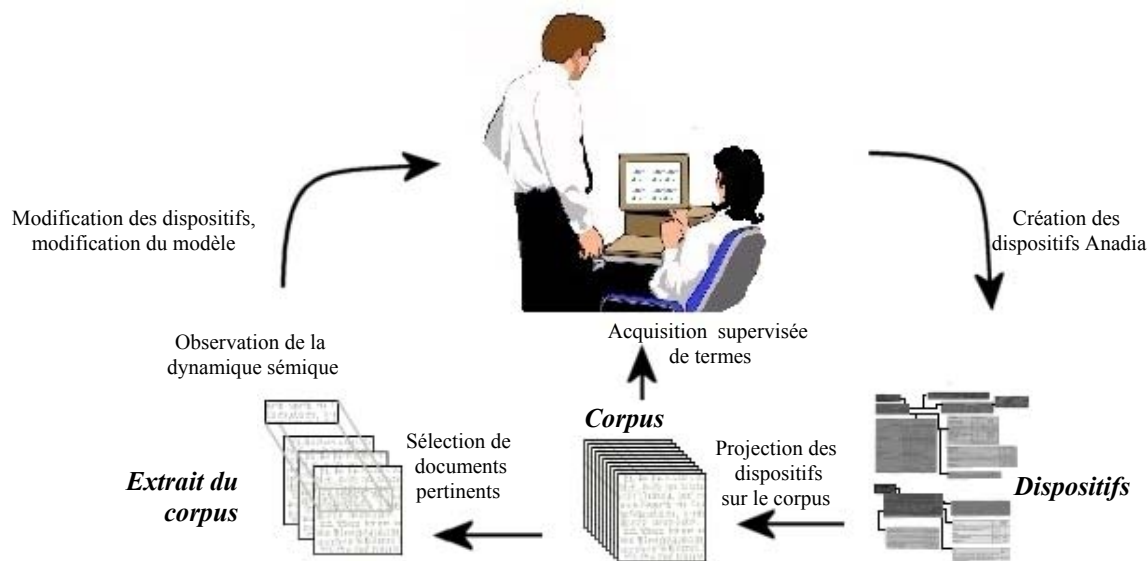


Figure 2. Utilisation du modèle Anadia pour l'étude d'un fait de langue sur corpus : une démarche cyclique.

Les lexies obtenues font ensuite l'objet d'une structuration en dispositifs Anadia grâce au logiciel *Anadia4.2* dont les principes fondateurs ont été décrits dans (Nicolle et al. 2001 : 58). Les dispositifs sont ensuite projetés sur les corpus. Il s'agit de repérer les lexies des dispositifs présentes dans les documents. Sur le corpus, la projection permet d'apprécier (ou non) l'homogénéité de l'ensemble par rapport à la répartition des domaines envisagés. Elle permet également de localiser les documents susceptibles de présenter des configurations sémiques remarquables pour le fait de langue envisagé.

La projection des dispositifs sur les documents jugés pertinents permet d'analyser ces derniers à un niveau macrosémantique. On y repère des isotopies et on étudie leur configuration. Ce type d'étude permet d'affiner les

⁴ Les propositions dans ce domaine sont très nombreuses : nous renvoyons le lecteur intéressé par ces techniques à la lecture de (Jacquemin 2001).

descriptions sémiques initiales et continuer ainsi la démarche cyclique en question. Au fur et à mesure, des conclusions peuvent être tirées pour l'étude en cours, des exemples notables peuvent analysés...

À la vue des tâches à réaliser, le besoin d'outils de visualisation efficaces et rapides se fait ressentir. Ils sont le sujet de la partie III.2.

III.2 Outils et visualisation des résultats

L'interface de visualisation créée pour nos travaux utilise les langages HTML, CSS et XML et XSLT. Cette interface est distribuée en open-source⁵ et fonctionne dans son état actuel avec *IExplorer5.0* et *JDK1.4* (et les versions ultérieures de ces gratuits). Le format des documents manipulables est le format TXT (d'autres formats sont en cours d'intégration). Cette interface permet d'assister chacune des étapes de la démarche cyclique décrite précédemment (Figure 3).

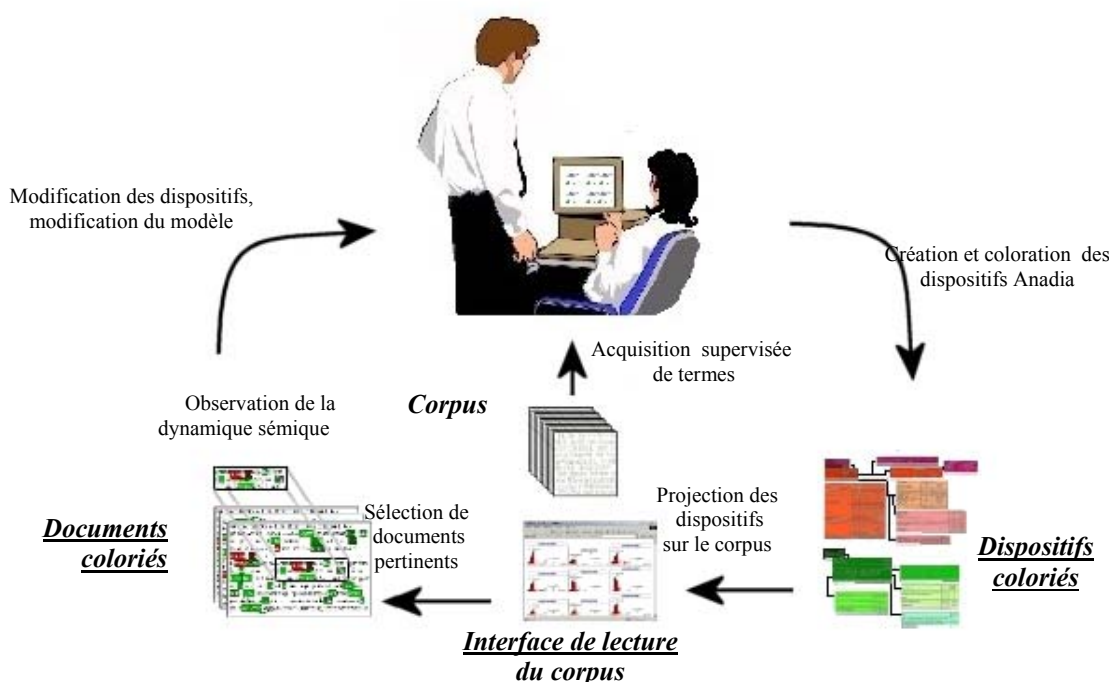


Figure 3. Utilisation du modèle Anadia pour l'étude d'un fait de langue sur corpus : outils de visualisation.

Le repérage des documents pertinents se fait par l'intermédiaire de la première partie de l'interface proposée : une page HTML regroupant l'ensemble des documents traités par les modules d'analyse d'*Anadia4.2* (Figure 4). Chaque document y est représenté sous la forme d'un graphique en histogrammes (Figure 5). Chacun des histogrammes correspond à une table d'un dispositif. Il est colorié en fonction de sa couleur dans les dispositifs et sa hauteur est proportionnelle au nombre de lexies du document qui appartiennent à cette table. Le passage de la souris sur l'un des histogrammes permet l'affichage du nom de la table ainsi que le nombre de lexies repérées lui appartenant. Un lien hypertexte permet d'accéder au document ainsi représenté. Cette page HTML permet d'apprécier la répartition thématique en termes de dispositifs et de tables Anadia à l'intérieur de chaque document du corpus. Elle donne donc une vision d'ensemble du corpus traité : une vision macroscopique de la répartition thématique.

⁵ Vous pouvez contacter les auteurs depuis <http://www.info.unicaen.fr/~perlerin/recherche/anadia.html>

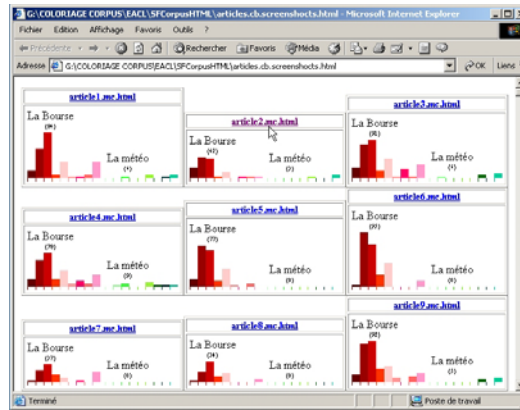


Figure 4. Interface permettant une vision macroscopique de la répartition thématique au sein du corpus.

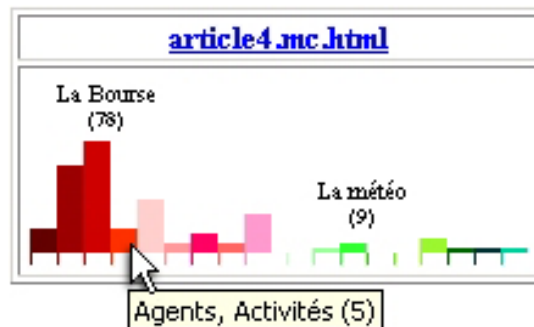


Figure 5. Représentation graphique de la répartition thématique au sein d'un document du corpus. Les chiffres entre parenthèses indiquent le nombre total de lexies en rapport avec les thèmes considérés. Dans l'exemple, 78 lexies du dispositif de la bourse et 9 du dispositif de la météorologie ont été repérées au sein de l'article représenté. Le pointeur de la souris est positionné sur l'un des histogramme : il permet l'affichage du nom de la table (« Agents, Activités ») et entre parenthèse le nombre de lexies repérées lui appartenant (5).

Une fois repérés, les documents pertinents doivent être analysés. Ces derniers sont aussi codés en HTML. Leur affichage dans un navigateur représente la deuxième partie de notre interface (Figure 6). Chacune des lexies appartenant aux dispositifs y est mise en valeur par une coloration en arrière plan avec la couleur de la table correspondant. Le passage de la souris sur une lexie mise en valeur permet l'affichage des représentations sémiques issues du dispositif Anadia (les jeux d'attributs-valeurs) (Figure 7). Cette technique permet une vision microscopique de la sémantique des documents.

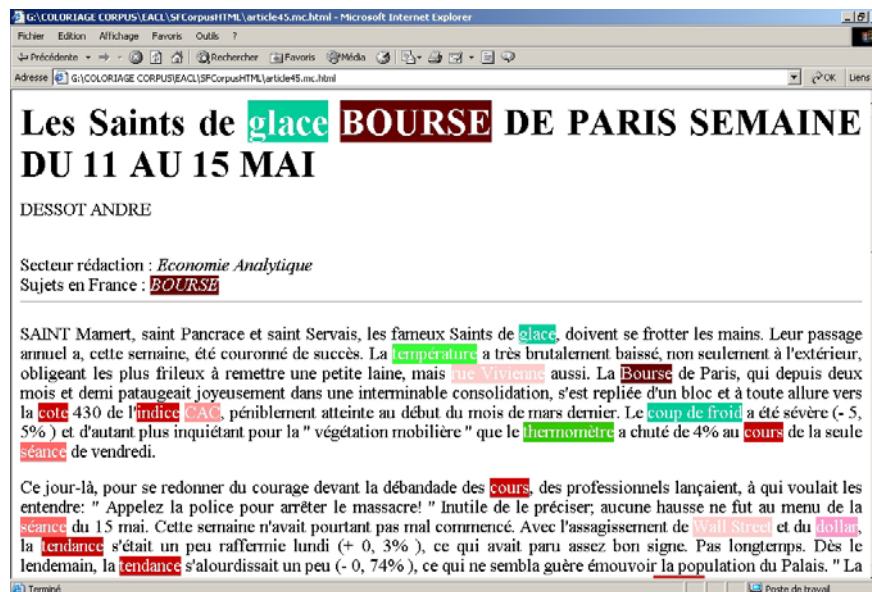


Figure 6. Article avec lexies mises en valeur.

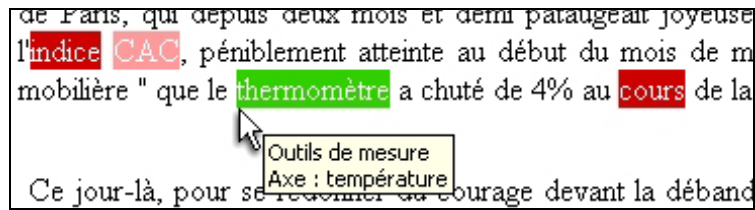


Figure 7. Passage de la souris sur une lexie mise en valeur : le navigateur affiche le nom de la table correspondante (« Outils de mesure ») ainsi que la (ou les) valeur(s) du (ou des) attribut(s) correspondant(s) (« Axe : température »).

L'interface de lecture des résultats d'analyse de corpus que nous proposons répond aux exigences de ce type d'application :

- elle est interactive (les couleurs et les groupes thématiques peuvent être modifiés par l'utilisateur, le passage de la souris sur certaines zones permet l'affichage d'informations ciblées),
- elle propose l'abstraction de certaines données pour rendre leur contenu plus explicite par l'utilisation des graphiques en histogrammes,
- elle propose la mise en valeur de données pertinentes pour l'étude par un moyen graphique facilement repérable,
- elle permet enfin un retour sur les documents du corpus par l'utilisation de liens hypertextes.

Si les atouts de cette interface sont suffisants pour notre étude (et d'autres similaires : Nicolle et al. 2001), un certain nombre d'informations intéressantes manquent pour d'autres travaux nécessitant par exemple, un repérage thématique dans des unités linguistiques plus petites que le texte (comme par exemple le paragraphe). Les graphiques en histogrammes présentés souffrent de ne pas refléter la linéarité des textes. Nous travaillons actuellement sur un projet tendant à remplacer ces graphiques par des représentations schématiques des documents en SVG (dessins vectoriels) permettant un repérage thématique dans des unités plus fines et la conservation de l'aspect général des documents. SVG permettra en outre l'utilisation de notre interface dans un plus grand nombre de navigateurs. Cette étude a également pour but d'appliquer ces propositions dans le domaine de la recherche documentaire personnalisée où les besoins en techniques de visualisation sont tout aussi cruciaux (ibid.).

Nous allons maintenant commenter des résultats d'expérience sur la métaphore. Ces résultats ont été obtenus à l'aide des outils décrits précédemment tels qu'ils se présentent dans leurs versions actuelles.

IV. Résultats

Les résultats de nos analyses sont de différentes natures : un retour sur le modèle de représentation lexicale (IV.1.) et une caractérisation de l'isotopie mise en jeu dans la métaphore (IV.2). Ces résultats ouvrent la voie à de nouvelles perspectives (IV.3.).

IV.1 Sèmes partagés et sèmes propres

L'expérimentation que nous avons menée a nécessité une première représentation des domaines de la bourse et de la météorologie à l'aide d'Anadia. Les tables correspondantes devaient servir d'amorce pour le processus cyclique décrit précédemment.

Descripteurs de température	Type	Caractère
surchauffe, torride, étouffant	chaud	fort
chaud, réchauffement, réchauffer	chaud	faible
froid, glacial, polaire, froid	froid	fort
frileux, frilosité, refroidir, petite laine, coup de froid	froid	faible

Figure 8. Exemple de représentation ayant servi d'amorce à la démarche cyclique

Certains manques sont apparus à l'issue de la projection de cette première représentation sur le corpus. Les outils de visualisation de la dynamique sémique développés (qui techniquement donnaient toute satisfaction) permettaient d'identifier les unités linguistiques où les deux domaines étaient co-présents (exemple 1) mais

n'offraient pas la possibilité de détecter des emplois métaphoriques ou d'en fournir une aide à l'interprétation satisfaisante.

(1) Extrait de l'article 42

L'**accalmie** intervenue ensuite sur le front **monétaire**, grâce aux interventions des **banques** centrales et aux déclarations apaisantes des officiels américains, a cependant **réchauffé** l'ambiance à la corbeille.

Les représentations proposées ne permettaient pas de trouver des isotopies remarquables entre les domaines source et cible de la métaphore conceptuelle étudiée. Un manque est apparu : certains sèmes devaient être partagés dans les représentations de ces domaines. L'évaluation qualitative de la première expérience nous a donc incité à envisager deux nouveaux types de sèmes⁶, les sèmes propres et les sèmes partagés. Les sèmes propres sont des sèmes dont l'emploi semble circonscrit au domaine du dispositif (comme par exemple [Axe : 'vent' vs. 'précipitations' vs. 'température' vs. 'pression'] dans le dispositif de la météorologie, Figure 9) ; les sèmes partagés sont des sèmes qui peuvent être utilisés dans plusieurs dispositifs (comme par exemple [Direction : 'monte' vs. 'descend'], Figure 11). Dans une seconde expérimentation, nous avons donc revu nos dispositifs en cherchant à utiliser certains sèmes partagés pour rendre compte de la métaphore conceptuelle entre les deux domaines.

Cette contrainte de construction des dispositifs a permis d'analyser des isotopies trans-dispositifs et de retrouver ainsi des résultats connus sur la métaphore mais jamais formalisés comme tels. L'absence de lexicalisations dans un domaine cible d'une métaphore conceptuelle pour certaines valeurs sémantiques est possiblement comblée par une lexicalisation décrite de manière similaire dans le domaine source. Cette absence donne lieu à des emplois métaphoriques. Par exemple, aucune lexie du domaine de la bourse désignant un phénomène dynamique relatif à une agitation très violente n'a été trouvée dans le corpus (i.e. la ligne de la table correspondante du dispositif était vide). Cependant les sèmes partagés définissant une agitation très violente dans le domaine de la météorologie permettent notamment de décrire la lexie *bourrasque*. Une utilisation métaphorique de *bourrasque* dans le domaine de la bourse peut donc renforcer certaines isotopies (par exemple celle du sème [Force : 'violent' vs. 'très violent']). De fait, nous avons trouvé dans le corpus de telles utilisations métaphoriques de *bourrasque* (exemple 2).

(2) Extrait de l'article 10

"**COURAGE**, fuyons... " *Tel était le slogan en vogue ces jours derniers sous les lambris du palais Brongniart, où la **bourrasque monétaire** a fait s'envoler nombre d'investisseurs et autant d'espoirs de nouveaux records.*

IV.2 Isotopie et lien métaphorique

Lors de la deuxième expérience, les dispositifs ont été construits avec neuf attributs partagés. Les exemples suivants montrent en quoi l'utilisation des sèmes partagés permet de mettre en évidence des configurations sémiques remarquables pour certains usages métaphoriques et quel type d'aide à leur interprétation nous pouvons proposer.

(3) Extrait de l'article n°126

Le **Dow Jones** par exemple, le **thermomètre** de la **Bourse** de New York, qui avait chuté de 508 points (...).

Dans cet exemple (3), la lexie *thermomètre* est source d'une métaphore *in praesentia* dont la cible est exprimée par la lexie *Dow Jones*. Cette métaphore trouve une explication à travers l'isotopie du sème [Rôle : 'intervenir' vs. 'étudier, analyser']. Les extraits des dispositifs suivants montrent les tables où sont définies les lexies *Dow Jones* et *thermomètre* :

⁶ Les aspects « propres » et « partagés » de ces sèmes sont distincts et peuvent se combiner avec les caractères « spécifiques » et « génériques » de la SI.

Objets du domaine	Rôle
ciel, pression, température	intervient
degrès, bar	étudie, analyse
Instruments de mesure	Axe
anémomètre	vent
pluviomètre	précipitations
thermomètre	température
mercure, baromètre	pression

Figure 9. Extrait du dispositif en rapport avec la météorologie (les tables ont été tronquées).

Objets du domaine	Rôle
marché, cours ...	intervient
graphiques, ratio, courbes, indices...	étudie, analyse
Indicateurs boursiers	Zone géographique
CAC, CAC40	France
Dow Jones, Nasdaq	U.S.A.
Nikkei	Japon
Dax	Allemagne
Footsie	Royaume Uni

Figure 10. Extrait du dispositif en rapport avec la bourse (les tables ont été tronquées).

Les sémèmes des lexies *Dow Jones* et *thermomètre* regroupent les sèmes de la table où la lexie est définie ainsi que ceux des tables de plus haut niveau dans la relation de sous-catégorisation. Les sémèmes de *Dow Jones* et *thermomètre* partagent donc un même sème [Rôle : ‘intervenir’ vs. ‘étudier, analyser’]. Cet héritage sémique permet de trouver l’isotopie du sème [Rôle] dans l’exemple 3. Une telle isotopie trans-dispositifs nous permet de suspecter un emploi métaphorique et de proposer une aide à son interprétation en indiquant que *thermomètre* est utilisé comme étant un objet qui sert à l’étude et l’analyse au même titre qu’un *graphique* ou une *courbe*. Une telle interprétation rend compte de la nature analogique du lien métaphorique.

Dans l’exemple suivant (4), la lexie *thermomètre* est source d’une métaphore *in absentia* dont la cible n’est donc pas lexicalisée dans le cotexte. Deux isotopies des sèmes [Direction : ‘monte vs. ‘descend’] et [Évaluation : ‘bien’ vs. ‘mal’ vs. ‘pas connoté’] peuvent cependant être repérées du fait de la présence de la lexie *krach* dans le même cotexte (Figure 11 et Figure 12). Celles-ci constituent un faisceau d’isotopies mettant en valeur la table « phénomènes dynamiques » du domaine de la bourse.

(4) Extrait de l’article n°153

Ce **krach** était dû (...) à la chute vertigineuse et incontrôlée du **dollar**, signe que la **tempête** affecte dorénavant les **marchés financiers**.

Phénomènes dynamiques	Direction	Evaluation
dépréciation, dévaluation, krach, dévaluer	descend	mauvais
baisse des cours, inflation	monte	pas connoté
hausse des cours, déflation	descend	pas connoté

Figure 11. Extrait du dispositif en rapport avec la bourse (les tables ont été tronquées).

Phénomènes dynamiques	Direction	Evaluation	Axe
gel, geler	descend	mauvais	température
intempéries	monte	mauvais	temps
accalmie	descend	bon	temps
Intempéries	Force		
rafale, tempête	violent		
cyclon, typhon	très violent		

Figure 12. Extrait du dispositif en rapport avec la météorologie (les tables ont été tronquées).

Les deux lexies sélectionnent la même valeur ‘mauvais’ du sème [Évaluation], cette appréciation est donc particulièrement pertinente dans l’interprétation de la métaphore. En revanche, les lexies sélectionnent deux valeurs distinctes du sème [Direction] (‘monte’ et ‘descend’). Ce dernier est donc moins saillant que le sème [Évaluation] dans l’interprétation. D’autre part, les sèmes [Axe] et [Force] de la source ne rentrent pas en jeu dans ce faisceau d’isotopies : ils sont selon nous virtualisés. Les sèmes actualisés constituent un premier élément de l’interprétation de la métaphore. Ici, la lexie *tempête* est utilisée pour décrire un *phénomène dynamique* évalué comme *mauvais*. Le sème [Force], non utilisé dans le dispositif de la bourse, apparaît comme partageable et caractérisant l’aspect créatif de cette métaphore. Nous sommes donc tentés de caractériser le phénomène dynamique boursier en question comme étant *violent*.

L'utilisation des sèmes partagés a ainsi permis d'observer par des isotopies trans-dispositifs la dynamique sémique mise en jeu dans les emplois métaphoriques des domaines envisagés. La détection de ces isotopies constitue un premier élément d'une aide à l'interprétation. Il reste maintenant à automatiser un certain nombre des opérations interprétatives que nous avons décrites.

IV.3 Perspectives

Comme l'exemple précédent l'a montré, l'introduction des sèmes partagés pourrait nous amener à vouloir structurer le plus possible le domaine cible en utilisant les mêmes sèmes que le domaine source. Le sème [Force] pourrait être ainsi utilisé dans une table sous-catégorisant certains phénomènes dynamiques boursiers. Cependant, une telle démarche ne pourrait rendre compte que du caractère analogique de la relation métaphorique. En n'utilisant pas systématiquement de tels sèmes dans les deux dispositifs, il devient possible de caractériser aussi l'aspect créateur de la métaphore. En effet, un sème tel que [Force] peut alors être transporté de la source vers la cible, bien que n'étant pas partagé entre les deux domaines, et n'intervenant pas dans une isotopie. Une perspective de nos travaux est de caractériser de telles afférences, tant en terme de dynamique sémique qu'en terme de nature de sème. Afin de mieux appréhender la nature des sèmes, il nous faudra également faire varier les métaphores conceptuelles et les corpus étudiés. La virtualisation et l'actualisation des sèmes devront également faire l'objet d'une automatisation.

A plus long terme, les propositions d'interprétations que l'on peut fournir peuvent devenir la base d'un système de génération automatique de métaphores⁷. Il apparaît par exemple, qu'un manque lexical pour une combinaison de sèmes donnée dans le domaine cible comblé dans le domaine source est un lieu propice à la métaphore. Cependant, certains problèmes restent posés : comment caractériser les tables sources et tables cibles pour un domaine donné, comment juger de la pertinence d'un emploi métaphorique dans un certain contexte ...

V. Conclusions

Le modèle Anadia permet de caractériser la dynamique sémique mise en jeu dans la métaphore. Nous avons montré que l'on pouvait retrouver la nature analogique autant que la nature créatrice de la relation entre la source et la cible d'une métaphore en étudiant les configurations sémiques englobant un représentant de ce phénomène. Pour rendre compte de l'aspect d'analogie, les techniques de structuration proposées pour les domaines sont suffisantes. Pour l'aspect créatif, la méthodologie doit être encore développée pour mieux rendre compte des transports possibles de traits sémantiques. En particulier, l'actualisation et la virtualisation (opérations interprétatives de la SI) doivent être systématisées au regard de la nature des sèmes mis en jeu. Nous considérons ici que l'interprétation de la métaphore n'est pas singulière mais qu'elle s'inscrit dans un cadre plus générique : celui des parcours interprétatifs.

L'évaluation de tels travaux n'est pas sans poser problème. Du fait même de la nature du fait de langue étudié et de l'absence de consensus autour de sa définition, une évaluation quantitative ne saurait passer que par le prétraitement manuel de corpora judicieusement sélectionnés. Ce type d'activité nécessite un investissement en temps extrêmement important : nous réfléchissons actuellement à un cadre de mise en œuvre optimal. L'évaluation qualitative quant à elle, est dépendante du caractère anthropocentré de nos représentations. Traiter de l'interprétation nécessite obligatoirement la prise en compte d'un utilisateur et de son objectivité par rapport à la tâche. Une série d'expérimentations a déjà commencé en ce sens (Perlerin et Beust 2002).

Dans la démarche cyclique présentée dans cet article, le corpus joue un rôle central : il est à la fois la source des ressources lexicales construites et le matériau de validation des propositions. Pour instrumentaliser cette démarche cyclique, nous avons développé des outils d'analyse de corpus à plusieurs niveaux. Au niveau global, une interface permet d'appréhender de façon macroscopique la répartition thématique au sein de l'ensemble de documents. Au niveau local, la microsémantique est abordée par le biais de l'utilisation de la couleur pour mettre en relief les lexies intéressantes. Depuis le début de nos travaux, les outils d'analyse de corpus que nous proposons ont été développés dans un souci de généricité. Si nous les avons utilisés dans le cadre d'une étude sur la métaphore, ils peuvent s'avérer utiles pour d'autres études nécessitant des interfaces de lecture rapide dédiées. C'est la raison pour laquelle ces outils sont développés à l'aide de langages assurant une grande portabilité et sont mis gratuitement à la disposition des chercheurs intéressés. Cependant, pour amorcer et poursuivre ce type d'études sur corpus, nous argumentons en faveur d'un processus de recherche et de développement en aller-

⁷ Ce domaine de recherche est déjà étudié pour, par exemple, rendre plus naturelles les interactions homme/machine.

retour entre des outils (des *logiciels d'étude*) et des corpus (des *corpus d'étude*), les premiers étant conditionnés par les seconds.

Bibliographie

Beust P. (1998). *Contribution à un modèle interactionniste du sens*. Thèse d'Informatique de l'université de Caen, France.

Falkenhainer B., Forbus K.D. et Gentner D. (1989). *The Structure-Mapping Engine : Algorithm and Examples*. Artificial Intelligence, 41/1, p. 1-63.

Fass, D. (1997). *Processing metaphor and metonymy*. Greenwich, Connecticut: Ablex Publishing Corporation.

Ferrari, S. (1997). *Méthode et outils informatiques pour le traitement des métaphores dans les documents écrites*. Thèse d'Informatique de l'université de Paris XI, France.

Gentner D. (1983). *Structure-Mapping: A Theoretical Framework for Analogy*. Cognitive Science, 7, p. 155-170.

Gineste, M.-D., Indurkha, B. et Scart-Lhomme, V. (1997). *Mental representations in understanding metaphors*. Rapport Technique, 97/2, Groupe Cognition Humaine, LIMSI-CNRS, Orsay.

Indurkha, B. (1992). *Metaphor and Cognition*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Jacquemin, C. (2001). *Spotting and Discovering Terms through NLP*. MIT Press, Cambridge MA.

Kintsch, W. (2000). *Metaphor comprehension: A computational theory*. Psychonomic Bulletin & Review, p. 257-266.

Lakoff G. and Johnson M. (1980). *Metaphors we live by*. University of Chicago Press, Chicago, U.S.A.

Landauer, T. K. and Dumais, S. T. (1997) *A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge*. Psychological Review, 104(2), p. 211-240.

Nicolle A., Beust P., Perlerin V. (2001). *Un analogue de la mémoire pour un agent logiciel interactif*. In Cognito 21, p. 37-66.

Perlerin V. et Beust P. (2002). *Pour une instrumentation informatique du sens*. Actes de l'atelier formation ARCo-CNRS. Île de Tatihou, Juillet 2002 – à paraître.

Rastier F. (1987). *Sémantique interprétative*. Presses Universitaires de France, Paris.

Saussure F. de (1915). *Cours de linguistique générale*. Editions Mauro-Payot : Paris (1986).

Théodore T. (1998). *Sémantique Interprétative Intertextuelle*. Thèse d'Informatique de l'université de Rennes I.

Wilks Y. (1978). *Making Preferences More Active*. Artificial Intelligence 11(3), p.197-223.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Addison-Wesley.