Actes de CORIA-TALN 2026
Nantes, France, 29 juin au 3 juillet 2026

Le corpus NeWMe : un corpus de référence pour l'étude de la négociation du sens des mots

Aina Garí Soler[1, 2, 3] Jenny Myrendal[4, 5] Chloé Clavel[1] Staffan Larsson[4]
(1) INRIA Paris, France
(2) LTCI, Télécom-Paris, Institut Polytechnique de Paris, France
(3) Miles Team, LAMSADE, Université Paris Dauphine-PSL, Paris, France
(4) Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg, Sweden
(5) Department of Education, Communication and Learning, University of Gothenburg, Sweden
aina.gari-soler@inria.fr

RÉSUMÉ

Les séquences de négociation du sens des mots (*Word Meaning Negotiation*, WMN) apparaissent lorsque des interlocuteurs se concentrent sur la clarification ou la négociation du sens d'un mot ou d'une expression. Ces interactions orientent temporairement la conversation vers l'exploration des nuances de sens, menant à des clarification rapides ou à des débats plus approfondis.
Cet article présente le plus grand corpus de WMN librement accessible annoté manuellement. Il comprend des conversations orales dyadiques et multipartites ainsi que des débats en ligne. Notre méthodologie combine une recherche par expressions régulières avec un schéma d'annotation distinguant plusieurs types de WMN, notamment celles déclenchées par une non-compréhension (NONs) ou un désaccord (DINs), ainsi que des phénomènes connexes.
Les analyses révèlent des tendances selon les contextes : les NONs sont plus fréquentes à l'oral, tandis que les DINs dominent en ligne. Cette ressource contribue à l'étude de l'alignement sémantique et à l'identification des WMN.

ABSTRACT

The NeWMe Corpus : A gold standard corpus for the study of Word Meaning Negotiation

Word Meaning Negotiation (WMN) sequences occur when participants focus on clarifying or negotiating the meaning of a word or phrase. These interactions temporarily shift the conversation to explore nuances of meaning, resulting in quick clarification or extended debates.
This paper presents the largest freely available manually annotated corpus of WMNs. It encompasses spoken dyadic and multiparty conversations as well as online debates. Our methodology combines searching for WMNs using regular expressions with a detailed annotation scheme that distinguishes several types of WMNs, such as those triggered by non-understanding (NONs) or disagreement (DINs), as well as related phenomena.
Our analyses reveal distinct patterns across contexts, with NONs prevalent in spoken interactions and DINs dominating online debates. This resource lays a foundation for the study of semantic alignment and the development of automated WMN detection.

MOTS-CLÉS : Négociation du sens des mots, WMN, alignement sémantique, coordination sémantique, linguistique interactionnelle, malentendu.

KEYWORDS: Word meaning negotiation, WMN, semantic alignment, semantic coordination, interactional linguistics, misunderstanding.

Actes des 33ème Conférence sur le Traitement Automatique des Langues Naturelles.
Volume 2 : articles déjà publiés, pages 27–28.
Cette œuvre est mise à disposition sous licence Attribution 4.0 International.

ARTICLE ACCEPTÉ À : Language Resources and Evaluation.
URL : https://link.springer.com/article/10.1007/s10579-026-09907-x

---