

Extraction semi-supervisée de couples d'antonymes grâce à leur morphologie

Didier Schwab, Mathieu Lafourcade et 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}>

Mots-clefs : antonymie, morphologie, antonymie complémentaire, antonymie scalaire, antonymie duale, répartition statistique

Keywords: antonymy, morphology, complementar antonymy, scalar antonymy, dual antonymy, statistic distribution

Résumé Dans le cadre de la recherche sur la représentation du sens en Traitement Automatique des Langues Naturelles, nous nous concentrons sur la construction d'un système capable d'acquérir le sens des mots, et les relations entre ces sens, à partir de dictionnaires à usage humain, du Web ou d'autres ressources lexicales. Pour l'antonymie, il n'existe pas de listes séparant les antonymies complémentaire, scalaire et duale. Nous présentons dans cet article une approche semi-supervisée permettant de construire ces listes. Notre méthode est basée sur les oppositions de nature morphologique qui peuvent exister entre les items lexicaux. À partir d'un premier ensemble de couples antonymes, elle permet non seulement de construire ces listes mais aussi de trouver des oppositions morphologiques. Nous étudions les résultats obtenus par cette méthode. En particulier, nous présentons les oppositions de préfixes ainsi découvertes et leur validité sur le corpus puis nous discutons de la répartition des types d'antonymie en fonction des couples opposés de préfixes.

Abstract In the framework of meaning representation in Natural Language Processing, we focus on enabling a system to autonomously learn word meanings and semantic relations from user dictionaries, web contents and other lexical resources. For antonymy, as a lexical semantic relation, no resource provides distinctions between complementary, scalar and dual antonymies. In this paper, we present a semi-supervised method to collate such lists, based on operating morphological opposition holding between lexical items. The approach presented here starts from a bootstrapped initial list. It is able to augment such lists but also to find out morphological oppositions. We scrutinize the obtained results and discuss the distribution of antonymy types.

1 Introduction

Dans le cadre de la recherche sur la représentation du sens en Traitement Automatique des Langues Naturelles, nous nous concentrons sur la construction d'un système capable d'acquérir le sens des mots, et les relations entre ces sens, à partir de dictionnaires à usage humain, du Web ou d'autres ressources lexicales. Pour le Français, qui constitue la langue de référence de notre expérimentation, des bases de données gratuites regroupant les principales fonctions lexicales n'existent pas ou ne correspondent pas exactement à nos attentes. En effet, les dictionnaires d'antonymes, regroupent indifféremment les trois types d'antonymie connus : *antonymie complémentaire*, *antonymie scalaire* et *antonymie duale*. Ces trois types seront expliqués dans le prochain chapitre. Afin de gagner en précision, nous cherchons donc à construire des listes d'antonymes basées sur cette typologie.

Nous présentons dans cet article une approche semi-supervisée permettant de construire ces listes. Notre méthode est basée sur les oppositions de nature morphologique qui peuvent exister entre les items lexicaux. Ainsi, les préfixes *ante-* et *post-* s'opposent sur une idée de durée. Cette méthode, à partir d'un premier ensemble de couples connus comme antonymes, extrait les préfixes susceptibles de s'opposer puis, cherche, dans un corpus constitué par les entrées de notre base lexicale, d'autres termes susceptibles d'être antonymes. Un expert valide les couples ainsi extraits et, par là même, les couples de préfixes qui leur correspondent. Une recherche automatique de nouveaux préfixes opposés est effectuée dans le corpus à partir de ces nouveaux couples. La méthode est itérée jusqu'au moment où il n'y a plus de préfixes candidats.

Nous étudions les résultats obtenus par cette méthode. En particulier, nous présentons les oppositions de préfixes ainsi découvertes et leur validité dans le corpus puis nous discutons de la répartition des types d'antonymie en fonction des couples opposés de préfixes.

2 L'antonymie

On considère que « *deux items lexicaux sont en relation d'antonymie si on peut exhiber une symétrie de leurs traits sémantiques par rapport à un axe.* » (Schwab et al., 2002). La symétrie se décline alors de différentes manières, selon la nature de son support. Il se dégage ainsi trois types d'antonymie : l'antonymie *complémentaire*, l'antonymie *scalaire* et l'antonymie *duale*, cette dernière regroupant les *conversifs* et les *duals propres*. Nous notons l'antonymie par un signe d'équivalence ayant subi une rotation de 90 degrés (*riche* \bowtie *pauvre*). Ce signe rappelle à la fois le signe marquant la synonymie, relation considérée comme opposée à l'antonymie, chez Polguère (Polguère, 2003) et la symétrie axiale existant entre les deux termes antonymes.

2.1 Antonymie complémentaire

Cette antonymie concerne les couples tels que *absent* \bowtie *présent*, ou *existence* \bowtie *inexistence*.

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, nous avons :

$\forall x \quad P(x) \Rightarrow \neg Q(x)$
 $\forall x \quad Q(x) \Rightarrow \neg P(x)$

$\forall x \quad \neg P(x) \Rightarrow Q(x)$
 $\forall x \quad \neg Q(x) \Rightarrow P(x)$

Nous reconnaissons ici une 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 présente deux types de symétrie : (1) une symétrie de valeur dans un système à deux valeurs seulement, comme dans l'exemple précédent, (2) 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 à toute combinaison de couleurs.

2.2 Antonymie scalaire

Les antonymes scalaires (ou gradables) concernent les systèmes échelonnés comme la taille (*grand*^{III_s}, *petit*[▷]) ou la température (*chaud*^{III_s}, *froid*[▷]). La symétrie se réalise par rapport à une valeur de référence du système qui n'est pas toujours représentée par un mot. Par exemple, pour *grand*^{III_s}, *petit*[▷], nous avons :

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*^{III}, *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, si R est la propriété ayant la valeur de référence (neutre ou médiane), par

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

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

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/grand que Pierre* », « *il avance/recule* » (le terme moyen étant *immobile*¹). 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 référera pas à la même idée de chaleur (donc à la même valeur de référence) qu'un habitant des fjords de Norvège.

2.3 Antonymie duale

Les antonymes *duals* 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.

2.3.1 Conversifs

Les conversifs (appelés aussi réciproques) sont des couples tels que *mari*^{III_d}, *femme*[▷], *acheter*^{III_d}, *vendre*[▷], *prêter*^{III_d}, *emprunter*[▷], *avant*^{III_d}, *après*[▷], *père*^{III_d}, *fil*[▷]. De nombreux linguistes comme Igor Mel'čuk (Mel'čuk *et al.*, 1995) ne les considèrent pas comme des antonymes. La fonction *anti* de son DEC désigne en réalité les antonymes complémentaires et les antonymes scalaires. Il dédie aux conversifs une autre fonction lexicale *conv*. Cependant, dans la mesure où pour nous, la modélisation de l'antonymie correspond à une étude complète des mécanismes de symétrie, nous avons considéré les conversifs comme un cas particulier de symétrie, et les avons naturellement associés à un processus antonymique "étendu".

Pierre est le père de Marc ↔ Marc est le fils de Pierre.

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

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

¹On peut remarquer que le neutre d'un type d'antonymie peut être opposable dans une autre antonymie. Ici, par exemple, *mobile*^{III_c}, *immobile*[▷]

Dans le cas des conversifs, si on remplace dans une phrase un terme x par son réciproque y , on peut systématiquement rétablir la synonymie entre les deux phrases à condition de permuter les arguments syntaxiques mis en relation par x 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).

2.3.2 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). Les duals sont des mots que la culture associe comme *soleil* et *lune*, ou qui ne vont pas, à priori, l'un sans l'autre comme *question* et *réponse* ou alors sont l'expression d'une antonymie temporelle i.e. qui exprime le passage d'un état à un autre comme *naissance* et *décès*. Dans ce troisième cas, on peut remarquer que ces deux événements marquent le passage entre deux antonymes complémentaires (*inexistence* et *existence* dans le cas de *naissance* et *décès* ou bien *présence* et *absence* dans le cas de *départ* et *arrivée*). L'antonymie duale propre présente naturellement une symétrie qui n'est pas relevée dans l'échange des places d'arguments puisqu'il s'agit de prédicats unaires. Elle exprime le fait que si l'un des deux prédicats est vrai, il existe une valeur pour laquelle l'autre l'est aussi nécessairement. Pour la modéliser, on écrira :

$$\exists x_0 P(x_0) \leftrightarrow \exists Q, Q(x_0)$$

avec Q dual de P qui modélise par exemple le fait que si x_0 a un début, alors il existe aussi une fin à x_0 ou :

$$\exists x_0 P(x_0) \leftrightarrow \exists Q, \exists y_0 Q(y_0)$$

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 descripteurs temporels. Ainsi, *avant* et *après* en prédicats unaires, sont 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* et *après* qui sont des scalaires avec comme valeur médiane *pendant*.

3 Construction de listes d'antonymes

3.1 Problématique

Dans le cadre de nos recherches sur la représentation du sens en TALN, notre modèle est basé sur un apprentissage effectué à partir de diverses sources comme des dictionnaires à usage humain, le Web, ou des dictionnaires de relations lexicales (Schwab *et al.*, 2004). S'il existe depuis quelques années pour le Français des dictionnaires de synonymes sur le web (celui du CRISCO² mais aussi une partie d'EuroWordNet), les dictionnaires d'antonymes étaient eux inexistantes. Seul, depuis quelques mois, le dictionnaire du CRISCO fournit des antonymes. Toutefois, il ne considère pas les différents types d'antonymies présentées dans la section 2. C'est pourquoi, dans le but d'améliorer nos représentations nous avons cherché à étudier des méthodes qui nous permettraient de construire le plus automatiquement possible des listes de couple d'antonymes classées suivant leur type. La méthode choisie va utiliser les oppositions de nature morphologique entre les termes.

²[urlhttp://elsap1.unicaen.fr/cgi-bin/cherches.cgi](http://elsap1.unicaen.fr/cgi-bin/cherches.cgi)

3.2 Morphologie et antonymie

Depuis Charlemagne, l'habitude a été prise de créer des termes à partir de racines latines mais aussi grecques (Walter, 1988). C'est pourquoi en français il n'est pas rare de trouver des termes dits *populaires* c'est-à-dire ayant subi des déformations normales dans une langue (‘*mère*’, ‘*ciel*’) à côté de termes dits *savants* qui eux sont directement construits à partir de morphèmes issus du latin ou du grec (‘*maternel*’, ‘*céleste*’).

Les morphèmes sont les unités minimales significatives qui constituent les mots. Par exemple, le mot “fleurs” est composé de deux morphèmes : le radical (ou base) correspondant à l’item ‘*fleur*’ et le suffixe marquant le pluriel *s*. Il existe deux types de morphèmes : (1) les *morphèmes lexicaux* qui correspondent aux items lexicaux ou à une légère variante ; (2) les *morphèmes grammaticaux*, autrement appelés *affixes*. Situés avant le radical, un affixe est dit *préfixe*, après, *suffixe*, dans le radical, *infixe*.

Les morphèmes sont porteurs de sens. Par exemple, le préfixe latin *bi-* correspond à *deux* (‘*binaire*’, ‘*bisexuel*’), le préfixe latin *semi-* et le préfixe grec *hémi-* à une idée de milieu (‘*semi-conducteur*’, ‘*semi-rigide*’, ‘*hémisphère*’) et le préfixe grec *péri-* à ‘*tour*’ (‘*périmètre*’, ‘*péricarde*’).

De nombreux mots “*savants*” ont donc été (et sont encore) créés “de toute pièce” en utilisant des préfixes ou des suffixes marquant une idée négative par rapport à la racine ou alors provenant de mots opposés en Latin, en Grec et aujourd’hui en Français. Ce sont ces deux types d’affixes que nous allons utiliser pour construire automatiquement des listes d’antonymes. Ainsi, les préfixes *poly-* (‘*plusieurs*’) et *mono-* (‘*un*’) s’opposent sur le *nombre*, *hyper-* (superlatif) et *hypo-* (au-dessous) s’opposent par rapport à une valeur de *référence* tandis que *non-* ou *més-* marquent la négation. Pour les suffixes, on remarquera l’opposition *-phobe*|||*-phile* (‘*homophobe*’|||‘*homophile*’) ou l’opposition *-dynamique*|||*-statique* (‘*hydrostatique*’|||‘*hydrodynamique*’). Notre article n’étudie ici que les préfixes mais la méthode choisie pour extraire les suffixes est la même.

Il existe des études comme (Béchade, 1992) sur les préfixes du Français et leur signification. Elles permettent de faire une première étude du comportement des préfixes mais ne permettent pas de façon rigoureuse d’opposer deux préfixes et encore moins de chercher à savoir de quel type d’antonymie ils sont marqueurs. Ainsi, une vérification en corpus s’avère impérative. C’est pour cette raison que nous avons mis au point un processus semi-automatique de construction de listes d’antonymes qui permet aussi d’extraire les préfixes opposés.

4 Processus

4.1 Principe

Notre méthode est proche de celle utilisée par (Morin, 1999) pour l’acquisition de schémas lexico-syntaxiques mais elle s’en différencie sur deux points principaux :

- Notre problématique première n’est pas de récupérer des morphèmes opposés mais bien de **construire**, le plus automatiquement possible, des listes d’antonymes. Nous avons tout de même conservé les informations concernant ces préfixes afin de savoir s’ils caractérisent plus particulièrement telle ou telle antonymie. Nous présentons ces résultats dans la section 5.
- Dans (Morin, 1999), les experts valident directement les schémas. Ici, les schémas sont validés indirectement si des couples d’items antonymes, caractérisés par ces schémas, le sont. Cette méthode semble augmenter le nombre de cas à examiner mais il nous paraît difficile d’éliminer des couples de préfixes sans observer leur comportement. Un filtrage automatique basé sur le nombre de couples validés permet de limiter ces cas.

La figure 1 présente le processus de construction de ces listes qui est composé de sept étapes. La méthode utilisant les suffixes est identique.

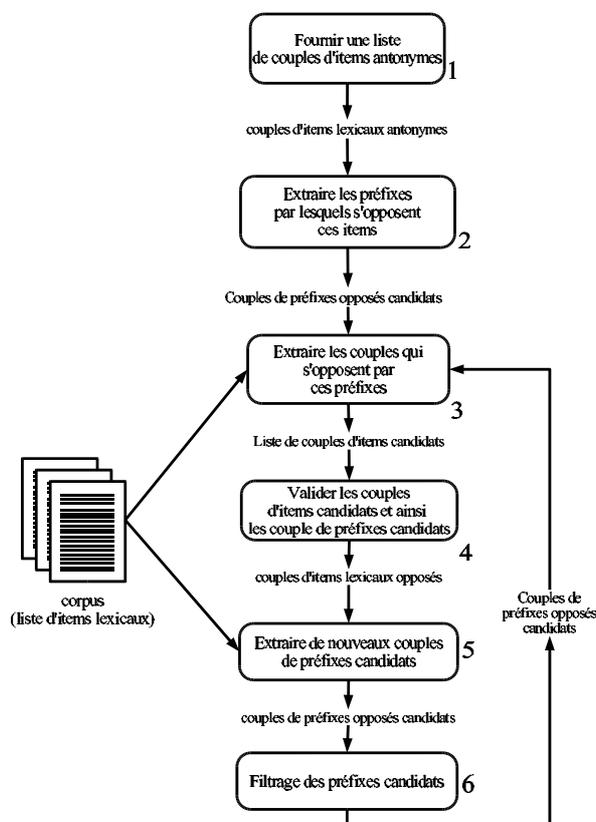


FIG. 1 – Processus d’acquisition de préfixes et de termes antonymes.

1. *Fournir une liste de couples d’items lexicaux antonymes* : cette liste peut être fournie par des dictionnaires ou bien spécifiées manuellement. Elle va permettre d’avoir un noyau de référence pour amorcer le processus.
2. *Extraire les préfixes par lequel s’opposent ces items* : la méthode consiste à enlever aux mots le plus long suffixe commun aux deux. Ainsi, avec *«monosémique»* et *«polysémique»*, on enlève le suffixe *sémique*. On obtient ainsi une liste de couples de préfixes candidats. Ces deux premières étapes sont facultatives. On peut directement fournir une liste de préfixes censés s’opposer.
3. *Extraire les couples qui s’opposent par ces préfixes* : on extrait du corpus ces couples d’items. On obtient ainsi une liste de couples d’items candidats.
4. *Valider les couples candidats* : on vérifie manuellement les termes candidats et on ne conserve que ceux qui sont effectivement des antonymes. Cette phase est réalisée par un "expert". Si au moins un des couples validés est caractérisé par un des couples de préfixes candidats, ce dernier est validé.
5. *Extraire des préfixes candidats* : parmi la liste des couples retenus, on extrait les suffixes pour trouver dans la base de nouveaux préfixes marquant l’opposition. Le principe consiste à sortir du corpus l’ensemble des termes ayant ce suffixe et d’extraire les préfixes de chacun.
6. *Filtrer les préfixes candidats* : il s’agit de ne retenir que les couples de préfixes qui caractérisent au moins *n* couples d’items lexicaux dans le corpus.

Le choix de n est important puisque si n est grand, on élimine un grand nombre de préfixes à vérifier manuellement ce qui facilite la tâche de l'expert mais ne garantit pas l'extraction de l'ensemble des oppositions de préfixes. En revanche, le choix d'un n trop petit multiplie le nombre de couples à vérifier et s'avère difficile à concevoir (un choix de $n = 2$ sur la liste des "a" privatifs entraîne la vérification de 3251 couples de préfixes !). Notre choix s'est porté sur un compromis de $n = 5$ qui élimine un nombre suffisant de candidats (Dans notre expérience, nous n'avons plus alors que 110 vérifications à effectuer).

4.2 Déroulement

Considérons le corpus très réduit suivant : *acyclique*, *anticyclique*, *anticyclone*, *antimoine*, *bisémiq*, *biphonie*, *cyclique*, *moine*, *monophonie*, *monosémiq*, *polyphonie*, *polysémiq*, *souris*. Du fait de la taille du corpus, ici $n = 2$.

1. *étape 1* : On prend une liste de couples que l'on considère comme antonymes :
monosémiq ||| *polysémiq*, *acyclique* ||| *cyclique*
2. *étape 2* : On extrait les préfixes par lesquels ces couples s'opposent :
mono- ||| *poly*-, *a*- ||| ϵ
3. *étape 3* : On extrait les couples qui s'opposent par ces préfixes :
pour *mono*- ||| *poly*-, *monosémiq* ||| *polysémiq*, *monophonie* ||| *polyphonie*
pour *a*- ||| ϵ , *acyclique* ||| *cyclique*
4. *étape 4* : L'expert valide les paires extraites :
monosémiq ||| *polysémiq* et *acyclique* ||| *cyclique* sont validées, les deux couples de préfixes sont validés.
monophonie ||| *polyphonie* est rejetée, cela n'entraîne aucune conséquence sur la validité des préfixes.
5. *étape 5* : On extrait de nouveaux préfixes candidats grâce aux paires validées et au corpus. *monosémiq* ||| *polysémiq* possèdent en commun le suffixe *sémiq*, on recherche dans le corpus les termes qui ont cette même caractéristique. On extrait ainsi *polysémiq*, *monosémiq* et *bisémiq*. En excluant l'opposition déjà considérée *mono*- ||| *poly*-, ces couples de termes nous permettent comme nouveaux couples de préfixes candidats *mono*- ||| *bi*- et *poly* ||| *bi*-. De même, *acyclique* ||| *cyclique* permet d'obtenir *anti*- ||| ϵ
6. *étape 6* : On filtre les couples de préfixes :
 - *mono*- ||| *bi*- apparaît dans deux couples (*monosémiq* ||| *bisémiq*, *monophonie* ||| *biphonie*), il est donc conservé.
 - *poly* ||| *bi*-, en revanche, n'apparaît qu'une fois dans le corpus, il est donc supprimé. Dans notre expérience, ces préfixes candidats n'ont pas été rejetés par le filtrage automatique mais par les experts qui ont invalidé les dix couples d'items extraits pour ce couple de préfixes. Il ne semble donc pas y avoir d'exception pour les préfixes *bi*- (marquant une idée de *deux*) et *poly* (marquant une idée de *plusieurs*) dont les idées sont incluses l'une dans l'autre.
 - *anti*- ||| ϵ apparaît lui dans trois couples *cyclone* ||| *anticyclone*, *cyclique* ||| *anticyclique* et *moine* ||| *antimoine*. En pratique la validation ne cherche bien sûr pas tous les couples et s'arrête dès qu'elle en a trouvé deux.

On réitère l'étape 3

7. *étape 3* : On extrait les couples qui s'opposent par ces préfixes :
pour *mono*- ||| *bi*-, *monosémiq* ||| *bisémiq*, *monophonie* ||| *biphonie*
pour *anti*- ||| ϵ , *cyclone* ||| *anticyclone*, *cyclique* ||| *anticyclique* et *moine* ||| *antimoine*.
8. *étape 4* : L'expert valide les paires extraites :
monosémiq ||| *bisémiq*, *monophonie* ||| *biphonie*, *cyclone* ||| *anticyclone*, et *cyclique* ||| *anticyclique* sont validées, les couples de préfixes *mono*- ||| *bi*- et *anti*- ||| ϵ sont validés.
moine ||| *antimoine* est rejetée, cela n'entraîne aucune conséquence sur la validité des préfixes.

9. *étape 5* : On cherche à extraire de nouveaux préfixes candidats grâce aux paires validées et au corpus.
Il n’y en a plus, le processus s’arrête.

5 Résultats

Le corpus que nous avons utilisé est constitué de 79 220 items lexicaux issus de notre base lexicale sémantique (Schwab *et al.*, 2004). Ces termes correspondent globalement aux entrées hors noms propres de dictionnaires sous forme électronique : dictionnaires classiques (Larousse, 2004) (Robert, 2000), dictionnaires de synonymes (CRISCO³), thésaurus (Larousse, 1992).

Lors de cette expérience, pour de simples raisons pratiques, les auteurs de cette publication ont joué le rôle d’expert validateur. On peut estimer le temps de réalisation de l’expérience à une cinquantaine d’heures utilisées à plus de 99% par la phase de validation (étape 4).

Cette méthode nous a permis d’extraire 49 couples de préfixes opposés. Le tableau de la figure 2 en présente quelques uns.

préfixe 1	préfixe 2	exemple	contre-exemple
a-	ε	‘chromatique’ ‘achromatique’	‘afin’ / ‘fin’
an-	ε	‘aérobie’ ‘anaérobie’	‘anatomiste’ / ‘atomiste’
anti-	ε	‘communiste’ ‘anticommuniste’	‘moine’ / ‘antimoine’
dés-	ε	‘accord’ ‘désaccord’	‘avouer’ / ‘désavouer’
in-	ε	‘imaginable’ ‘inimaginable’	‘incas’ / ‘cas’
il-	ε	‘licite’ ‘illicite’	‘illustre’ / ‘lustre’
pré-	post-	‘préface’ ‘postface’	ε
hyper-	hypo-	‘hyperonymie’ ‘hyponymie’	‘hyperstatique’ / ‘hypostatique’
syno-	anto-	‘synonymie’ ‘antonymie’	ε
méro-	holo-	‘méronymie’ ‘holonymie’	ε
mono-	poly-	‘polysémique’ ‘monosémique’	‘monophonie’ / ‘polyphonie’
mono-	stéréo	‘monophonie’ ‘stéréophonie’	‘monotype’ / ‘stéréotype’

FIG. 2 – Exemples de préfixes antonymes extraits

La figure 3 présente le pourcentage de paires validées par l’expert, c’est-à-dire le nombre de paires considérées comme valides pour chaque couple de préfixes antonymes candidats. On peut constater que si le taux de validation est très important pour la plupart des couples de préfixes il est, en revanche, très faible pour le *a privatif* (*a-* et *an-*).

Finissons par la typologie des paires d’antonymes extraites. La figure 4 présente les résultats obtenus. On constate que globalement la morphologie permet de relativement bien connaître le type d’antonymie. Ainsi, on peut donner quelques indications sur les couple de préfixes suivant leur sémantique :

- *temporels* (*anté-*|||ε, *post-*|||ε, *anté-*|||*post-*, *pré-*|||*post-*) : Tous sont scalaires sauf le couple dual ‘christ’|||‘antéchrist’. La définition de (Larousse, 2004) nous donne « *Imposteur qui, suivant l’Apocalypse, doit venir quelque temps avant la fin du monde pour essayer d’établir une religion opposée à celle de Jésus-Christ.* ». Dans ce cas, l’opposition sémantique s’explique

³url<http://elsap1.unicaen.fr/cgi-bin/cherches.cgi>

préfixe 1	préfixe 2	nombre de paires extraites	nombre de paires validées	nombre de paires invalidées	pourcentage
a-	ε	288	71	217	24,6%
an-	ε	61	15	41	24,6%
anti-	ε	156	147	9	94,2%
dés-	ε	195	179	16	91,7%
in-	ε	616	539	77	87,5%
il-	ε	31	18	13	58%
anté-	ε	15	10	5	80%
post-	ε	37	33	4	89%
anté-	post-	3	2	1	66,6%
pré-	post-	11	11	0	100%
hyper-	hypo-	25	24	1	96%
syno-	anto-	4	4	0	100%
méro-	holo-	2	2	0	100%
mono-	poly-	28	25	3	89,2%
mono-	stéréo-	5	4	1	80%

FIG. 3 – Extraits des résultats d'extraction

par l'opposition «*dieu*»/«*démon*» tandis que la construction de l'opposition morphologique exprime l'idée de l'arrivée du démon avant le retour du christ.

- *médicaux (hyper-|||hypo-)* : Ils caractérisent des mesures qui sont donc au-dessus ou au-dessous de la normale («*hyperthyroïdie*»/«*hypothyroïdie*»).
- *nombre (bi-|||ε, tri-|||ε, mono-|||poly-, mono-|||stéréo-)* : Ils s'opposent par une propriété possédée une fois (*mono*) ou plusieurs (*bi-*, *tri-*, *poly*, *stéréo-*). Ils sont complémentaires.
- *absence de propriété* : *il-|||ε* («*illicite*»/«*licite*», «*illimité*»/«*limité*»), *a-|||ε* («*typique*»/«*atypique*», «*sociabilité*»/«*asociabilité*»), *an-|||ε* («*anencéphale*»/«*encéphale*») ils sont tous complémentaires sauf «*anion*»/«*ion*» qui relève plutôt de l'antonymie duale.
- *opposition culturelle ou produit permettant de lutter contre quelque chose* : *anti-|||* («*anticléric*»/«*cléric*», «*antiviral*»/«*viral*»). Ils sont duals.

6 Conclusion

Dans cet article, nous avons présenté la méthode semi-supervisée qui nous a permis de construire le plus efficacement possible trois listes de couples d'antonymes, chacune correspondant à un des trois types : complémentaire, scalaire et dual. Cette méthode est basée sur les oppositions de nature morphologique qui peuvent exister entre les items lexicaux et utilise un corpus constitué des entrées de notre base lexicale. À partir d'un premier ensemble de couples antonymes, cette méthode permet à la fois de construire ces listes et de trouver de nouveaux couples de préfixes antonymes. Nous avons ainsi pu étudier la distribution des différents types d'antonymie en fonction du couple de préfixe.

préfixe 1	préfixe 2	complémentaires	scalaires	duals
a-	ε	71 (100 %)		
an-	ε	14 (92,8%)		1 (7,2%)
anti-	ε			147 (100 %)
dés-	ε		3 (1,6 %)	176 (98,3 %)
in-	ε	539 (100 %)		
il-	ε	18 (100%)		
anté-	ε		9 (90%)	1 (10%)
post-	ε		23 (100 %)	
anté-	post-		2 (100%)	
pré-	post-		11 (100%)	
hyper-	hypo-		24 (100 %)	
mono-	poly-	25 (100%)		
mono-	stéréo-	4 (100 %)		

FIG. 4 – Répartition des schémas suivant le type

Références

Hervé BÉCHADE. *Phonétique et morphologie du français moderne et contemporain*. Presses Universitaires de France, 1992.

LAROUSSE, *Thésaurus Larousse - des idées aux mots, des mots aux idées*. Larousse, 1992.

LAROUSSE, *Le Petit Larousse Illustré 2004*. Larousse, 2004.

Igor MEL'ČUK, André CLAS, et Alain POLGUÈRE. *Introduction à la lexicologie explicative et combinatoire*. Duculot, 1995.

Emmanuel MORIN. « *Extraction de liens sémantiques entre termes à partir de corpus techniques* ». Thèse de doctorat, Université de Nantes, 1999.

Alain POLGUÈRE. *Lexicologie et sémantique lexicale*. Les Presses de l'Université de Montréal, 2003.

Le ROBERT, *Le Nouveau Petit Robert, dictionnaire alphabétique et analogique de la langue française*. Éditions Le Robert, 2000.

Didier SCHWAB, Mathieu LAFOURCADE, et Violaine PRINCE. « Vers l'apprentissage automatique, pour et par les vecteurs conceptuels, de fonctions lexicales. L'exemple de l'antonymie ». Dans les actes de *TALN 2002*, volume 1, Nancy, Juin 2002.

Didier SCHWAB, Mathieu LAFOURCADE, et Violaine PRINCE. « Hypothèses pour la construction et l'exploitation conjointe d'une base lexicale sémantique basée sur les vecteurs conceptuels ». Dans les actes de *JADT 2004*, Louvain-La-Neuve, Belgique, Mars 2004.

Henriette WALTER. *Le Français dans tous les sens*. Livre de poche, 1988.