

Compréhension automatique du langage naturel : quelques applications récentes des liens entre logique et langage naturel

Christian Retoré, Université de Montpellier & Equipe Texte, LIRMM-CNRS

J'essaierai dans cette présentation de montrer comment les grammaires formelles, la logique, le lambda calcul typé et les graphes de relations sémantiques entre mots permettent de développer des analyseurs syntaxiques et sémantiques du langage naturel. Ces derniers, à partir de phrases ou de textes construisent automatiquement des représentations du sens utilisables par des applications comme la recherche d'information, la réponse automatique à des questions, ou le dialogue homme-machine. Ce domaine de l'informatique et plus particulièrement de l'intelligence artificielle est apparu après la seconde guerre mondiale (Turing), et il connaît actuellement certains succès. Je parlerai plus particulièrement des aspects logiques des avancées récentes.

Dès ses débuts, la logique est étroitement liée à des considérations linguistiques : la logique se pratique en langage naturel et les logiciens de l'Antiquité et du Moyen-Âge soulignent de nombreux parallèles : une phrase est une proposition, le groupe verbal un prédicat dont le sujet grammatical est aussi le sujet logique, etc. Les grammaires catégorielles (Lambek 1958, cf. [6]) formalisent et automatisent la correspondance entre la structure syntaxique de la phrase, vue comme une preuve formelle de sa correction grammaticale, et son sens vu comme une proposition logique : cela peut se voir comme une conséquence de l'isomorphisme de Curry-Howard (1969) entre déductions et lambda termes. D'un point de vue pratique, il existe aujourd'hui des grammaires catégorielles à large échelle, acquises automatiquement, qui traduisent les phrases en propositions logiques, notamment pour le français [3] ou l'anglais [5]

Une partie des recherches récentes [1,7] cherche à enrichir le versant sémantique des grammaires catégorielles par des considérations lexicales : comment reconstituer l'événement qui se termine dans la phrase « *Anaïs a fini son livre* » ? Comment rejeter « *Bertille a mangé ce livre.* » et accepter « *Céline a dévoré ce livre.* » — en l'interprétant correctement ? Comment savoir si « *livre* » renvoie à l'objet matériel ou à un contenu informationnel ? « *J'ai perdu ce livre de Murakami que j'avais tant aimé.* » Pour ce faire, la grammaire doit intégrer les données issues des réseaux lexicaux (par ex. *JeuxDeMots* [4]), ou de l'analyse statistique des textes (par ex. [8]).

Quelques références :

1. Nicholas Asher. *Lexical Meaning in context – a web of words*. Cambridge University press, 2011.
2. Johan Bos (2011): *A Survey of Computational Semantics: Representation, Inference and Knowledge in Wide-Coverage Text Understanding*. *Language and Linguistics Compass* 5(6): 336–366.
3. Mathieu Lafourcade, Alain Joubert. Computing trees of named word usages from a crowdsourced lexical network. In *IMCSIT*, volume Computational Linguistics – Applications (CLA'10), pages 439–446, 2010.
4. Richard Moot. Wide-coverage French syntax and semantics using Grail. In *Proceedings of Traitement Automatique des Langues Naturelles (TALN)*, Montreal, 2010.
5. Richard Moot, Christian Retoré The logic of categorial grammars: a deductive account of natural language syntax and semantics July 2012 Springer LNCS 6850
6. Christian Retoré [The Montagovian Generative Lexicon \$\Lambda\$ T_n: a Type Theoretical Framework for Natural Language Semantics](#) in Ralph Matthes and Aleksy Schubert (eds) 19th International Conference on Types for Proofs and Programs (TYPES 2013) LIPICS vol 26 pp. 202--229. <http://dx.doi.org/10.4230/LIPIcs.TYPES.2013.202>
7. Tim van de Cruys: Mining for Meaning: The Extraction of Lexico-Semantic Knowledge from Text. PhD thesis, University of Groningen (2010).