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

Polysémie et ambiguité : le cas du verbe Devoir

Anna Colli[1] Delphine Battistelli[1]
(1) Laboratoire MoDyCo (UMR 7114 Université Paris Nanterre-CNRS)
acolli@parisnanterre.fr, dbattist@parisnanterre.fr

RÉSUMÉ

Dans cet article nous présentons une approche qui vise à représenter les valeurs sémantiques (ou sens) des occurrences des marqueurs polysémiques qui ne peuvent pas (ou n'ont pas besoin de) être désambiguïsées. Nous appelons cette tâche "représentation (multi-)sens" et nous nous focalisons ici sur le verbe modal *devoir*. Nous affinons un modèle (CAMEMBERT) pour une tâche de classification multi-label, intégrant une boucle d'active learning pour améliorer le processus d'annotation. En outre, nous démontrons que la combinaison de l' *embedding* de la phrase et de l'occurrence de *devoir* permet au modèle d'obtenir les meilleurs résultats par rapport à une classification basée exclusivement sur l'*embedding* du token. Finalement, nous avons appliqué le modèle à deux corpus distincts en termes de *médium* (oral vs écrit) et genre textuel, démontrant ainsi que l'analyse automatique des valeurs modales de *devoir* permet de mieux les comparer.

ABSTRACT

Polysemy and Ambiguity : the case of the French modal verb Devoir

This article focus on a methodology for representing the semantics of polysemous markers whose meanings cannot (or do not have to) be disambiguated, even in context. We name this task (multi-)sense representation and present here the French modal verb *devoir* as a case study. In order to fine-tune our model (CAMEMBERT) on a multi-label classification task, we implement an active learning loop to enhance the annotation process and we demonstrate that combining global (sentence) and local (token) representations yields the best results. We then applied the model on two distinct corpora, showing that the automatic analysis of *devoir*'s modal senses facilitates comparisons across corpora differing in *medium* (spoken vs. written) or genre.

MOTS-CLÉS : modalité, polysémie, ambiguité, sémantique, classification.

KEYWORDS: modality, polysemy, ambiguity, semantics, classification.

ARTICLE ACCEPTÉ À : ISA 2026@LREC 2026 : The 22nd Joint ACl - ISO Workshop on Interoperable Semantic Annotation and Representation, Palma, Mallorca, Spain, May 11-16, 2026..
URL : https://sigsem.uvt.nl/isa22/ISA22-Proceedings.pdf

Actes des 33ème Conférence sur le Traitement Automatique des Langues Naturelles.
Volume 2 : articles déjà publiés, pages 40–40.
Cette œuvre est mise à disposition sous licence Attribution 4.0 International.