

L'antonymie observée avec des méthodes de TAL : une relation à la fois syntagmatique et paradigmatisque ?

1 - Objectif

Tester l'hypothèse d'un **double fonctionnement syntagmatique/paradigmatique** de la relation d'antonymie (Murphy 06) parmi les adjectifs du français en mobilisant des outils de TAL :

- un jeu de patrons lexico-syntaxiques pour la dimension syntagmatique,
- une base de voisins distributionnels pour la dimension paradigmatisque.

Le recoupement de ces deux dimensions permet-il un meilleur repérage de la relation d'antonymie lors d'une tâche d'extraction ?

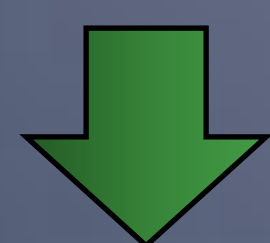
2 - Méthodologie

Patrons lexico-syntaxiques Axe syntagmatique

Inspirés de Fellbaum 95 et Jones 02

PATRON	EXEMPLE
X ou Y	Cette connexion peut être temporaire ou définitive .
à la fois X et Y	Sa production est à la fois fermière et industrielle .
entre X et Y	les différences entre les vins blancs et les vins rouges
plus/plutôt/moins/autant/aussi (bien) X que Y	Il se déguste aussi bien chaud que froid .
X plutôt que Y	Il décrit une Terre sphérique plutôt que plate .
soit X soit Y	Les coups francs sont soit directs soit indirects .
ni X ni Y	Il n'est donc ni explicite ni implicite .

Projection sur corpus (Wikipedia, ≈200 millions de mots)



13 295 couples uniques d'adjectifs
→ restriction aux **907** couples qui apparaissent dans, au moins, 2 patrons différents (réduction du bruit)

Ces deux méthodes permettent de capter deux propriétés des antonymes :

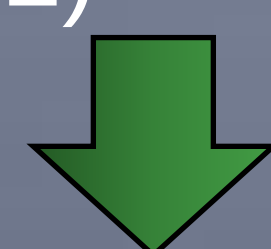
- leur tendance à apparaître dans des structures récurrentes,
- le fait qu'ils sont substituables.

→ complémentarité de ces deux approches (67% des 907 paires retenues apparaissent dans la base de voisins)

Voisins distributionnels Axe paradigmatisque

Base lexicale *Les voisins de Wikipedia* constituée par analyse distributionnelle automatique du corpus Wikipedia :

- 1.extraction des relations syntaxiques (Bourigault 07) sous forme de triplets <gouverneur, relation, dépendant>
- 2.rapprochement des mots partageant les mêmes contextes syntaxiques (Bourigault 02)



≈**4 millions de couples** (dont 205 000 Adj/Adj) → relations sémantiques diversifiées, bruit

Auteurs

François MORLANE-HONDÈRE, Cécile FABRE
CLLE-ERSS, Université de Toulouse
{francois.morlane,cecile.fabre}@univ-tlse2.fr



3 - Évaluation

Première évaluation - Utilisation d'une ressource de référence, le dictionnaire de synonymes et d'antonymes du CRISCO.

	Antonyme	Synonyme	Absent	
Voisins	32%	4%	64%	100%
Non-voisins	27%	6%	67%	100%

→ La comparaison avec le dictionnaire montre qu'environ 1/3 des 907 couples, voisins et non-voisins confondus, relève de l'antonymie.

→ Le calcul du χ^2 indique que la **différence entre la proportion d'antonymes parmi les voisins et les non-voisins n'est pas significative**.

Parmi les autres paires :

- quelques-unes sont identifiées comme synonymes (*simple/facile*, *accidentel/fortuit*),
- la plupart sont tout simplement absentes de la ressource.

Deuxième évaluation - Recours au jugement de locuteurs

- Constitution de 3 jeux de 100 paires parmi les 907 (50% de voisins chacun)
- Chaque jeu est proposé à deux locuteurs → identification de la relation entre les paires : *opposition forte/faible*, *synonymie*, *hyperonymie*, etc.

Proportion de couples identifiés comme des opposés *forts* par les deux locuteurs :

Jeu 1		Jeu 2		Jeu 3	
Patrons seuls	Patrons + voisins	Patrons seuls	Patrons + voisins	Patrons seuls	Patrons + voisins
12 %	11 %	16 %	16 %	14 %	16 %

→ Les proportions sont à peu près similaires pour les voisins et les non-voisins : le critère de la substituabilité n'est pas déterminant pour le jugement d'opposition forte.

4 - Analyse

- Certaines paires ont des fréquences trop faibles pour être captées par le système d'analyse distributionnelle utilisé (*dioïque/monoïque*).
- D'autres paires, malgré des fréquences relativement élevées, ne sont pas reconnues comme des voisins distributionnels : pour des couples comme *domestique/professionnel* ou *fixe/portable*, qui ont été identifiés comme des opposés fort, le critère de la cooccurrence prime sur celui de la substituabilité.

Références

BOURIGAULT D.(2002). UPERY : un outil d'analyse distributionnelle étendue pour la construction d'ontologies à partir de corpus. In *Actes de la 9^e conférence sur le Traitement Automatique de la Langue Naturelle*, p. 75–84, Nancy.
BOURIGAULT D.(2007). *Un analyseur syntaxique opérationnel : SYNTAX*. Mémoire d'habilitation à diriger des recherches. Université Toulouse II – Le Mirail.
FELLBAUM C.(1995). Co-occurrence and antonymy. *International journal of lexicography*, 8.
JONES S.(2002). *Antonymy: a corpus-based perspective*. Routledge.
M URPHY L. (2006). Antonyms as lexical constructions: or, why paradigmatic construction is not an oxymoron. *Constructions all over : case studies and theoretical implications*. Special volume of *Constructions*, SV1(8).