

---

# Antonymie et vecteurs conceptuels

**Didier Schwab — Mathieu Lafourcade — Violaine Prince**

*LIRMM - Laboratoire d'informatique, de Robotique  
et de Microélectronique de Montpellier  
MONTPELLIER - FRANCE.*

*{schwab,lafourca,prince}@lirmm.fr  
http://www.lirmm.fr/~{schwab,lafourca,prince}*

---

*RÉSUMÉ. Dans le cadre de recherches sur la représentation du sens en traitement automatique du langage, nous nous concentrons sur la représentation de l'aspect thématique des segments textuels sous la forme de vecteurs conceptuels. Le système d'apprentissage des vecteurs conceptuels se base sur l'analyse de définitions issues de dictionnaires à usage humain. Cette analyse ne gère pas les phénomènes liées à la négation. Cet article vise à étudier l'antonymie comme expression de l'opposition d'idées et ce comme partie de la négation. Nous abordons d'abord l'antonymie du point de vue linguistique et introduisons un modèle basé sur une idée de symétrie compatible avec les vecteurs conceptuels. Nous définissons ensuite des fonctions d'antonymies qui permettent d'obtenir un vecteur antonyme ou les items antonymes. Nous présentons des mesures permettant de savoir si un vecteur peut raisonnablement posséder un antonyme ou si un mot est l'antonyme d'un autre. Nous concluons sur les réalisations effectuées et les améliorations potentielles.*

*ABSTRACT. In the framework of research in meaning representations in NLP, we focus our attention on thematic aspects and conceptual vectors. The learning approach of conceptual vectors relies on the morphosyntactic analysis linked to vector propagation of human usage dictionary definitions. This analysis currently doesn't take into account negation phenomena. This work aims at studying the antonymy aspects of negation, in the larger goal of its integration into the thematic analysis. After a linguistic presentation of the antonymy, we build a model based on the idea of symmetry compatible with conceptual vectors. Then, we define antonymy functions which allow the construction of an antonymous vector and enumerate potentially antinomic lexical items. Finally, we introduce some measures functions, which evaluate how a given vector might accept an antonym and how a given word is an acceptable antonym of another term. We conclude with a presentation of the realisations and implementations of these functions and also with possible future improvements that would be desirable.*

*MOTS-CLÉS : représentation thématique, vecteurs conceptuels, antonymie complémentaire, scalaire et duale, mesure d'antonymie potentielle, mesure d'évaluation de l'antonymie*

*KEYWORDS: thematic representation, conceptuals vectors, complementary, scalar and dual antonymy, potential antonymy mesure, antonymy evaluation mesure*

---

## 1. Introduction

Dans le cadre de recherche sur la représentation du sens en traitement automatique du langage (TALN), nous nous concentrons sur la représentation de l'aspect thématique des segments textuels (documents, paragraphes, syntagmes, etc). Nous nous basons sur une approche mixte (symbolique et vectorielle) qui vise à combiner les informations déductibles des structures syntaxiques et les informations issues des représentations de sémantique lexicale. Certaines formes syntaxiques sont indirectement porteuses de sens et d'une façon générale peuvent être modélisées à l'aide de la théorie sens-texte et des fonctions lexicales [Melch'čuk and all 95]. Par exemple, la négation, très fréquente dans les textes, peut permettre, entre autres de produire des énoncés dont la forme n'est pas lexicalement avérée comme, par exemple, les syntagmes *il n'est pas sérieux*, *il n'est pas aimable*. Les mots 'sérieux' ou 'aimable' n'ont pas de contraires bien avérés. Les termes 'léger' et 'désagréable' ne sont tout au plus que des approximations, ambiguës. De plus, la négation ne signifie pas toujours le contraire d'une affirmation, comme dans le cas de la phrase, *elle n'est pas belle*, *elle est superbe*. Par contre, dans le cas, *il n'est pas mort* la négation exprime, a priori, l'idée opposée *il est vivant* avec cependant les problèmes de la polysémie et des sens figurés. On peut parler de 'vivant' dans le sens *gai, tonique*.

En particulier, nous nous intéressons à l'antonymie comme une des facettes de la négation et plus précisément à son aspect d'opposition d'idées. L'être humain apprend pour une grande part par analogie et lorsqu'il rencontre une nouvelle lexie, il tend à l'associer à un son, un graphème<sup>1</sup> ou à une idée déjà rencontrés. Ce sens peut, entre autres, être voisin (synonymie), être en relation de type méronymie (partie de), hyponymie (sorte de) ou bien contraire (antonymie). On en veut pour preuve qu'il n'est pas rare de trouver dans les dictionnaires des définitions expliquant un mot grâce à ces relations et en particulier, par un ou plusieurs antonymes. Par exemple, le mot 'inexistant' dans [Larousse 2001] est défini comme *Qui n'existe pas*. Dans [Robert 2000], 'inaction' est défini comme *absence d'action, d'occupation*. L'équipe TALN du LIRMM développe actuellement un système d'analyse de textes dans leurs aspects thématiques et de désambiguïsation lexicale basé sur les vecteurs conceptuels. Ces derniers représentent des sens associés à des mots ou à des expressions par le biais d'activation et d'inhibition sur un grand nombre de concepts. Le système d'apprentissage des vecteurs conceptuels, construit ou révisé automatiquement des vecteurs, à partir de définitions en langage naturel issues de dictionnaires à usage humain [Lafourcade 2001b]. Actuellement, l'analyse de définitions ne gère pas la question des négations. Cela a pour effet d'obtenir des activations conceptuelles quasi-identiques pour des mots dont la définition se réduit à une négation comme 'existant' et 'inexistant'. En termes de vecteurs, la prise en compte dans l'analyse thématique de la négation doit passer par la définition d'une fonction vectorielle d'antonymie. Dans notre premier exemple, le vecteur d'*inexistence* ne serait plus calculé directement à l'aide du vecteur 'exister' mais grâce au vecteur opposé à 'exister' calculé par cette fonction d'antony-

---

1. élément écrit.

mie. Cela peut permettre d'améliorer considérablement le système d'apprentissage et donc la précision de l'ensemble des vecteurs avec pour conséquence un gain qualitatif dans le processus global d'analyse de textes.

La fonction d'antonymie peut aussi servir à trouver une thématique contraire qui pourra être utilisée dans toute application génératrice de textes : recherche des idées inverses, paraphrases (par négation de l'antonyme), résumés de texte, etc. De plus, l'application de la fonction d'antonymie rend possible une forme de vérification de la cohérence de la base des vecteurs conceptuels. Durant les phases d'expérimentation, les résultats obtenus avec certains mots susceptibles de posséder des antonymes n'ont pas été satisfaisants. Cela peut résulter de plusieurs facteurs, dont en premier la polysémie. Le sens, véritablement recherché, du mot a été mal ciblé. La solution est de redéfinir le contexte thématique pour que la sélection soit correcte. La définition du mot (ou de certains de ses sens) peut aussi être ambiguë et l'apprentissage n'a pas permis de produire un vecteur thématiquement adéquat. Un ajustement manuel suivi d'une réindexation est en général nécessaire. On remarquera que la correction de vecteurs a pour effet global l'amélioration de la pertinence de l'ensemble de la base. Dans la suite, nous nous contenterons d'exhiber certains vecteurs défectueux sans pour autant nous concentrer sur leur correction.

Dans un premier temps, nous présenterons l'état actuel des recherches sur les vecteurs conceptuels. Nous les définirons ainsi que les diverses fonctions linguistiques qui leur sont associées. Nous présentons l'antonymie en linguistique et nous développons une théorie de la symétrie compatible avec les vecteurs conceptuels. Nous montrons ensuite la manière dont nous avons spécifiés les éléments calculables de cette théorie en termes de vecteurs conceptuels. En particulier, il est nécessaire de caractériser comment, à partir de mots, obtenir les mots opposés. Nous avons défini diverses fonctions d'antonymie, notamment la fonction *Anti* qui, à partir de vecteurs, renvoie un vecteur opposé est l'élément clé de ce travail. Nous l'avons étendu par la fonction *AntiLex*, incluant *Anti*, qui gère les passages délicats, du fait de la polysémie, des items aux vecteurs et des vecteurs aux items. Nous avons aussi défini deux mesures : une *mesure d'antonymie potentielle* qui permet d'évaluer si un item peut avoir un antonyme et une *mesure d'évaluation de l'antonymie* qui permet, elle, d'estimer raisonnablement si un item peut être l'antonyme d'un autre. Nous avons implémentés ces fonctions. Elles sont accessibles via les pages web des auteurs et peuvent être appliquées sur tout item de la base lexicale (environ 60000 items). Les résultats sont obtenus avec un taux de confiance qui est celui de la mesure d'évaluation.

## 2. Définitions et caractérisation de l'antonymie

La définition de l'antonymie fournie par [Larousse Universel 91] est la suivante : *Un antonyme est un mot qui a un sens opposé à celui d'un autre. Les antonymes, ou contraires, sont des mots appartenant obligatoirement à la même classe grammaticale ("grand" est l'antonyme de "petit" et non celui de "petitesse") et s'opposant par un ou plusieurs traits sémantiques, les autres étant communs. Par exemple, "monter" et*

"descendre" possèdent en commun le trait "déplacement vertical" et s'opposent par les traits "vers le haut" et "vers le bas". L'antonymie peut donc se définir comme une relation d'incompatibilité entre deux termes. Elle est, à cet égard, l'exact opposé de la synonymie. Cette définition de l'antonymie doit être revue à la lumière de la modélisation vectorielle utilisée. En effet, si on veut caractériser la construction de l'antonyme d'un concept, il est préférable d'utiliser la notion de *symétrie* plutôt que celle d'incompatibilité. Cette discussion, ainsi que ses résultats scientifiques sont proposés dans La symétrie se décline alors de différentes manières, selon la nature de son support. On distingue, comme supports :

- une propriété affectant une valeur étalonnable (valeur élevée, valeur faible) : par exemple, *chaud, froid* sont des valeurs symétriques de température ;
- l'application d'une propriété (applicable/non applicable, présence/absence) : par exemple, *informe* est antonyme de tout ce qui a une forme, *insipide, incolore, inodore, etc.* de tout ce qui pourrait avoir saveur, couleur, odeur. . .
- l'existence d'une propriété ou d'un élément considérés comme symétriques par l'usage (e.g. *soleil/lune*), ou par des propriétés naturelles ou physiques des objets considérés (e.g. *mâle /femelle, tête/pied, . . .* .

Notre idée est que les constructions d'antonymes sont dépendantes du type de support de symétrie ; il peut alors exister plusieurs types d'antonymes pour un même terme, comme il peut ne pas en exister d'évidents, si la symétrie n'est pas immédiatement décelable. En tant que fonction lexicale, comparée à la synonymie, on peut dire que si la synonymie est la recherche de la ressemblance avec comme test la substitution (*x est synonyme de y si x peut "remplacer" y*), l'antonymie est la recherche de la symétrie avec comme test la recherche du support de la symétrie (*x est antonyme de y s'il existe un support de symétrie t tel que x symétrique de y par rapport à t*).

De même que pour la synonymie [Lafourcade and Prince 2001], il n'existe pas d'antonymes absolus, i.e. deux mots qui seraient antonymes l'un de l'autre quel que soit le contexte. L'antonymie s'apprécie toujours en contexte. Par exemple, *frais* peut être le contraire de *tiède, chaud, racorni, flétri, maladif, rassis, confit, sec, surgelé, pourri, . . .*

## 2.1. Antonymie et linguistique

En général, les linguistes comme Palmer ou Lyons considèrent deux différents types de relation entre antonymes, les complémentaires et les scalaires.

### 2.1.1. Antonymes complémentaires

Les antonymes complémentaires sont ceux qui des couples tels que *pair/impair, présence/absence*

il est présent  $\Rightarrow$  il n'est pas absent  
il est absent  $\Rightarrow$  il n'est pas présent

il n'est pas absent  $\Rightarrow$  il est présent  
il n'est pas présent  $\Rightarrow$  il est absent

En termes de logique,

$$\begin{array}{ll} \forall x \ P(x) \Rightarrow \neg Q(x) & \forall x \ \neg P(x) \Rightarrow Q(x) \\ \forall x \ Q(x) \Rightarrow \neg P(x) & \forall x \ \neg Q(x) \Rightarrow P(x) \end{array}$$

Cela correspond à la relation de disjonction exclusive. Dans ce cadre, l'affirmation d'un des termes implique nécessairement la négation de l'autre. Sur le plan de la symétrie, l'antonymie complémentaire de la linguistique présente deux types de symétrie :

- une symétrie de valeur dans un système à deux valeurs seulement, comme dans l'exemple ci-dessus

- une symétrie par rapport à l'application d'une propriété : le *noir* est l'absence de couleur, il est donc "opposé" à toute couleur, et toute combinaison de couleurs.

Le mélange des deux types de symétrie peut parfois introduire des divergences entre la linguistique et une modélisation qui se voudrait cohérente avec le modèle vectoriel. Pour en donner un exemple, en linguistique, on considère aussi le couple *vivant/mort* comme relevant de l'antonymie complémentaire, alors que sur un plan logique

$$\forall x \ \neg \text{vivant}(x) \Rightarrow \text{mort}(x)$$

est falsifiable puisque l'inanimé n'est ni vivant ni mort. Une bonne façon ne pas falsifier ces propriétés est de circonscrire l'application des propriétés aux termes pour lesquelles elles sont pertinentes, ce qui peut se faire par une conjonction où C est la condition préalable pour l'application de P.

$$\forall x \ C(x) \wedge P(x) \Rightarrow \neg Q(x)$$

$$\forall x \ \text{animé}(x) \wedge \text{vivant}(x) \Rightarrow \neg \text{mort}(x)$$

En pratique, l'inanimé est complémentaire du vivant et du mort, par non application de la propriété de "vie", alors que *vivant/mort* est un couple de symétriques, en valeur, dans un système à deux valeurs. En rendre compte sur le plan logique nécessite des précautions qui, comme nous le montrerons, ne sont pas nécessaires dans la modélisation vectorielle.

Une modélisation logique peut aussi être remise en cause par la possibilité d'usage linguistique. Si la logique peut accepter ce qui est ni vivant ni mort (l'inanimé en l'occurrence) en cherchant à isoler le type de symétrie, la langue peut aussi accepter le paradoxe, qui défie la logique, et qu'elle atteste parfois par l'usage. C'est le cas des antonymes de notre exemple, puisque la forme *mort-vivant* existe pour exprimer ce qui possède les deux propriétés.

### 2.1.2. Antonymes scalaires

Les antonymes scalaires (ou gradables) sont ceux dont la symétrie est typiquement la symétrie par rapport à une valeur, dans un système dont la valeur de référence n'est pas toujours représentée par un mot. On trouve dans cette catégorie des antonymes tel que *grand/petit* :

Cet homme est grand	⇒ Cet homme n'est pas petit
Cet homme est petit	⇒ Cet homme n'est pas grand
Cet homme n'est pas grand	⇒ Cet homme est petit ∨ cet homme est de taille moyenne
Cet homme n'est pas petit	⇒ Cet homme est grand ∨ cet homme est de taille moyenne

Cet homme est "ni grand ni petit" qui désigne en général la taille moyenne, mais qui ne signifie pas dans le cas présent (comme dans le cas de *vivant/mort*) que la propriété ne s'applique pas. C'est simplement qu'il existe ici une "valeur neutre" à partir de laquelle les autres s'échelonnent. En logique classique, on pourrait l'exprimer par

$$\forall x P(x) \Rightarrow \neg Q(x) \wedge R(x)$$

si R est la propriété ayant la valeur de référence (neutre ou médiane)

$$\begin{array}{ll} \forall x Q(x) \Rightarrow \neg P(x) \vee R(x) & \forall x \neg Q(x) \not\Rightarrow P(x) \\ \forall x, R(x) \Rightarrow \neg Q(x) \wedge \neg P(x) & \forall x \neg P(x) \not\Rightarrow Q(x) \end{array}$$

La valeur de référence peut ne pas être la seule valeur possible, mais un des éléments remarquables de l'échelle (pour des propriétés multi-valuées par exemple). L'usage de termes gradables implique toujours une évaluation et donc une comparaison. Celle-ci peut être explicite : *Jean est plus petit/plus grand que Pierre, il avance/recule* (le terme moyen étant *immobile*). Elle peut aussi être implicite et renvoyer à des normes tacitement admises par l'individu ou la communauté à laquelle il appartient : *il fait chaud* dit par un habitant d'un pays équatorial ne se réfèrera pas à la même idée de chaleur (donc à la même valeur de référence) qu'un habitant des fjords de Norvège.

L'usage de deux termes antonymes n'est pas toujours symétrique. On ne dit pas *la mer est basse de 137 mètres (sous entendu par rapport à la colline)* mais *la colline est haute de 137 mètres (sous entendu par rapport au niveau de la mer)*. L'usage consacre celui des antonymes qui est en isotopie, dans son acception linguistique introduite par Greimas et Pottier, et développée par Rastier, avec le terme que l'on veut évaluer (ici la colline qui est une protubérance, et qui renforce le terme de hauteur).

## 2.2. Extensions de l'antonymie

Nous considérons, dans notre étude, un type d'antonymes supplémentaire : les antonymes *duals*. Ils sont composés de deux sous-familles : les antonymes conversifs et les duals propres. Ils correspondent au troisième type de symétrie, celui que l'usage et la nature même des objets peut introduire.

### 2.2.1. Conversifs

Les conversifs (ou réciproques) sont des couples tels que *acheter/vendre, prêter/emprunter, mari/femme, avant/après, père/fils*. De nombreux linguistes comme Igor Mel'čuk [*Melch'čuk and all 95*] ne les considèrent pas comme des antonymes. La fonction *anti* de son Dictionnaire Explicatif et Combinatoire [*Melch'čuk and all 84*]

désigne en réalité les antonymes complémentaires et les antonymes scalaires. Il dédie aux conversifs une autre fonction lexicale *conv*.

Pierre est le père de Marc  $\leftrightarrow$  Marc est le fils de Pierre.

Ce qui s'exprime, en terme de logique, comme :

$$\forall x, y P(x, y) \leftrightarrow Q(y, x)$$

Dans le cas des conversifs, si on remplace dans une phrase un terme A par son réciproque B, on peut systématiquement rétablir la synonymie entre les deux phrases à condition de permuter les arguments syntaxiques mis en relation par A comme le montre la formule. Ainsi, pour les conversifs, il y a symétrie par rapport à la place des arguments (P est réciproque de Q).

### 2.2.2. Duals

Les duals propres sont une notion d'antonymie que nous introduisons pour rendre compte d'un effet particulier de mise en relation de termes où la symétrie porte cette fois-ci sur des fonctions culturelles (symétrie consacrée par l'usage) et spatio-temporelles (propriétés particulières de l'espace-temps). L'antonyme dual d'un mot est le pendant de celui-ci. Les duals sont des mots que la culture associe comme *soleil/lune*, ou qui ne vont pas, à priori, l'un sans l'autre comme *question/réponse* ou alors sont l'expression d'une antonymie temporelle i.e. qui exprime le passage d'un état à un autre comme *naissance/décès*. Dans ce troisième cas, on peut remarquer que ces deux événements marquent le passage entre deux antonymes complémentaires (*inexistence/existence* dans le cas de *naissance/décès* ou bien *présence/absence* dans le cas de *départ/arrivée*). L'antonymie duale propre présente naturellement une symétrie qui n'est pas relevée dans l'échange des places d'argument puisqu'il s'agit de prédicats unaires. Elle exprime le fait que si un des deux prédicats est à vrai, il existe une valeur pour lequel l'autre l'est aussi nécessairement. Pour la modéliser, on écrira

$$\forall x P(x) \leftrightarrow \exists Q, Q(x)$$

avec Q dual de P qui modélise par exemple le fait que si x a un début, alors il existe aussi une fin à x ou

$$\forall x P(x) \leftrightarrow \exists y, \exists Q, Q(y)$$

avec Q dual de P qui exprime que si x est une question, il existe un objet y et il existe un prédicat réponse, tel que y est une réponse à x.

Cette nécessité du prédicat dual peut rendre compte de certains couples de description temporelle (*avant/après* en prédicats unaires, linguistiquement différenciés sur le plan de la catégorie grammaticale, comme dans *il y a un avant et un après* à ne pas confondre avec *avant/après* qui sont des scalaires avec comme valeur médiane *pendant*).

### 3. Vecteurs Conceptuels

Dans le cadre de recherche sur la représentation du sens en traitement automatique du langage (TALN), nous nous concentrons sur la représentation de l'aspect thématique des segments textuels (documents, paragraphes, syntagmes, etc) sous la forme de vecteurs conceptuels. Cette approche tire son origine de [Chauché 1990] pour l'utilisation d'un jeu de concepts prédéterminé, mais s'inspire aussi du modèle vectoriel [Salton and MacGill 1983] et du modèle LSI [Deerwester and all. 1990] pour la reconnaissance et l'exploitation de l'interdépendance des concepts.

Le modèle de vecteurs conceptuels s'appuie sur la projection de la notion linguistique de champ sémantique dans un modèle mathématique d'espace vectoriel. L'idée est qu'à partir d'un espace de notions élémentaires, des concepts, il est possible de construire des vecteurs (dit conceptuels) et de les associer à des items lexicaux<sup>2</sup>. Les mots polysémiques fusionnent les multiples vecteurs correspondant à leurs différents sens. Cette approche par vecteurs repose sur des bases mathématiques élémentaires bien connues et il est ainsi possible d'effectuer des manipulations formellement bien fondées auxquelles sont attachées des interprétations linguistiques raisonnables.

Les concepts sont définis selon un thésaurus (en ce qui nous concerne, il s'agit de la langue française [Larousse 1992] où 873 concepts sont répertoriés). L'hypothèse principale est que cet ensemble forme un espace générateur pour les mots de la langue (espace qui n'est probablement pas libre). Dès lors, tout mot se projette sur cet espace selon le principe énoncé ci-après.

#### 3.1. Principe

Soit  $\mathcal{C}$  un ensemble fini de  $n$  concepts, un vecteur conceptuel  $V$  est une combinaison linéaire des éléments  $c_i$  de  $\mathcal{C}$ . Pour un sens  $A$ , le vecteur  $V_A$  est la description, en extension, des activations des concepts de  $\mathcal{C}$ . Par exemple, les sens de 'ranger' et de 'couper' peuvent être projetés sur les concepts suivants (les  $CONCEPT[intensité]$  sont ordonnés par intensité décroissante) :

$$V_{porte} = (OUVERTURE[0.3], BARRIÈRE[0.31], LIMITE[0.32], FERMETURE[0.33], EXTÉRIEUR[0.35], \\ (CONTENANT[0.36], INTÉRIEUR[0.37], CENTRE[0.375]$$

$$(JEU[0.8], LIQUIDE[0.8], CROIX[0.79], PARTIE[0.78], MÉLANGE[0.78] FRACTION[0.75], SUP- \\ PLICE[0.75], BLESSURE[0.75], BOISSON[0.74], \dots).$$

En pratique, plus  $\mathcal{C}$  est grand, plus fines sont les descriptions de sens offertes par les vecteurs, mais plus leur manipulation informatique est lourde (on rappelle que dans

---

2. Les items lexicaux regroupent les mots ou les expressions qui constituent les entrées du lexique. Par exemple, 'voiture' tout comme 'pomme de terre', 'moulin à vent' et même des termes techniques comme 'pompe bivalve à échappement central' sont des items lexicaux. Par la suite, par abus de langage, nous utiliserons parfois mot ou terme pour qualifier un item lexical.



nos expérimentations,  $\dim(\mathcal{C}) = 873$ , ce qui correspond au niveau 4 des concepts définis dans [Larousse 1992]) Il est clair, que pour des vecteurs denses<sup>3</sup>, l'énumération des concepts activés est vite fastidieuse et surtout difficile à évaluer. On préférera en général, procéder par sélection de termes thématiquement proches. Par exemple, les termes proches de même catégorie lexicale et ordonnés par distance thématique décroissante des mots 'ranger' et 'couper' sont :

'ranger' : 'trier', 'cataloguer', 'sélectionner', 'classer', 'distribuer', 'grouper', 'ordonner', 'répartir', 'aligner', 'caser', 'arranger', 'nettoyer', 'distribuer', 'démêler', 'ajuster' ...

'couper' : 'cisailier', 'émincer', 'scier', 'tronçonner', 'ébarber', 'entrecouper', 'baptiser', 'recouper', 'subsectionner', 'bêcher', 'hongrer', 'essoriller', 'rogner', 'égorger', 'écimer', ...

Afin de pouvoir rechercher les mots les plus proches, il faut définir une distance. Dans le cadre des vecteurs conceptuels, il s'agit d'une distance angulaire.

### 3.2. Distance angulaire

Il est souhaitable de pouvoir mesurer la proximité entre les sens représentés par deux vecteurs (et donc celle de leur mot associé). Soit  $Sim(X, Y)$  la mesure de *similarité* entre deux vecteurs  $X$  et  $Y$ , utilisée habituellement en recherche d'informations [Morin 1999]. Cette fonction est définie selon la formule ci-dessous

$$\vartheta^2 \rightarrow [0, 1] : Sim(X, Y) = \cos(\widehat{X, Y}) = \frac{X \cdot Y}{\|X\| \times \|Y\|} \quad [1]$$

où "·" est le produit scalaire. Il est supposé ici que les composantes des vecteurs sont toujours positives ou nulles (ce qui n'est pas nécessairement le cas).

Enfin, nous définissons une fonction de *distance angulaire*  $D_A$  entre deux vecteurs  $X$  et  $Y$  selon la formule :

$$\vartheta^2 \rightarrow [0, \frac{\pi}{2}] : D_A(X, Y) = \arccos(Sim(X, Y)) \quad [2]$$

Intuitivement, cette fonction constitue une évaluation de la *proximité thématique* et est en pratique la mesure de l'angle formé par les deux vecteurs. En général, pour une distance, si  $D_A(X, Y) \leq \frac{\pi}{4}$  alors  $X$  et  $Y$  sont considérés comme sémantiquement proche et partagent des concepts. Si  $D_A(X, Y) \geq \frac{\pi}{4}$ , la proximité sémantique de  $A$  et  $B$  est considérée comme faible et aux alentours de  $\frac{\pi}{2}$ , les sens sont sans rapport. La synonymie (dans son acception la plus large) est incluse dans la proximité thématique, cependant elle exige de plus une certaine forme de concordance des catégories morphosyntaxiques. Par exemple, *vite* a pour synonyme *rapide* et non *rapidité*. L'inverse n'est évidemment pas vrai. Pour plus de détails sur la synonymie, [Lafourcade and Prince 2001].

3. Les vecteurs denses sont ceux qui ont très peu de coordonnées nulles.

La métaphore de la nuit étoilée peut aider à appréhender cette idée de distance angulaire pour calculer la proximité thématique. Nous pouvons nous représenter l'espace des sens comme un ciel rempli d'étoiles. Les étoiles sont les items lexicaux. Les mots, tout comme les étoiles, forment des constellations. Certaines parties de l'espace sont très denses tandis que d'autres sont sous-peuplées. Un sens est une direction de l'espace et non un point. Un observateur ne peut connaître exactement la distance entre une étoile et la terre mais il connaît sa direction. Dans le ciel, la distance entre deux étoiles est la distance apparente, l'angle entre deux étoiles. Il en est de même avec les items lexicaux.

Pourquoi la norme n'est-elle pas considérée ? Prenons un vecteur représentant l'idée de couleur rouge. Prenons un autre vecteur colinéaire représentant l'idée de quelque chose de plus rouge. Dans ce cas, cela signifie, que le premier vecteur est moins rouge, ce qui signifie qu'il peut être plus bleu (ou plus jaune ou plus vert,...). Il ne pointe donc pas dans la même direction ce qui n'était pas notre première supposition. La norme d'un vecteur peut être utilisée comme une idée d'intensité mais pas directement comme un évaluateur des activations thématiques et des rapprochement entre vecteurs.

$D_A$  est, contrairement à la mesure de similarité, une vraie distance et elle vérifie les propriétés de réflexivité (3), symétrie (4) et inégalité triangulaire (5) :

$$D_A(X, X) = 0 \quad [3]$$

$$D_A(X, Y) = D_A(Y, X) \quad [4]$$

$$D_A(X, Y) + D_A(Y, Z) \geq D_A(X, Z) \quad [5]$$

Par définition, nous posons,  $D_A(\vec{0}, \vec{0}) = 0$  et  $D_A(X, \vec{0}) = \frac{\pi}{2}$  avec  $\vec{0}$  dénotant le vecteur nul<sup>4</sup>.

Par la suite, nous ne parlerons de distance que lorsque les propriétés de réflexivité, symétrie et inégalité triangulaire sont vérifiées. Nous parlerons de mesures dans le cas inverse.

La distance angulaire entre quelques items de la base<sup>5</sup> donne les angles suivants exprimés en radians :

$$\begin{array}{ll} D_A(V_{mésange}, V_{mésange}) = 0 & D_A(V_{mésange}, V_{passereau}) = 0,358 \\ D_A(V_{mésange}, V_{oiseau}) = 0,552 & D_A(V_{mésange}, V_{train}) = 1,288 \\ D_A(V_{mésange}, V_{insecte}) = 0,577 & D_A(V_{mésange}, V_{couleur}) = 0,598 \end{array}$$

La première a une interprétation simple, l'item *mésange* ne peut être plus proche d'autre chose que de lui-même. Le deuxième et le troisième ne sont pas surprenant

4. Le vecteur n'est sans doute pas représenté par un mot de la langue. Il s'agit d'une idée qui n'active ... aucun concept ! C'est l'idée vide.

5. Les exemples de cet article sont extrait de <http://www.lirmm.fr/~schwab> pour la partie concernant plus spécifiquement l'antonymie et <http://www.lirmm.fr/~lafourca> pour tous les autres exemples.

si nous nous souvenons que la mésange est de la famille des passereaux qui sont des oiseaux. Une mésange a beaucoup moins de points communs avec un train ce qui explique l'angle relativement proche de  $\pi/2$ . Pour expliquer le peu de différence entre  $D_A(V_{mésange}, V_{oiseau})$  et  $D_A(V_{mésange}, V_{insecte})$ , il faut se souvenir que  $D_A$  est une distance thématique et non une distance ontologique ou du type hypéronymie (*sorte de*) ou méronymie (*partie de*). Si nous examinons la définitions de *mésange* à partir de laquelle le vecteur conceptuel est calculé, l'interprétation semble plus claire.

**mésange** : *Oiseau passériforme insectivore, long de 10 à 14 cm, au plumage coloré, aux mouvements vifs, dont la plupart des espèces appartiennent au genre Parus.*

On considèrera, en toute généralité, l'extension du domaine image de  $D_A$  à  $[0, \pi]$  afin de comparer des vecteurs ayant des composantes négatives. Cette généralisation ne change pas les propriétés de  $D_A$ . On remarquera, de plus, que la distance angulaire est insensible à la norme des vecteurs ( $\alpha$  et  $\beta$  étant des scalaires) :

$$D_A(\alpha X, \beta Y) = D_A(X, Y) \quad \text{si } \alpha\beta > 0 \quad [6]$$

$$D_A(\alpha X, \beta Y) = \pi - D_A(X, Y) \quad \text{si } \alpha\beta < 0 \quad [7]$$

### 3.3. Construction des vecteurs conceptuels

La construction des vecteurs conceptuels se fait à partir de définitions extraites de diverses sources (dictionnaires, listes de synonymes, indexations manuelles,...). Ces définitions sont analysées et le vecteur conceptuel correspondant à une définition calculée. Cette méthode d'analyse forme, à partir de vecteurs conceptuels déjà existants et de définitions, de nouveaux vecteurs. Il est nécessaire d'effectuer l'amorçage du système d'apprentissage à partir d'un noyau constitué de vecteurs calculés au préalable. Nous l'avons vu, tout vecteur conceptuel est une combinaison linéaire des vecteurs des concepts générateurs, les vecteurs génératifs. Un ensemble de vecteurs est ensuite indexé manuellement pour les termes les plus courants ou les plus délicats. Cet ensemble constitue la base d'items lexicaux sur laquelle se basera l'apprentissage. Les vecteurs génératifs sont donc la base du système. Leur construction est fondamentale.

#### 3.3.1. Vecteurs conceptuels génératifs

Les vecteurs génératifs sont les vecteurs associés aux concepts générateurs. Ils sont la base du système et sont les seuls à être construits manuellement. La méthode de construction de ces vecteurs se base sur la structure hiérarchique de l'ontologie des concepts (structure d'arbre).

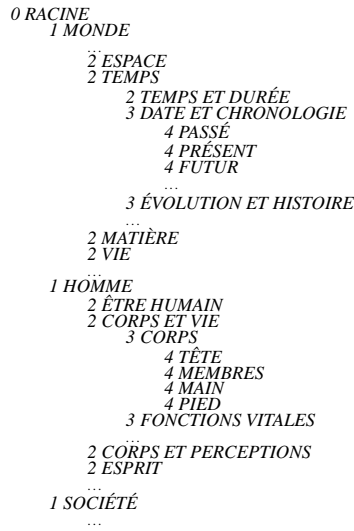
Le point de départ de cette construction est un vecteur booléen. Le vecteur booléen du concept  $i$  est le vecteur dont tous les éléments sont à 0 sauf la composante  $i$  qui elle est à 1. Cette construction est simple et elle obtient souvent de bons résultats mais semble inadéquate dans plusieurs cas. Il paraît curieux que deux concepts proches comme le sont *GUERRE* et *PAIX* partagent quantitativement autant d'idées que *PAIX* et *CHAMPIGNON*. Nous l'avons déjà dit, les concepts ne sont clairement pas indépendants

et leurs vecteurs respectifs doivent en tenir compte. L'ensemble des concepts définis selon [Larousse 1992] est hiérarchiquement ordonné selon un arbre ontologique. La construction des vecteurs génératifs est basée sur cette structure et plus particulièrement sur la distance ultramétrique entre deux concepts. Il s'agit du chemin minimal à parcourir dans l'arbre des concepts pour aller d'un concept à l'autre. Cette distance est définie par :

$$D_u(C, C) = 0 \quad [8]$$

$$D_u(C_1, C_2) = \max \left[ \begin{array}{l} D_u(\text{Sup}(C_1), C_2) + 1 \\ D_u(C_1, \text{Sup}(C_2)) + 1 \end{array} \right] \quad [9]$$

où  $\text{Sup}(X)$  est le père du concept  $X$ . Par définition, on a  $\text{sup}(\text{racine}(\text{arbre})) = \text{racine}(\text{arbre})$ . La figure 1 est un extrait de l'ontologie de [Larousse 1992]. Par exemple,



**FIG. 1.** Extrait de l'ontologie de [Larousse 1992]

nous avons  $D_u(\text{tête}, \text{jambe}) = 2$ . Tous les concepts frères de tête sont à une distance ultramétrique égale à 2. Nous avons également  $D_u(\text{tête}, \text{corps}) = 1$ ,  $D_u(\text{tête}, \text{fonctions vitales}) = 3$  et  $D_u(\text{tête}, \text{présent}) = 8$ . (Cette valeur est la plus grande possible entre deux concepts dans cette ontologie)

Munis de cette distance ultramétrique, nous pouvons construire les vecteurs génératifs. Appelons  $X_i$  le vecteur booléen correspondant au  $i$ ème concept de la base  $\mathcal{C}$ .  $Y_i$  est le vecteur conceptuel hiérarchiquement augmenté par addition de tous les vecteurs booléens  $X_j$  pondérée par un poids égal inversement proportionnel à la distance ultramétrique.

$$Y_i = X_i \oplus \bigoplus_{j=0}^{\dim(\mathcal{C})} \frac{1}{1 + D_u(C_i, C_j)} \times X_j \quad [10]$$

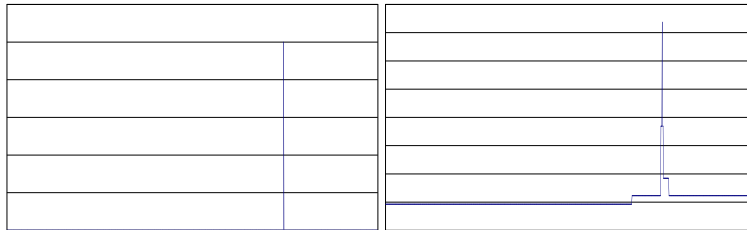
Le vecteur original  $X_i$  est ajouté à la somme normalisée des vecteurs afin que de tous les vecteurs  $Y_j$  le plus proche de  $Y_i$  soit toujours  $X_i$ . La figure 2 montre, pour *PAIX*, le vecteur booléen et le vecteur hiérarchiquement augmenté.

Bien qu'augmenter un vecteur par ses voisins améliore sa qualité, il faut admettre que la hiérarchie des concepts n'est qu'une vue particulière de la façon selon laquelle ils peuvent être s'organisé. D'autres liens spécifiques peuvent être exhiber. C'est le cas entre *CHAMPIGNON* et *TOXICITÉ* ou *GASTRONOMIE* par exemple. L'augmentation transversale d'un concept  $C$  est une opération manuelle réalisée une seule fois, à des ajustements près, qui consiste à énumérer les concepts relatifs à  $C$  qui ne sont pas représentés dans l'arbre ontologique. Le nouveau vecteur est appelé vecteur transversalement augmenté.

Par exemple, le concept *PAIX* a comme concepts transversalement associés les concepts *CONCORDE*, *GUERRE*, *CALME*, *SÉCURITÉ*, *REPOS*, *ÉQUILIBRE*. Ces concepts transversaux sont sélectionnées manuellement et peuvent être trouvés généralement dans la partie thésaurus de [Larousse 1992]

Si  $Y_i$  est le vecteur hiérarchiquement augmenté du concept  $i$  défini sur  $C$ , nous pouvons calculer le  $i$ ème vecteur augmenté transversalement  $Z_i$  en faisant la somme de tous les vecteurs  $Y_j$  avec  $Y_i$ . Cette construction assure que le vecteur  $Z_j$  le plus proche de  $Y_i$  reste  $Z_i$ .

$$Z_i = \bigoplus_{j=0}^{\dim(C)} (Y_j \oplus Y_i) \quad [11]$$



**FIG. 2.** Vecteur du concept *PAIX* et vecteur hiérarchiquement augmenté du concept *PAIX*

### 3.3.2. Noyau

Le noyau est l'indexation manuelle de mots qui est utilisée pour amorcer le système d'apprentissage. Le sens des mots est décrit par la sélection d'un nombre limité de concepts pondérés. Une telle indexation peut être trouvée dans la partie index d'un thésaurus. La description conceptuelle du  $i$ ème sens d'un mot est la liste pondérée des concepts  $\mathcal{D}_{item.i} = \{\alpha_1 c_1, \dots, \alpha_n c_n\}$ .

$$V_{item.i} = \bigoplus_j^{\mathcal{D}_{item.i}} \alpha_j Z_j \quad [12]$$

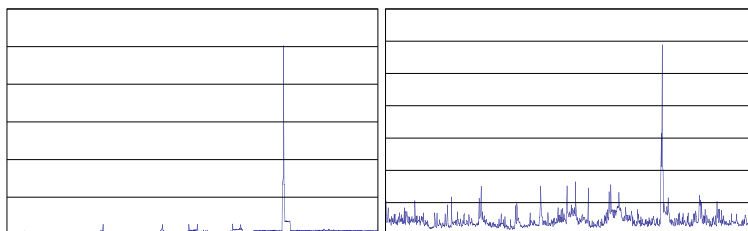


FIG. 3. Vecteurs transversalement augmenté du concept PAIX and pour l'item 'paix'

Par exemple, nous avons pour l'item lexical 'paix' l'indexation manuelle suivante :

**paix.1** [ADJ] {PAIX, 1/2 GUERRE, SECURITÉ, 1/2 ACCORD}

**paix.2** [ADJ] {REPOS, CALME, SILENCE, 1/2 ÉQUILIBRE}

Deux sens ont été indexés pour 'paix'. Le premier se réfère à une *absence de guerre* et la deuxième à une *situation de calme*. L'énumération des concepts pondérés est une tâche difficile car subjective. Seuls les mots parmi les plus importants sont ainsi décrits. Nous nous reposons sur l'apprentissage automatique pour l'indexation en masse. Toutefois, les séparations délicates de sens nécessitent des ajustements manuels.

### 3.3.3. Apprentissage automatique

L'apprentissage, à partir de définition issues de dictionnaire à usage humain, présente plusieurs difficultés telles que le métalangage utilisé (*partie de*, *se dit de*, etc) ou d'ambiguïtés que même un humain ne peut lever. C'est la cas, par exemple, de la définition de cannibale : *se dit de ceux qui mangent de la viande humaine en parlant de l'homme*. Faut-il comprendre qu'il s'agit d'humains qui mangent des humains ou de ceux qui mangent de la viande humaine tout en discourant à propos d'un être humain. Ces problèmes particuliers et des solutions sont exposés dans [Lafourcade 2001b].

L'analyse se base sur la structure morpho-syntaxique de la phrase. À partir de chaque nœud de l'arbre, qui correspondent à un item lexical constituant les définitions, sont effectués des aller-retours successifs qui permettent l'élimination d'un certain nombre d'ambiguïté lexicales. Finalement, un vecteur conceptuel est obtenu correspondant aux idées contenues dans la définition. Pour plus de détails sur la méthode d'analyse, [Lafourcade 2001c]. C'est cette analyse que les fonctions d'antonymie peuvent améliorer en affinant la valeur du vecteur conceptuel de l'item antonyme.

## 4. Fonctions d'antonymie

### 4.1. Principes et définitions

L'objectif de nos travaux est de créer une fonction qui à partir d'items lexicaux renverra les N plus proches antonymes à la manière de la fonction des N plus proches thé-

matiquement déjà développée par l'équipe. Nous l'avons vu dans la partie précédente, un antonyme s'apprécie toujours en contexte. Dans certains cas, ce contexte seul ne nous semble pas suffisant pour déterminer un axe de symétrie à l'antonymie. Prenons l'exemple de l'item lexical 'père'. Dans le contexte 'famille', il peut être opposable à 'mère' ou 'enfants'. Il peut être pertinent, dans les cas où il ne sert pas d'axe de symétrie, d'affiner le contexte par un vecteur conceptuel qui peut, lui, jouer ce rôle de référent. Dans notre exemple, il faudrait prendre alors comme référent 'filiation', 'mariage' ou 'homme'. C'est la raison pour laquelle nous avons défini la fonction d'antonymie  $AntiLex_S$  qui renvoie les  $n$  plus proches antonymes du mot  $X$  dans le contexte défini par le mot  $C$  en référence au mot  $R$  de la manière suivante :

$$w^3 * \mathbb{N} \rightarrow w^* : X, C, R, n \rightarrow Z = AntiLex_S(X, C, R, n) \quad [13]$$

où  $w$  représente l'ensemble des items lexicaux.

$$w^2 * \mathbb{N} \rightarrow w^* : X, C, n \rightarrow Z = AntiLex_R(X, C, n) = AntiLex_S(X, C, C, n) \quad [14]$$

$$w^2 * \mathbb{N} \rightarrow w^* : X, R, n \rightarrow Z = AntiLex_B(X, R, n) = AntiLex_S(X, R, R, n) \quad [15]$$

$$w * \mathbb{N} \rightarrow w^* : X, n \rightarrow Z = AntiLex_A(X, n) = AntiLex_S(X, X, X, n) \quad [16]$$

Nous avons défini la fonction partielle  $AntiLex_R$  car, dans la plupart des cas, le contexte suffit à déterminer un axe de symétrie.  $AntiLex_B$  peut être définie si on souhaite déterminer un axe de symétrie et pas un contexte. Dans la pratique, on a  $AntiLex_B = AntiLex_R$ . La dernière fonction est la fonction d'antonymie absolue. Son usage est délicat dans le cas de mots polysémiques. Il faudra admettre cependant que, dans les cas réels, elle sera la fonction la plus sollicitée. Elle va considérer le sens d'un item dans son propre contexte. Cela augmente la probabilité de ne pas obtenir une réponse satisfaisante. Ces fonctions existent bien sûr sous la forme complémentaire, scalaire et duale. La fonction globale tente de regrouper les trois formes d'antonymie. Ces trois fonctions nécessitent toutes le même parcours 4 :

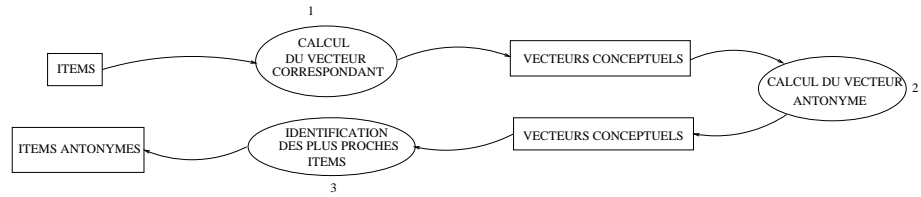


FIG. 4. parcours des fonctions  $AntiLex$

Tout d'abord, nous allons nous intéresser plus particulièrement au calcul du vecteur antonyme, c'est à dire à la fonction d'antonymie au niveau des vecteurs conceptuels, la base de ce système de traitement du langage. Elle se nomme  $Anti$  et se définit comme :

$$\vartheta^2 \rightarrow \vartheta : X, C \rightarrow Z = Anti(X, C)$$

où  $\vartheta$  représente l'ensemble des vecteurs conceptuels.

Comme nous l'avons déjà expliqué, il existe trois type d'antonymie. Nous avons donc défini trois fonctions,  $Anti_c$  pour l'antonymie complémentaire,  $Anti_s$  pour l'antonymie scalaire,  $Anti_d$  pour l'antonymie duale.

Certains items n'ont pas d'antonymes. C'est le cas, par exemple, des objets matériels comme voiture, bouteille, bateau, etc. Nous devons donc nous poser la question de la continuité des fonctions d'antonymie dans le domaine des vecteurs conceptuels. Faut-il que les fonctions soient définies pour ces mots ? Nous avons répondu favorablement à cette question. En effet, les vecteurs conceptuels sont des combinaisons linéaires de concepts. Ne pas définir un antonyme pour les concepts qui n'en possède pas serait ne pas définir d'antonymes pour l'ensemble des vecteurs qu'ils composent. Par exemple, considérons que *'enterrer'* est la composition des vecteurs de *FIN*, *TERRE* et *FUNERAILLES*. La fonction d'antonymie ne pourrait pas être définie, si l'on considère *TERRE* et *FUNERAILLES* sans antonymes. Il est pourtant clair que *'enterrer'* peut avoir comme antonyme *'déterrer'*. Comment définir l'antonyme d'un item lexical qui n'a pas d'antonyme ? Plusieurs possibilités s'offraient à nous.

- considérer que l'antonyme d'un item non opposable est  $\vec{0}$ , le vecteur nul.
- considérer que l'antonyme d'un item non opposable est tout sauf le mot lui-même. C'est à dire, en terme de vecteurs,  $\vec{1} - \vec{V}$  où  $\vec{V}$  est le vecteur correspondant à l'item et  $\vec{1}$  le vecteur correspondant à la somme de tous les vecteurs des concepts générateurs.
- considérer l'objet comme un point fixe de la fonction, c'est à dire que l'antonyme d'un item lexical qui ne possède pas d'antonyme est l'item lexical lui-même.

La première approche ne semble pas pertinente. En terme de linguistique, cela reviendrait à considérer que le contraire d'un item non opposable est l'idée vide. Cela n'est pas acceptable avec la méthode de construction que nous définirons par la suite. En effet, le contraire d'un vecteur composé de *GRAND* et de *AUTOMOBILE* ne serait composé que de *PETIT*, le concept *AUTOMOBILE* aurait disparu ce qui est incompatible avec la théorie sur l'antonymie. La deuxième méthode, est difficilement compréhensible tant du point de vue linguistique que du point de vue logique. Par exemple, le contraire de *'voiture'* serait quelque chose qui est à la fois *'gastronomie'*, *'funérailles'*, *'état'*, *'homme'*, etc. Nous avons donc retenu la dernière approche. Elle offre l'avantage que les points fixes peuvent être considérés comme étant sur l'axe de symétrie, ce qui est compatible avec l'approche théorique précédente. Dans la suite de notre exposé, nous parlerons indifféremment d'item lexical qui ne possède pas d'antonymes ou d'item lexical qui est son propre antonyme.

Comme les fonctions de synonymies (voir [Lafourcade and Prince 2001]), les diverses fonctions *Anti* sont dépendantes du contexte mais, contrairement à elles, elles ne peuvent pas être indépendantes de l'organisation des concepts. Le contexte est ici représenté par un vecteur qui sert de cadre de référence. Elles nécessitent d'identifier pour chaque concept, pour chaque contexte et pour chaque type d'antonymie un vec-



teur qui sera considéré comme son opposé. Il faut donc construire trois listes de triplets  $\langle \text{concept}, \text{contexte}, \text{vecteur} \rangle$  que nous appellerons, dans la suite, listes d'antonymes.

## 4.2. Listes d'antonymes

### 4.2.1. Notation

Le vecteur antonyme d'un concept de base sera noté à la manière des fonctions lexicales de Mel'čuk [Melch'čuk and all 95],  $AntiC_\alpha(C_i, V_{\text{contexte}})$  où  $\alpha \in \{comp, scal, dual\}$ ,  $C_i$  désigne le  $i$ ème concept de l'espace générateur<sup>6</sup> et  $V_{\text{contexte}}$  le vecteur représentant le contexte. Afin de simplifier l'écriture, l'indice  $\alpha$  sera omis lorsque nous parlons indifféremment de l'une ou l'autre des fonction *Anti*.

### 4.2.2. Construction de la liste d'antonymes

Ces vecteurs antonymes sont construits, manuellement, uniquement à partir de vecteurs conceptuels de concepts générateurs y compris le vecteur lui-même le cas échéant, puisque comme nous l'avons déjà défini, l'antonyme d'un item lexical non opposable est lui même. Ainsi, pour l'antonymie complémentaire, nous pouvons avoir par exemple :

$$\begin{aligned} AntiC_c(EXISTENCE, V) &= V(INEXISTENCE) && \forall V \\ AntiC_c(INEXISTENCE, V) &= V(EXISTENCE) && \forall V \\ AntiC_c(AGITATION, V) &= V(INERTIE) \oplus V(REPOS) && \forall V \\ AntiC_c(JOUET, V) &= V(JOUET) && \forall V \end{aligned}$$

où  $\oplus$  représente la somme normée.

Comme nous l'avons vu dans le chapitre 2, des items lexicaux peuvent avoir suivant le contexte un antonyme différent. Il en est de même pour les concepts même s'ils sont censés ne pas être polysémiques. Ainsi, *DESTRUCTION* peut avoir comme antonyme *PRÉSERVATION*, *CONSTRUCTION*, *RÉPARATION* ou *PROTECTION*. Nous avons ainsi défini pour chacun d'eux un vecteur conceptuel qui permettra la sélection de l'antonyme le mieux adapté à la situation (cf 4.2.3). Un exemple de liste d'antonymes est :

concept	contexte	vecteurs constituant le vecteur antonyme.
<i>EXISTENCE</i>	$\forall V$	$V(INEXISTENCE)$
<i>INEXISTENCE</i>	$\forall V$	$V(EXISTENCE)$
<i>AMOUR</i>	$\forall V$	$V(DÉSACCORD) \oplus V(AVERSION) \oplus V(INIMITÉ)$
<i>DÉSORDRE</i>	$V(ORDRE) \oplus V(DÉSORDRE)$	$V(ORDRE)$
<i>DÉSORDRE</i>	$V(ORDRE) \oplus V(CLASSIFICATION)$	$V(CLASSIFICATION)$
<i>DÉSORDRE</i>	$V(ORGANISATION) \oplus V(DÉSORGANISATION)$	$V(ORGANISATION)$

6. Rappelons que dans notre expérience, les concepts sont 873 et proviennent du niveau 4 de l'ontologie de [Larousse 1992].

Par exemple, le concept *EXISTENCE* a pour antonyme le vecteur *INEXISTENCE* quelque soit le contexte. Le concept *DÉSORDRE* a pour antonyme le vecteur *ORDRE* dans le contexte constitué des vecteurs *ORDRE* et *DÉSORDRE*.

La création de cette liste est un travail long et fastidieux. Sur un concept, deux personnes ne sont pas toujours d'accord sur la construction du vecteur antonyme. L'une peut considérer qu'il est formé de certains concepts tandis qu'une autre personne en considèrera d'autres ou même n'en considèrera aucun. Par exemple, l'antonyme de *HASARD* peut être *BUT* pour quelqu'un, *DÉTERMINISME* pour un autre et ne pas en avoir pour un troisième. La création de la liste est donc très subjective. C'est la partie délicate mais nécessaire de notre méthode car il s'agit de la base sur laquelle sera construit le vecteur conceptuel antonymique (cf 4.3). Il convient de préciser toutefois, qu'elle est modifiable à tout moment si des ajustements sont nécessaires par exemple si l'usage consacre de nouveaux antonymes.

#### 4.2.3. Fonction *AntiC*

La fonction *AntiC* basée sur la liste  $L_{opposés}$  construite en 4.2.2 renvoie un concept, en fonction du type d'antonymie  $\alpha$  et du contexte, son vecteur antonyme. Pour cela, elle explore la liste d'antonymes construite en 4.2 selon les algorithmes 1 et 2.

---

#### Algorithme 1: *AntiC*

---

```

Entrée : concept CONCEPT, vecteur contexte, liste  $L_{opposés}$ 
Sortie :  $V_{ant}$ 
 $l \leftarrow L.chercher(concept)$  % la méthode 'chercher' renvoie la liste des anto-
nymes possibles
si  $l.estVide()$  alors
|   retourner  $V_{concept}$  % le vecteur conceptuel du concept lui-même
sinon
|   si  $l.AUnSeulAntonyme()$  alors
|   |   retourner  $l.vecteurAntonyme()$ 
|   sinon
|   |   retourner  $meilleurChoix(l, contexte)$ 
|   |   % il y a plusieurs solutions, il faut renvoyer l'antonyme correspondant
|   |   le mieux au contexte.

```

---

L'algorithme 1 parcourt la liste d'antonyme et renvoie l'antonyme correspondant au concept et au contexte choisi. L'algorithme 2 permet simplement de choisir dans la liste l'antonyme dont le contexte est le plus proche du contexte général (cf 4.2.2). C'est à dire qu'elle calcule la distance angulaire entre le contexte général et le contexte de chaque antonyme et choisit l'antonyme dont cette distance est la plus faible. Muni de la fonction *ANTI*, nous allons maintenant pouvoir étudier la méthode de construction d'un Vecteur Antonyme.

### 4.3. Construction du vecteur antonyme : fonction *Anti*

#### 4.3.1. Définitions

Nous pouvons définir mathématiquement la fonction d'antonymie relative  $Anti_R$  :

$$\vartheta^2 \rightarrow \vartheta : X, C \rightarrow Z = Anti_R(X, C)$$

où  $\vartheta$  représente l'ensemble des vecteurs conceptuels.

La fonction d'antonymie absolue  $Anti_A$  peut être définie mathématiquement comme :

$$\vartheta \rightarrow \vartheta : X \rightarrow Z = Anti_A(X) = Anti_R(X, X)$$

$Anti_A$  est d'usage difficile car l'item lexical est considéré comme étant son propre contexte. Nous verrons dans 4.4.1 que cela pose de réels problèmes dus à la sélection de sens.

Nous allons maintenant expliciter la fonction en montrant comment à partir de deux vecteurs conceptuels, un pour l'item lexical dont nous voulons l'antonyme, l'autre pour le contexte, nous construisons le vecteur conceptuel opposé.

#### 4.3.2. Construction du vecteur conceptuel antonyme

Nous appellerons dans la suite  $V_{item}$  le vecteur dont nous souhaitons l'antonyme et  $V_c$  le vecteur constituant le contexte.

L'idée de base de la construction est d'insister sur les notions saillantes à la fois dans  $V_{item}$  et  $V_c$ . Si ces idées peuvent être opposées alors l'antonyme doit posséder les idées inverses en proportion identique. Nous en déduisons la formule :

$$Anti_R(V_{item}, V_c) = \bigoplus_{i=1}^N P_i \times Anti_C(C_i, V_c) \quad [17]$$

---

#### Algorithme 2: meilleur choix

---

**Entrée** : liste  $l$ , Vecteur *contexte*

**Sortie** :  $V_{ant}$

$min \leftarrow \frac{\pi}{2}$  % la distance entre deux vecteurs conceptuels ne peut excéder  $\frac{\pi}{2}$

**pour chaque**  $i \in l$  **faire**

**si**  $D_A(V_{contexte}, l.contexteAt(i)) \leq min$  **alors**  

 $min \leftarrow dist_{angulaire}(V_{contexte}, l.contexteAt(i))$   
 $V_{res} \leftarrow l.VecteurAt(i)$

**retourner**  $V_{res}$

---

avec  $\oplus$  qui représente la somme vectorielle normée et  $P_i$  le poids du ième concept, c'est à dire l'importance que possède ce concept à la fois dans  $V_{item}$  et  $V_c$ . En première approximation, nous posons :

$$P_i = V_{item_i} \times V_{c_i} \quad [18]$$

Nous rencontrons cependant un problème. Avec la formule (18), la fonction  $Anti_R$  est symétrique  $Anti_R(X, C) = Anti_R(C, X)$  or elle ne doit pas. L'antonyme de 'froid' par rapport à 'température' est 'chaud', mais l'antonyme de 'température' par rapport à 'froid' n'est certainement pas 'chaud'. Cela vient du fait de l'importance du contexte dans la symétrisation. Le contexte est le pivot de la symétrie. Il faut donc désymétriser cette fonction et renforcer l'influence des composantes de  $V_{item}$  sur celles de  $V_c$ . Multiplier  $V_{item_i}$  par une valeur  $\beta$  ne modifierait pas, bien sûr, ce problème de symétrie. Une solution non-linéaire intéressante est d'introduire une puissance dans l'opération. En effet, il semble pertinent de donner d'avantage de poids à une composante selon son importance.

$$P_i = V_{item_i}^\beta \times V_{c_i} \quad [19]$$

avec  $\beta \in \mathbb{R} \mid \beta > 1$ .

La valeur  $\beta$  doit donc être supérieure à 1 afin de privilégier les concepts de item sur ceux du contexte. Nous avons cherché expérimentalement quelle valeur donner à  $\beta$ . Nous nous sommes vite aperçu que lui donner une valeur trop grande favorisait déraisonnablement la composante la plus importante. Cela est approprié pour un vecteur qui n'a qu'une composante majeure<sup>7</sup> comme le vecteur d'un concept générateur mais pour d'autres, l'effet est néfaste. Dans le cas d'un vecteur qui a plusieurs composantes importantes, un  $\beta$  important appuierait uniquement sur  $C_i$  la composante la plus forte et le vecteur antonyme serait extrêmement proche de  $Anti_\alpha(c_i)$ . Par exemple, considérons que le vecteur de 'ferrari' est la combinaison linéaire des vecteurs de *AUTOMOBILE*, de *ROUGE* et de *RAPIDITÉ* avec la composante *AUTOMOBILE* relativement majoritaire. Dans le cas où  $\beta$  serait grand, on aurait  $Anti(V_{ferrari}) = V_{AUTOMOBILE}$ , si l'on considère que l'antonyme de *AUTOMOBILE* est *AUTOMOBILE*. Or, l'antonyme serait plus raisonnablement un vecteur dont les composantes importantes seraient plutôt *AUTOMOBILE*, *VERT* et *LENTEUR* ce qui pourrait correspondre, par exemple à une deux-chevaux verte.

La valeur de  $\beta$  pourrait donc être calculé en fonction des caractéristiques des composantes du vecteur. Par exemple, en utilisant le coefficient de variation de  $V_{item}$ . Le coefficient de variation  $CV(V)$  est une mesure statistique normalisée (sans unité) de la "conceptualité" du vecteur  $V$ . Il est d'autant plus important que les composantes du vecteur sont contrastés et vaut 0 si elles ont toutes la même valeur. Il est calculé en divisant l'écart-type  $ET(V)$  du vecteur par sa moyenne  $MOY(V)$ .

$$CV(V) = \frac{ET(V)}{MOY(V)} \quad [20]$$

---

7. Sont considérées comme importantes ou majeures les composantes du vecteur dont la valeur est supérieure à la moyenne de l'ensemble des composantes du vecteur.

Puisque la valeur de  $\beta$  doit être supérieure à 1, nous posons :

$$\beta = 1 + CV(V_{item}) \quad [21]$$

Nous obtenons donc, la formule suivante :

$$P_i = V_{item_i}^{1+CV(V_{item})} \times V_{c_i} \quad [22]$$

Nous remarquerons que pour le vecteur  $\vec{1}$  dont toutes les composantes sont identiques, nous sommes ramenés à  $\beta = 1$ . Dans ce cas, le contexte prend autant d'importance que l'item. Il s'agit d'un cas totalement artificiel<sup>8</sup>. Le résultat du contraire de  $\vec{1}$  serait de renvoyer un vecteur égal à celui du contexte, ce qui n'est pas le résultat escompté. Il convient de préciser toutefois que dans ce cas, l'indice de pertinence de l'antonymie, que nous détaillerons en ??, est nul. Le vecteur rendu est donc inexploitable.

Dans le cas de  $\vec{0}$ , le coefficient de variation n'est pas défini. Nous pouvons l'étendre et considérer qu'il est nul. Dans ce cas,

$$Anti_R(\vec{0}, C) = \vec{0} \quad \forall C \quad [23]$$

Expérimentalement, la formule (23) est satisfaisante dans de nombreux cas. Toutefois, il arrive que certains résultats soient catastrophiques. Ainsi,  $V(\text{'froid'})$  dans le contexte  $V(\text{'température'})$  ne nous rend pas un vecteur proche de  $V(\text{'chaud'})$ . Le problème vient du fait que le contexte n'allume pas les mêmes idées que  $V(\text{'froid'})$ . La faiblesse du concept de *FROID* dans  $V(\text{'température'})$  entraîne l'écrasement de la composante *'chaleur'* dans le vecteur conceptuel antonyme par rapport aux autres composantes. D'où la nécessité de conserver les idées importantes de  $V_{item}$  quelque soit le contexte (le contexte ne servant plus alors qu'à insister sur les idées qui lui sont plus fortes).

La formule actuelle de construction du vecteur conceptuel antonyme se base sur les notions saillantes à la fois dans  $V_{item}$  et  $V_c$  tout en ne négligeant pas celles qui sont importantes seulement dans  $V_{item}$ . Si elles peuvent être opposées alors l'antonyme doit posséder l'idée inverse en proportion identique. D'où la redéfinition du poids :

$$P_i = V_{item_i}^{1+CV(V_{item})} \times \max(V_{item_i}, V_{c_i}) \quad [24]$$

En fait, nous ajoutons proportionnellement à son importance dans  $V_{item}$  et  $V_c$  chaque concept de base. Nous construisons ainsi un vecteur conceptuel qui possède l'inverse des propriétés de  $V_{item}$ .

Pour résumer, le vecteur antonyme  $Anti_R$  de  $V_{item}$  dans le contexte  $V_c$  est,

$$Anti_R(V_{item}, V_c) = \bigoplus_{i=1}^N P_i AntiC(C_i, V_c)$$

$$P_i = V_{item_i}^{1+CV(V_{item})} \times \max(V_{item_i}, V_{c_i})$$

8. ce vecteur n'est certainement pas représentable dans une langue, il signifierait tout et son contraire.

Nous présenterons dans la partie 4.4.2 des exemples de résultats obtenus à partir de ces formules. Nous allons maintenant exposer le passage 1 entre les items lexicaux et les vecteurs .

#### 4.4. Items lexicaux et vecteurs : problèmes et solutions

L’objectif des fonctions *AntiLex* est, à partir d’un item lexical, de donner son antonyme. Il faut utiliser la fonction :  $Anti : \vartheta_c^2 \rightarrow \vartheta_c$  définie en 4.3. Nous devons donc utiliser des outils permettant le passage des items lexicaux vers les vecteurs conceptuels. Ce passage pose de nombreux problèmes dû à la polysémie c’est à dire choisir la bonne correspondance entre un item lexical et un vecteur conceptuel, choisir le bon sens du mot.

##### 4.4.1. Passage items lexicaux $\rightarrow$ Vecteurs conceptuels

Nous l’avons vu, notre définition de l’antonymie est relative à un contexte. Dans certains cas, ce contexte seul ne suffit pas à déterminer un axe de symétrie pour l’antonymie. Prenons l’exemple de l’item lexical ‘*père*’. Dans le contexte ‘*famille*’, il peut être opposable à ‘*mère*’ ou ‘*enfants*’. Il peut être pertinent, dans les cas où il ne sert pas d’axe de symétrie, d’affiner le contexte par un vecteur conceptuel qui peut, lui, jouer ce rôle de référent. Dans notre exemple, il faudrait prendre alors comme référent ‘*filiation*’, ‘*mariage*’ ou ‘*homme*’.

La fonction *AntiLex<sub>S</sub>* prend trois items lexicaux en arguments : l’item lexical *item* dont on cherche l’antonyme et l’item lexical ‘*contexte*’ qui définit le contexte qui nous permet de sélectionner le sens de *item* plus particulièrement recherché. Par exemple, le mot ‘*femme*’ peut avoir comme acception épouse ou bien personne du sexe féminin. Dans le premier cas, un antonyme pourrait être ‘*mari*’, dans le second ‘*homme*’. Le troisième item sert d’axe de symétrie. Plusieurs approches ont été tentées pour réaliser la sélection du sens de l’item et, s’il est différent du contexte, réaliser la sélection du sens du référent :

- utiliser la définition<sup>9</sup> la plus proche du contexte. Cette solution, bien quelle gomme complètement l’effet de la polysémie, est beaucoup trop brutale. Elle ne va considérer qu’un seul vecteur (le plus proche du contexte). Avoir plusieurs sources d’information vient de la volonté d’affiner les sens des items lexicaux, de les compléter et non de les mettre en opposition. Cette méthode ne semble pas adéquate car elle se base sur une confusion importante : confondre les sèmes et les définitions.

- contextualiser le vecteur, c’est à dire obtenir via la méthode de mise en contextualisation forte  $\Gamma : w^2 \rightarrow w$  [Lafourcade 2001b] les vecteurs conceptuels paramètres de *Anti*.

9. Chaque item lexical a plusieurs définitions provenant soit de dictionnaires différents soit du fait même de la polysémie. A chacune de ces définitions correspond un vecteur conceptuel. Le vecteur conceptuel correspondant à l’item lexical est une somme pondérée de ces vecteurs. cf [Lafourcade 2001b] pour plus de détail.

Cette méthode a l'avantage de ne pas omettre un sens : tous sont présents dans le vecteur résultat, mais le contexte permet de privilégier certains sens au détriment d'autres. Le principe général consiste à classer les différents vecteurs constitués par les définitions en les pondérant proportionnellement à leur distance à un contexte. Un contexte thématique peut prendre la forme d'un vecteur conceptuel, le poids étant en proportion inverse à la distance. D'autres contextes peuvent également se rajouter : contexte morphologique, fréquence d'utilisation, etc. On peut se demander pourquoi le vecteur contexte est aussi contextualisé. Une fois de plus, le problème de la polysémie, est à l'origine de la nécessité de contextualiser le vecteur contexte. Cette contextualisation permettra de favoriser le sens le plus proche de '*item*' dont on souhaite l'antonyme et donc donnera un meilleur vecteur conceptuel qui sera utilisé pour la construction du vecteur conceptuel antonyme (cf 4.3.2). Par exemple, si on cherche l'antonyme du mot '*frégate*' dans le contexte de '*bateau*'. Le mot '*frégate*' peut être un oiseau ou justement ce que l'on cherche, un bateau. Le mot bateau lui aussi est polysémique, il peut prendre le sens de '*facile*', de '*dépression de trottoir devant une porte*' ou bien le sens qui nous intéresse dans le cas présent, '*navire*'. La contextualisation forte va insister sur le coté '*navire*' de '*frégate*' et '*bateau*' et nous fournira par ce biais des vecteurs mieux exploitables. On peut envisager ici une contextualisation itérée qui permettrait un plus grand affinement du sens (cf [Lafourcade 2001b]).

L'algorithme 3 montre l'utilisation de la contextualisation forte pour le calcul du vecteur antonyme.

Cette mise en contexte, montre le problème de la fonction  $Anti_{\alpha_R}$  d'antonymie absolue. Dans ce cas, la méthode va calculer le vecteur du mot '*item*' dans le contexte '*item*'. Ce n'est pas un problème dans le cas d'un mot qui n'a qu'une définition dans la base puisque la contextualisation forte est sans effet sur le vecteur. Dans le cas contraire, le vecteur conceptuel rendu, appuiera plus sur le sens le plus présent dans le vecteur d'origine, sens qui n'est pas forcément le sens souhaité.

#### 4.4.2. Passage Vecteurs conceptuels $\rightarrow$ items lexicaux

Ce passage est, a priori, plus aisé. Il suffit de comparer le vecteur conceptuel antonyme  $V_{ant}$  aux vecteurs conceptuels de la base. Les plus proches, au sens de la distance angulaire, c'est à dire ceux qui sont les plus synonymes de  $V_{ant}$ , seront en antonymie thématique de  $V_{item}$  (cf [Lafourcade and Prince 2001]). Avec notre méthode, nous obtenons, par exemple,

$$- Anti_{c_R}(\text{mort}, \text{'mort'} \text{ et } \text{'vie'}) = (VIE[0.4]) \quad (ASSASSINEUR[0.4493]) \\ (HASCHISCHIN[0.4671]) \quad (ASSASSIN[0.4673]) \quad (SUCEUR \quad DE \quad SANG[0.4716]) \\ (STRIGE[0.4716])(PAMPHILE[0.4763]) \quad (MOUCHE[0.4764])(VITAL[0.4765]) \quad (SE \quad SUF- \\ FIRE[0.4825]) \quad (MOURIR[0.4844])(\text{ÊTRE ANIMÉ}[0.4858]) \quad (VIVRE[0.486])(ISSANT[0.4891]) \\ (JEUNET[0.4891]) \quad (JEUNOT[0.4892]) \quad (JEUNE[0.4894]) \quad (NAISSANT[0.4895])$$

$$- Anti_{c_R}(\text{vie}, \text{'mort'} \text{ et } \text{'vie'}) = (MORT :N[0.3367]) \quad (\text{ÊTRE GUÉRI DE TOUS LES} \\ \text{MAUX}[0.3573]) \quad (C4 :MORT[0.3573]) \quad (HASCHISCHIN[0.3672]) \quad (ASSASSIN[0.3675]) \quad (ASSAS- \\ SINEUR[0.3774]) \quad (MORT :ADJOINT[0.407]) \quad (C3 :LES ÂGES DE LA VIE[0.481]) \quad (TYRANNI-$$

*CIDE*[0.5161]) (*ASSASSINER*[0.5797]) (*MORT*[0.5821])

–  $AntiC_{c_A}(VIE) = (C4 :MORT[0.0344]) (ÊTRE\ GUÉRI\ DE\ TOUS\ LES\ MAUX[0.0344])$   
 $(MORT :N[0.4272]) (MORT :ADJOINT[0.4747]) (C3 :LES\ AGES\ DE\ LA\ VIE[0.551]) (HASCHISCHIN[0.568])$   
 $(ASSASSIN[0.5683]) (ASSASSINEUR[0.5885]) (MORTEL :ADJOINT[0.6631]) (TYRANNICIDE[0.6991])$   
 $(C2 :L'ÊTRE\ HUMAIN[0.7357]) (ASSASSINER[0.7485]) (MORT[0.7704])$

Tout d’abord, rappelons qu’il est inutile de mettre le concept  $vie_c$  en contexte puisque il fait parti des concepts générateurs. L’item lexical le plus proche du vecteur antonyme est le concept  $mort_c$ , ce qui est très acceptable en antonymie complémentaire (cf chapitre 2). On peut s’apercevoir que la distance entre le vecteur antonyme calculé et le vecteur antonyme réel n’est pas nulle. Ce phénomène est explicable par deux facteurs :

– la construction des vecteurs : les vecteurs générateurs ne sont pas indépendants les uns des autres. Leur construction est basée sur l’ontologie. De plus, ils se renforcent les un, les autres (voir 3.3). Avec des vecteurs booléens, où seule la composante correspondant au concept ne serait pas nulle, la fonction *Anti* rendrait exactement le vecteur antonyme.

– le bruit dû à la méthode de construction : nous travaillons en logique floue, la méthode de construction d’un vecteur antonyme, basée sur des définitions, n’est pas

---

### Algorithme 3: items vers vecteur antonyme

---

**Entrée** : item, contexte, référent

**Sortie** : vecteur antonyme

**si** *contexte.estVide()*  $\wedge$  *référent.estVide()* **alors**

    % cas de l’antonymie absolue  
     $V_{item} = vecteur(item)$   
     $V_{cont} = V_{item}$   
    **retourner**  $Anti_R(V_{item}, V_{cont})$

**si** *référent.estVide()* **alors**

    % on considère que l’axe de symétrie du contexte est suffisant  
     $V_{item} = \Gamma(item, contexte)$   
     $V_{cont} = \Gamma(contexte, item)$   
    **retourner**  $Anti_R(V_{item}, V_{cont})$

**si** *contexte.estVide()* **alors**

    % cas de “sécurité” la logique voudrait que l’utilisateur sélectionne un  
    contexte plutôt qu’un référent  $V_{item} = \Gamma(item, référent)$   
     $V_{cont} = \Gamma(référent, item)$   
    **retourner**  $Anti_R(V_{item}, V_{cont})$

**sinon**

    % cas où les trois sont définis  
     $V_{item} = \Gamma(item, contexte)$   $V_{cont} = \Gamma(référent, contexte)$

---



d'une précision absolue. Elle a pour but de construire un vecteur s'approchant le plus possible du vecteur antonyme mais il est impossible de le construire exactement.

Plus le mot est polysémique, plus ce phénomène se retrouve comme en témoignent les deux premiers exemples  $Anti_{c_R}$  (mort, 'mort' et 'vie') où l'antonyme est à 0.4 et  $Anti_{c_R}$  (vie, 'mort' et 'vie') dont l'antonyme est à 0.3 radians du vecteur calculé.

On ne peut considérer comme antonyme, même si la mesure d'antonymisabilité est correcte, l'item lexical le plus proche du vecteur antonyme. Pour simplifier, on peut dire qu'un antonyme doit avoir la même catégorie morphologique que 'item'. Si on cherche l'antonyme d'un verbe, le résultat le plus pertinent sera de rendre un verbe. Éventuellement, certaines variantes peuvent être considérées. Ainsi l'antonyme d'un adjectif pourrait être un adverbe. Par exemple, le contraire de vite peut être 'rapide' ou 'rapidement'.

La technique de recherche des items les plus proches du vecteur antonyme souffre, elle aussi, du problème de la polysémie. En effet, la proximité thématique entre le vecteur antonyme et les mots de la base ne considère pas le contexte. Si nous cherchons l'antonyme de 'chaud' dans le contexte de température, le vecteur conceptuel rendu par la fonction  $Anti_{\alpha}$  sera proche de 'froid' mais pas du terme 'froid' de la base, formé des vecteurs des différents sous-sens de 'froid', mais du sous-sens de froid correspondant à la température. Un système qui fasse la proximité thématique en contexte est très facilement réalisable mais extrêmement coûteux en temps à exécuter. Cela nécessiterait pour chaque terme de le mettre en contextualisation puis de calculer la distance de son vecteur contextualisé au vecteur antonyme. Faire cela pour chaque vecteur prendrait à l'heure actuelle des heures. Sur un iMac, cadencé à 300 Mhz, la fonction de contextualisation forte met environs une seconde à se réaliser pour un mot possédant quatre définitions ce qui est légèrement inférieur à la moyenne de définitions pour un item à l'heure actuelle. La base comporte environs 60000 items. Cette technique mettrait donc environs  $60000/3600 = 16.66$  heures. La technique exposée ci-dessus est une bonne approximation de l'antonymie.

On pourrait imaginer de combiner les deux techniques, c'est à dire de chercher les n plus proches item, et de faire une contextualisation sur chacun d'eux avant de calculer leur distance au vecteur antonyme. Cela n'est qu'une heuristique, elle ne donnera pas la bonne solution si, par exemple, il avait fallu prendre le n+1ème item. Ce cas est certainement fréquent, dans certaines parties de l'espace, la densité lexicale étant beaucoup plus forte que dans d'autres. Nous en avons un bon exemple ci-avant avec  $Anti_{c_R}$  ('vie', 'mort' et 'vie') dont les voisins sont proches et  $Anti_{c_A}$  (VIE) dont les voisins sont beaucoup plus éloignés.

Nous avons montré une méthode de construction des vecteurs conceptuels dans le cas des antonymies complémentaire, scalaires et duales. Nous allons, dans la partie suivante, définir la fonction d'antonymie générale qui se base sur une mesure, la mesure d'antonymie potentielle  $Manti_{Pot}$ , qui permet de déterminer si un vecteur est susceptible de posséder un antonyme.

## 5. Mesures et fonction d'antonymie générale.

### 5.1. Mesures d'antonymie potentielle

#### 5.1.1. Définitions

L'objectif premier de cette mesure est d'éviter le calcul du vecteur antonyme d'un vecteur qui ne doit pas en avoir. La mesure d'antonymie potentielle  $M_{ant_\alpha}$  doit permettre de savoir, dans quelle mesure, un vecteur conceptuel, est susceptible de posséder au moins un antonyme. Il s'agit d'une mesure à priori.

$$\vartheta \rightarrow \mathbb{R} : X \rightarrow Z = M_{ant_\alpha}(X) \quad [25]$$

La mesure est basé sur l'idée qu'un vecteur possède un antonyme si ses idées importantes peuvent être inversées et n'en possède pas dans le cas contraire. Le calcul de la mesure est réalisé par l'algorithme 4.

---

#### Algorithme 4: Mesure d'antonymie potentielle

---

**Entrée :** vecteur  $V_{item}$

**Sortie :**  $M_{ant_\alpha} \in \mathbb{R}$

$d_{anto} \leftarrow 0$

$d_{nonAnto} \leftarrow 0$

$moy \leftarrow moyenne(V_{item})$  % 'moyenne' calcule la moyenne des composantes du vecteur.

$cv \leftarrow CV(V_{item})$  % 'CV' calcule le coefficient de variation du vecteur.

**pour chaque concepts  $i$  faire**

$E = V_{item_i} - moy$

**si**  $Anti(C_i) \neq V_{C_i}$  **alors**

% le concept a un antonyme différent de lui-même.

$d_{anto} \leftarrow d_{anto} + max(0, sgn(E) \times |E|^{cv})$  % ce cas est favorable à l'antonymie.

%  $sgn(E)$  renvoie le signe de E.

$d_{nonAnto} \leftarrow d_{nonAnto} - min(0, sgn(E) \times |E|^{cv})$  % ce cas est défavorable à l'antonymie.

**sinon**

% le concept n'a pas antonyme.

$d_{nonanto} \leftarrow d_{nonAnto} - min(0, sgn(E) \times |E|^{cv})$  % ce cas est défavorable à l'antonymie.

$d_{anto} \leftarrow d_{anto} + max(0, sgn(E) \times |E|^{cv})$  % ce cas est favorable à l'antonymie.

**retourner**  $calcul(d_{anto}, d_{nonanto})$

---

Le principe de l'algorithme 4 est de considérer l'importance des idées suivant l'écart à la moyenne des composantes. Si un concept qui possède un antonyme est

**Algorithme 5:** Calcul de la mesure d'antonymie potentielle

---

```

Entrée :  $d_{anto}, d_{nonanto}$ 
Sortie :  $M_{ant_\alpha} \in \mathbb{R}$ 
si  $d_{nonAnto} \neq 0$  alors
  | si  $d_{anto} = 0$  alors
  | | retourner  $-\infty$  % il n'y a aucune chance que le vecteur ait un anto-
  | | nyme
  | sinon
  | | retourner  $\log(d_{anto}/d_{nonAnto})$  % cas général
sinon
  | si  $d_{anto} = 0$  alors
  | | retourner  $0$  % on est dans le cas d'un vecteur plat, il n'a pas d'anto-
  | | nyme
  | sinon
  | | retourner  $+\infty$  % on est sûr que vecteur est antonymable

```

---

une idée importante il doit renforcer l'indice, sinon il doit l'affaiblir. L'inverse est appliqué si le concept ne possède pas d'antonyme. l'écart à la moyenne est renforcé par une mise à la puissance 'cv' qui correspond au coefficient de variation du vecteur. Les écart à la moyenne doivent avoir d'autant plus d'influence qu'il y a des écarts importants entre les composantes. En effet, la composante la plus importante du vecteur d'un concept générateur doit avoir plus de poids que la composante importante d'un terme quelconque. A priori,  $M_{ant_\alpha}$  devrait être très facilement interprétable, si  $M_{ant_\alpha} \leq 0$  le vecteur n'a pas d'antonyme et si  $M_{ant_\alpha} > 0$  le vecteur possède un antonyme.

L'expérimentation nous amène à nuancer cette vision des choses. On peut considérer trois zones définies empiriquement :

- si  $M_{ant_\alpha} < 0.4$  le vecteur n'a pas d'antonyme.
- si  $M_{ant_\alpha} > 0.4$  le vecteur possède un antonyme.
- si  $M_{ant_\alpha} \in [-0.4, 0[ \cup ]0, 0.4]$  il s'agit d'une zone de flou où on trouve des mots sans antonymes dont la note est positive et des mots avec antonyme dont la note est négative. Précisons que cette note négative n'empêche pas la fonction *Anti* de renvoyer un vecteur acceptable (mais par l'humain pas par la machine).
- le cas  $M_{ant_\alpha} = 0$  est particulier. Il ne peut survenir que si aucune composante correspondant à un concept qui possède un antonyme n'est supérieure à la moyenne. On peut donc en déduire que, dans ce cas, le vecteur n'a pas d'antonyme.

## 5.1.2. Exemples

Nous avons les résultats suivants :

$M_{ant_c}(V_{existence})$	= +0.82	$M_{ant_c}(V_{femme})$	= +0.54
$M_{ant_s}(V_{existence})$	= -1.47	$M_{ant_c}(V_{FEMME})$	= +3.79
$M_{ant_d}(V_{existence})$	= -1.33	$M_{ant_c}(V_{automobile})$	= -2.15
$M_{ant_c}(V_{EXISTENCE})$	= +2.81	$M_{ant_c}(V_{AUTOMOBILE})$	= -8.16
$M_{ant_s}(V_{EXISTENCE})$	= -6.02	$M_{ant_d}(\gamma(père, homme))$	= -0.14

avec  $\gamma$  qui calcule la contextualisation forte de ‘père’ dans le contexte ‘homme’. Nous voyons que le mot ‘existence’ possède un antonyme complémentaire mais ni antonyme scalaire ni antonyme dual. Le concept ‘existence’ obtient une mesure plus importante que le mot en antonymie complémentaire. Ce phénomène est dû à la polysémie des mots. Nous retrouvons ce phénomène qui agit de manière inverse avec l’antonymie scalaire, le concept obtient dans ce cas une moins bonne note que le mot. Nous traiterons ce problème dans la partie 4.4.1. On peut remarquer le même phénomène avec ‘femme’ qui possède un antonyme et ‘automobile’ qui lui n’en possède pas. Le cas de ‘père’ est très intéressant. La mesure  $M_{ant_d}$  du vecteur de ‘père’ dans le contexte ‘homme’ est négative. Nous nous trouvons dans le cas évoqué précédemment où la mesure est comprise entre -0.4 et 0. En effet, le vecteur antonyme rendu par la fonction est proche du vecteur de mère (la distance est de 0.3, cf 4.4.2) ce qui semble une réponse très acceptable.

## 5.2. Fonction d’antonymie globale

Il s’agit d’essayer de simuler la fonction d’antonymie classique dont les gens ont l’intuition. En effet, seules les personnes qui ont réellement étudié le sujet savent qu’il existe plusieurs types d’antonymie. Pour toutes les autres et même pour les dictionnaires, elles sont confondues. Nous avons donc voulu réaliser une méthode qui renvoie les antonymes d’un item. La solution qui consiste à faire la somme des vecteurs résultant n’est pas satisfaisante. En effet, ‘action’ a pour antonyme complémentaire  $V_{inaction}$ ,  $V_{réaction}$  en antonymie duale et lui même en antonymie scalaire. Faire la somme des trois vecteurs donnerait un vecteur dont la direction ne semble pas clairement correspondre aux antonymes. Évidemment, ceci est d’autant plus vrai que le terme est polysémique. Nous avons donc préféré une méthode qui n’intervient qu’après le calcul des trois vecteurs et la recherche de leur plus proches voisins. Elle se base aussi sur la mesure d’antonymie potentielle qui renforce les antonymes. Par exemple, si nous cherchons les antonymes de *CHALEUR* dans le contexte ‘température’ en référence à *CHALEUR*, nous obtenons :

– En antonymie complémentaire : (*FROID* 0.0341) (‘solidifié’ 0.3213) (c3 :propriétés de la matière 0.4647) (‘algide’ 0.6342) (‘froid’ 0.6345) (*HUMIDITÉ* 0.6534) (‘station de pompage’ 0.657) (*SÉCHERESSE* 0.657) (‘calorimétrique’ 0.6611) (*CHALEUR* 0.6611) (chaud :N 0.6611) (conducteur :ADJOINT 0.6611) (*COMBUSTIBILITÉ* 0.6629) (*SOLIDITÉ* 0.6637) ...

avec  $M_{antiPot_c} = 3.02$

– En antonymie scalaire : (*FROID* 0.0338) (*solidifié* 0.3209) (*c3 :propriétés de la matière* 0.4644) (*algide* 0.6353) (*froid* 0.6356) (*station de pompage* 0.6548) (*SÉCHERESSE* 0.6548) (*HUMIDITÉ* 0.6553) (*calorimétrique* 0.661) (*CHALEUR* 0.661) (*chaud :N* 0.661) (*conducteur :ADJOINT* 0.661) (*COMBUSTIBILITÉ* 0.662) ...

avec  $Manti_{Pot_d} = 2.727$

– En antonymie duale : (*calorimétrique* 0.0317) (*CHALEUR* 0.0317) (*chaud :N* 0.0317) (*conducteur :ADJOINT* 0.0317) (*c3 :propriétés de la matière* 0.4237) (*solidifié* 0.5624) (*s'atomiser* 0.5864) (*se vaporiser* 0.5866) (*thermo-* 0.6065) (*station de pompage* 0.6183) (*SÉCHERESSE* 0.6183) (*HUMIDITÉ* 0.6236) (*COMBUSTIBILITÉ* 0.627) ...

avec  $Manti_{Pot_s} = -6.326$

L'idée est d'interclasser les items obtenus dans les trois listes de couples (item, distance) en fonction de la pondération P. Ce poids est calculé grâce à  $Manti_{Pot}$  et à des distances aux vecteurs antonymes dans chacune des trois listes.

---

**Algorithme 6:** fonction d'antonymie générale
 

---

**Entrée :**  $L_{comp}, Manti_{Pot_c}, L_{scal}, Manti_{Pot_d}, L_{dual}, Manti_{Pot_s}$

**Sortie :** liste des antonymes triée

$associations \leftarrow \emptyset$  **pour chaque** item  $i$  **présent au moins dans une des listes**

**faire**

$\sqcup$  associer( $i$ , poids( $i$ ))

**retourner**  $classement(associations)$

*% la fonction 'classement' trie les associations suivant leur poids*

---



---

**Algorithme 7:** calcul de poids d'un item
 

---

**Entrée :** item  $i$ ,  $L_{comp}, Manti_{Pot_c}, L_{scal}, Manti_{Pot_d}, L_{dual}, Manti_{Pot_s}$

**Sortie :** poids de l'item  $i$

$poids \leftarrow distance(i, L_{comp}) \times e^{Manti_{Pot_c}}$

$poids \leftarrow poids + distance(i, L_{scal}) \times e^{Manti_{Pot_s}}$

$poids \leftarrow poids + distance(i, L_{dual}) \times e^{Manti_{Pot_d}}$

**retourner**  $poids \times 1 / (e^{Manti_{Pot_c}} + e^{Manti_{Pot_d}} + e^{Manti_{Pot_s}})$

---

L'algorithme 6 renvoie une liste des items triée en fonction de leurs poids respectifs rangés par ordre croissant.  $Manti_{Eval}$  est donc une moyenne pondérée par les mesures de potentialité des distances.

Dans la pratique, les listes ne sont pas totale, c'est à dire que nous ne récupérons pas, pour des raisons d'espace (la base comporte actuellement près de 60000 items) l'ensemble des items ordonnés par distance mais seulement les  $n$  premiers. Nous avons donc légèrement modifié l'algorithme 7 afin que si l'item n'est pas présent dans une liste on considère sa distance comme égale à la dernière de la liste. Nous avons choisi cette solution car nous considérons que l'item recherché peut se situer à la  $n+1$ ème

place dans la liste. Dans l'exemple précédent, nous obtenons : (*FROID* 0.034) (*solidifié* 0.32) (*c3 :propriétés de la matière* 0.4645) (*algide* 0.635) (*froid* 0.6349) (*HUMIDITÉ* 0.6542) (*station de pompage* 0.656)

### 5.3. Mesure d'évaluation de l'antonymie

Il nous a semblé pertinent de pouvoir savoir si deux items lexicaux pouvaient être l'antonyme l'un de l'autre afin de posséder un outil comparable à la synonymie relative (cf. [Lafourcade and Prince 2001]). Nous avons, pour apporter une réponse à ce problème, créé une mesure d'évaluation de l'antonymie .

Soit les items lexicaux  $V_{I1}$  et  $V_{I2}$ . La question est de savoir dans quelle mesure on peut dire qu'ils sont antonymes dans le contexte  $V_c$ . La distance d'antonymie  $M_{ant_\alpha}$  est l'angle formé par la somme par les vecteurs  $V_{I1}$  et  $V_{I2}$  et la somme de leur opposés  $Anti_{cR}(V_{I1}, V_c)$  et  $Anti_{cR}(V_{I2}, V_c)$ . Soit :

$$\begin{aligned} \vartheta_c^3 \rightarrow R : \quad & V_{I1}, V_{I2}, V_c \rightarrow M_{ant} \\ M_{ant_\alpha} = & D_A(V_{I1} \oplus V_{I2}, Anti_R(V_{I1}, V_c) \oplus Anti_R(V_{I2}, V_c)) \end{aligned}$$

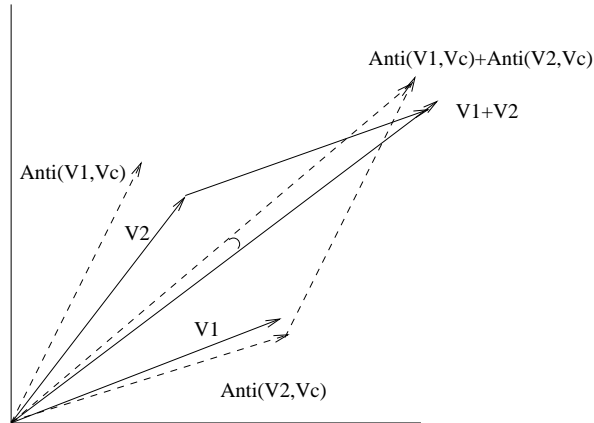


FIG. 5. mesure d'évaluation de l'antonymie

La mesure d'antonymie n'est pas une distance. Ce n'est qu'une pseudo-distance. Elle vérifie les propriétés de réflexivité, symétrie et inégalité triangulaire uniquement dans le sous ensemble des items qui n'ont pas d'antonymes où elle est, au bruit près, égale à la distance angulaire. Dans le cas général, elle ne vérifie pas la réflexivité. Les composantes des vecteurs conceptuels sont positives, nous avons la propriété :  $Dist_{anti} \in [0, \frac{\pi}{2}]$  Plus la mesure est petite, plus les deux items lexicaux sont antonymes dans le contexte. Par contre ce serait une erreur de considérer que deux synonymes seraient à une distance avoisinant  $\pi/2$ . Deux items lexicaux à  $M_{ant} = \pi/2$

l'un de l'autre n'ont aucune idée en commun<sup>10</sup>. Nous pouvons plutôt voir ici l'illustration que deux antonymes ont certaines idées en communs, celles qui ne sont pas opposable ou celles qui le sont mais dont l'activation est proche. Ils ne s'opposent que par certaines activations de concepts. Une distance de  $\pi/2$  entre deux items lexicaux devrait être plutôt interprété comme le fait que ces deux items lexicaux n'ont que peu d'idées en commun, une sorte d'anti-synonymie. Ce résultat confirme le fait que l'antonymie n'est pas exactement l'inverse de la synonymie mais lui est très liée. L'antonyme d'un item '*m*' n'est pas un mot qui ne partage aucune idée avec '*m*' mais un item qui s'oppose à '*m*' sur certaine idées !

### 5.3.1. Exemples

Dans les exemples qui suivent le contexte a été quelques fois omis. Dans ces cas, le contexte est constitué par la somme des vecteurs des deux items.

$M_{ant_c}(EXISTENCE, INEXISTENCE)$	= 0.032	$M_{ant_s}('existence', 'existence')$	= 0.58
$M_{ant_s}(EXISTENCE, INEXISTENCE)$	= 0.67	$M_{ant_c}(AUTOMOBILE, AUTOMOBILE)$	= 0.006
$M_{ant_d}(EXISTENCE, INEXISTENCE)$	= 0.67	$M_{ant_c}('automobile', 'automobile')$	= 0.407
$M_{ant_c}('existence', 'inexistence')$	= 0.44	$M_{ant_c}('existence', 'automobile')$	= 1.06
$M_{ant_c}('existence', 'existence')$	= 0.58	$M_{ant_c}(EXISTENCE, AUTOMOBILE)$	= 1.45

Les exemples ci-dessus illustre bien ce que nous disions auparavant. *EXISTENCE* et *INEXISTENCE* sont très fortement antonymes en antonymie complémentaire. Par contre, il le sont moins en antonymie scalaire et duale. L'effet de la polysémie explique que les items '*existence*' et '*inexistence*' sont moins antonymes que les concepts. En antonymie complémentaire, *AUTOMOBILE* est son propre antonyme. La mesure de l'antonymie entre *AUTOMOBILE* et *EXISTENCE* montre bien ce que nous expliquions auparavant. Aux alentours de  $\frac{\pi}{2}$ , cette mesure se comporte comme la distance angulaire. D'ailleurs, nous avons  $D_A(existence, automobile) = 1.464$ . On pourrait envisager d'utiliser cette fonction pour chercher dans le lexique conceptuel les meilleurs antonymes mais le coût en temps (environs une minute pour une mesure) serait prohibitif.

## 6. Conclusions et perspectives

Ce mémoire présente une modélisation et une réalisation informatique de l'antonymie à l'aide de vecteurs conceptuels. Dans le cadre du TALN, l'antonymie est un aspect fondamental dès que l'on cherche à évaluer et représenter la sémantique de segments textuels. Les applications majeures en sont l'analyse thématique de textes et la construction de grandes bases lexicales. Nous nous sommes basés sur une théorie linguistique calculable ainsi que vectoriellement représentable et nous avons mené notre analyse dans la perspective de la symétrie. Ce travail préliminaire nous a permis d'exprimer en termes de vecteurs conceptuels diverses fonctions d'antonymie, à

10. ce cas de figure est purement théorique, il n'existe dans aucune langue deux items lexicaux qui ne partagent aucune idée, ils partagent aux moins l'idée d'être des item lexicaux.

savoir les antonymies complémentaire, scalaire et duale. Ces fonctions permettent, à partir d'un vecteur conceptuel et d'informations contextuelles, de calculer un vecteur antonyme. Des extensions ont aussi été réalisées pour que ces fonctions soient définies et utilisables à partir des items lexicaux. Des mesures ont été introduites : l'une permet d'évaluer si un vecteur est susceptible de posséder un antonyme, l'autre d'apprécier si deux items peuvent être antonymes l'un de l'autre. Ces mesures d'antonymie sont à la base de la détection des phénomènes d'opposition dans les textes. La construction de vecteurs, elle, est nécessaire pour la sélection des items lexicaux contraires dans le cas de la génération de textes.

De nombreuses améliorations sont possibles à commencer par la révision des listes d'antonymes. Elles n'ont été actuellement construites que par une seule personne et il est souhaitable que d'autres, en particulier des linguistes, les affinent. L'utilisation de ces fonctions mettra certainement en évidence des problèmes que nous n'avons pas encore rencontrés. Des corrections devront probablement être apportées à nos fonctions et mesures pour qu'elles soient plus précises. Nous pensons en particulier à la mesure de potentiel d'antonymie dont les valeurs autour de 0 ne sont pas exploitables. Une piste possible est celle de l'automatisation de ces ajustements par apprentissage sur des corpus.

Un des objectifs des travaux exposés dans cet article est d'améliorer le système d'apprentissage des vecteurs conceptuels en permettant la prise en compte, dans l'analyse thématique, de la négation. Dans un cadre plus général, il reste à étudier de nombreuses fonctions lexicales [Melch'čuk and all 95] couramment utilisées. Le processus d'analyse thématique peut être globalement amélioré via l'identification dans les textes (et les définitions) des phénomènes de sémantique lexicale que décrivent ces fonctions. Il reste à les implémenter dans la perspective des vecteurs conceptuels pour améliorer l'ensemble de la chaîne d'analyse. Ces fonctions, environ une soixantaine (toutes ne sont manifestement pas calculables), permettent de caractériser les comportements linguistiques liés à l'usage. L'équipe a déjà commencé à étudier la synonymie (mots de même sens) [Lafourcade and Prince 2001] et la publication de ce travail sur l'antonymie [?] en constitue une suite logique. D'autres fonctions lexicales, non moins importantes, doivent être considérées. Parmi elles citons, par exemple, l'hyponymie qui est une relation du type *sorte de* (la pomme est un fruit, la frégate est un oiseau) ou diverses fonctions métaphoriques. En effet, l'emploi d'un terme à la place d'un autre se fonde sur le fait que tous deux partagent certains traits de sens. Par exemple, la phrase *il a filé comme une flèche* doit être comprise comme *il a filé rapidement* mais pas comme *il a filé comme un "trait qu'on lance avec un arc ou une arbalète et dont l'extrémité est ordinairement en forme de fer de lance"*. La métaphore comme la synonymie et l'antonymie peut faire l'objet d'une mesure continue (et non discrète) dont la définition reste à étudier.



## 7. Bibliographie

- [Chauché 1990] Chauché J. *Détermination sémantique en analyse structurale : une expérience basée sur une définition de distance*. TA Information, 1990, vol 31/1, p 17-24.
- [Deerwester and all. 1990] Deerwester S. and S. Dumais, T. Landauer, G. Furnas, R. Harshman, *Indexing by latent semantic analysis*. In Journal of the American Society of Information science, 1990, 416(6), p 391-407.
- [Ducrot et Todorov. 1972] Oswald Ducrot et Tzvetan Todorov, *Dictionnaire encyclopédique des sciences du langage*. Seuil 1972.
- [Ferrari 1997] Stéphane Ferrari. *Méthode et outils informatiques pour le traitement des métaphores dans les documents écrits*. thèse de doctorat de l'université Paris XI, 17 Décembre 1997.
- [Lafourcade and Sandford 1999] Lafourcade M. et E. Sandford, *Analyse et désambiguïsation lexicale par vecteurs sémantiques*. In Proc. of TALN'99 (Cargèse, July 1999) pp 351-356.
- [Lafourcade and Prince 2001] Lafourcade M. et V. Prince *Synonymies et vecteurs conceptuels*. To appear in In Proc. of Traitement Automatique du Langages Naturel (TALN'2001) (Tours, France, Juillet 2001) 10 p.
- [Lafourcade 2001] Lafourcade M. *Lexical sorting and lexical transfer by conceptual vectors*. In Proc. of the First International Workshop on MultiMedia Annotation (MMA'2001) (Tokyo, January 2001) 6 p.
- [Lafourcade 2001b] Mathieu Lafourcade. *Robust Thematic Lexical Database Construction with Conceptual Vector*. NLE (forthcoming), 33 p .
- [Lafourcade 2001c] Mathieu Lafourcade. *Robust Thematic Analysis with Conceptual Vector*. computationnal linguistics (forthcoming), 33 p .
- [Larousse 2001] Larousse. *Le Petit Larousse Illustré 2001*. Larousse, 2001.
- [Larousse 1992] Larousse. *Thésaurus Larousse - des idées aux mots, des mots aux idées*. Larousse, ISBN 2-03-320-148-1, 1992.
- [Larousse Universel 91] Larousse. *Grand Larousse Universel*. Larousse, 1982-1991.
- [Melch'čuk and all 95] Igor Melch'čuk, André Clas et Alain Polguère *Introduction à la lexicologie explicative et combinatoire*. ,éditions Duculot, 1995.
- [Melch'čuk and all 84] Igor Melch'čuk, avec Nadia Arbatchewsky-Jumarie, Léo Elnitsky, Lidija Iordanskaja et Adèle Lessard. rédaction : André Clas. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques I*, les presses universitaires de Montréal, 1984.
- [Melch'čuk and all 88] Igor Melch'čuk, avec Nadia Arbatchewsky-Jumarie, Louise Dagenais, Léo Elnitsky, Lidija Iordanskaja Marie-Noëlle Lefebvre et Suzanne Mantha. rédaction : André Clas. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques II*, les presses universitaires de Montréal, 1988.
- [Morin 1999] Morin, E. *Extraction de liens sémantiques entre termes à partir de corpus techniques*. Thèse de doctorat de l'Université de Nantes, 1999.
- [Ploux and Victorri1998] Ploux S. et B. Victorri *Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes*. In TAL, Vol 39/1, 1998 pp 161-182.
- [Resnik 1995] Resnik P. *Using Information contents to evaluate semantic similarity in a taxonomy*. In Proceedings of IJCAI-95, 1995.

[Robert 2000] Larousse. *Le Nouveau Petit Robert, dictionnaire alphabétique et analogique de la langue française*. Hachette, 2000.

[Salton and MacGill 1983] Salton G and M. J. MacGill. *Introduction to Modern Information Retrieval*. McGraw-Hill computer science series. McGraw-Hill, New-York, 1983.

[Univ Caen 1998] Université de Caen. *Dictionnaire de Synonymes*. 1998 <http://el-sap1.unicaen.fr>.

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Définitions et caractérisation de l’antonymie</b>	<b>3</b>
2.1	Antonymie et linguistique . . . . .	4
2.1.1	Antonymes complémentaires . . . . .	4
2.1.2	Antonymes scalaires . . . . .	5
2.2	Extensions de l’antonymie . . . . .	6
2.2.1	Conversifs . . . . .	6
2.2.2	Duals . . . . .	7
<b>3</b>	<b>Vecteurs Conceptuels</b>	<b>8</b>
3.1	Principe . . . . .	8
3.2	Distance angulaire . . . . .	9
3.3	Construction des vecteurs conceptuels . . . . .	11
3.3.1	Vecteurs conceptuels génératifs . . . . .	11
3.3.2	Noyau . . . . .	13
3.3.3	Apprentissage automatique . . . . .	14
<b>4</b>	<b>Fonctions d’antonymie</b>	<b>14</b>
4.1	Principes et définitions . . . . .	14
4.2	Listes d’antonymes . . . . .	17
4.2.1	Notation . . . . .	17
4.2.2	Construction de la liste d’antonymes . . . . .	17
4.2.3	Fonction <i>AntiC</i> . . . . .	18
4.3	Construction du vecteur antonyme : fonction Anti . . . . .	19

	Antonymie et vecteurs conceptuels	35
4.3.1	Définitions . . . . .	19
4.3.2	Construction du vecteur conceptuel antonyme . . . . .	19
4.4	Items lexicaux et vecteurs : problèmes et solutions . . . . .	22
4.4.1	Passage items lexicaux → Vecteurs conceptuels . . . . .	22
4.4.2	Passage Vecteurs conceptuels → items lexicaux . . . . .	23
<b>5</b>	<b>Mesures et fonction d'antonymie générale.</b>	<b>26</b>
5.1	Mesures d'antonymie potentielle . . . . .	26
5.1.1	Définitions . . . . .	26
5.1.2	Exemples . . . . .	27
5.2	Fonction d'antonymie globale . . . . .	28
5.3	Mesure d'évaluation de l'antonymie . . . . .	30
5.3.1	Exemples . . . . .	31
<b>6</b>	<b>Conclusions et perspectives</b>	<b>31</b>
<b>7</b>	<b>Bibliographie</b>	<b>33</b>