

MODÉLISATION DE PRÉFÉRENCES POUR L'INTERPRÉTATION D'ÉNONCÉS AMBIGUS

Jean-Philippe Prost et Christian Retoré
(Université de Montpellier et LIRMM)

Ce sujet de linguistique computationnelle concerne l'introduction de préférences, comme cela a été fait pour la syntaxe [1], pour la sémantique logique d'un énoncé intégrant le sens lexical [2]. Il ouvre la porte à l'étude cognitive du langage en confrontant la modélisation à des expériences.

Un énoncé en langue naturelle fait souvent l'objet de plusieurs lectures. La multiplicité des interprétations peut provenir d'ambiguïtés lexicales, syntaxiques ou sémantiques, ainsi que d'irrégularités syntaxiques (erreurs de grammaire, ouverture lexicale, fragment bien formé, langage familier, etc.). Tous ces aspects contribuent de concert à l'interprétation d'une phrase. L'objet de cette thèse est de calculer les sens possibles (identifiés avec des formules d'une logique multisorte) d'un énoncé présentant de telles lectures multiples, et de modéliser la préférence humaine pour certaines d'entre elles dans un cadre formel unifié intégrant les différents niveaux de langue mentionnés plus haut.

Cette approche ouvre la porte à l'étude cognitive du langage en confrontant la modélisation à des expériences. L'étude couvrira notamment des phénomènes linguistiques comme les suivants :

- (1) la quantification : *Il y a un gardien devant toutes les portes du musée*
→ un gardien pour chaque porte, ou un seul gardien pour toutes les portes ?
- (2) la portée de la négation : *Elle n'a pas offert de livre à Marie*
→ ¬livre ou ¬elle ou ¬offert ou ¬Marie ou ¬{phrase} ?
- (3) des catégories d'énoncés non standards¹ :
 - a. *Une jeune père de 13 ans* (ambiguïté lexicale)
→ un jeune père ou une jeune mère ?
 - b. *Quoi a-t-il écrit un livre sur ?* (ambiguïté structurelle)
→ *sur quoi (...)* ? Ou *Qu'a-t-il écrit **sur** un livre ?*

Cette modélisation se fera sur la base de la prise en compte d'un degré d'acceptabilité [3], dans un cadre compositionnel intégrant syntaxe, structure logique et sémantique lexicale [2]. Dans le cas de langage non-standard, il sera notamment envisagé que la structure syntaxique ne fasse pas nécessairement l'objet de corrections. On préférera, lorsque c'est possible, s'appuyer pour cela sur le réseau de contraintes syntaxiques qui décrit les propriétés grammaticales de l'énoncé, auquel seront appliquées des préférences, comme dans [1]. Ces degrés de préférence pourront également prendre en compte le sens des quantificateurs et le sens lexical comme dans [2] et [4].

Ce sujet s'inscrit dans l'actualité de la Recherche en linguistique computationnelle, comme en témoignent différents colloques internationaux récemment organisés, ainsi que les projets en cours au sein de l'équipe TEXTE.

References

- [1] J-Ph. Prost. Jugement exact de grammaticalité d'arbre syntaxique probable. In *Proc. of TALN 2014 (Vol. 1: Long Papers)*, pp. 352–362 2014. ATALA.
- [2] C. Retoré. The Montagovian Generative Lexicon ΛT_{yn} : a Type Theoretical Framework for Natural Language Semantics. In R. Matthes and A. Schubert, eds, *Proc. of TYPES 2013*, vol. 26 of *Leibniz Int'l Proc. in Informatics (LIPIcs)*, pp. 202–229, 2014. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- [3] K. Nagao. A Preferential Constraint Satisfaction Technique For Natural Language Analysis. *IEICE TRANSACTIONS on Information and Systems*, 77(2):161–170, 1994.
- [4] J-Ph. Prost and M. Lafourcade. Pairing Model-Theoretic Syntax And Semantic Network For Writing Assistance. In *Proc. of the CSLP@Context'11*, 2011.

¹Les exemples sont extraits du Projet Voltaire, "*Petite liste des différentes fautes d'orthographe que l'on risque de commettre*" (<http://www.projet-voltaire.fr/blog/actualite/petite-liste-des-differentes-fautes-dorthographe-que-lon-risque-de-commettre>) et de la sémanticlopédie (<http://www.semantique-gdr.net/dico/index.php/Portée>).