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

Couplage du contexte local et de prototypes sémantiques globaux via une architecture hiérarchique pour l'étiquetage des rôles rhétoriques

Anas Belfathi[1] Nicolas Hernandez[1] Laura Monceaux[1] Warren Bonnard[2]
Mary Catherine Lavissiere[1] Christine Jacquin[1] Richard Dufour[1]
(1) LS2N, UMR CNRS 6004, Nantes Université, Nantes, France
(2) Université de Lorraine, France
anas.belfathi@univ-nantes.fr

RÉSUMÉ

L'étiquetage des rôles rhétoriques (Rhetorical Role Labeling, RRL) consiste à identifier le rôle fonctionnel de chaque phrase dans un document ; il s'agit d'une tâche clé pour la compréhension du discours dans des domaines tels que le droit et la médecine. Bien que les modèles hiérarchiques capturent efficacement les dépendances locales, ils restent limités pour modéliser des caractéristiques globales à l'échelle du corpus. Pour remédier à cette limite, nous proposons deux méthodes fondées sur des prototypes qui intègrent le contexte local à des représentations globales. La régularisation fondée sur des prototypes (Prototype-Based Regularization, PBR) apprend des prototypes souples au moyen d'une perte auxiliaire basée sur la distance afin de structurer l'espace latent, tandis que la modulation conditionnée par prototypes (Prototype-Conditioned Modulation, PCM) construit des prototypes à l'échelle du corpus et les injecte durant l'entraînement et l'inférence. Compte tenu de la rareté des ressources pour le RRL, nous introduisons SCOTUS-LAW, le premier jeu de données d'opinions de la Cour suprême des États-Unis annotées en rôles rhétoriques selon trois niveaux de granularité : catégorie, fonction rhétorique et étape. Des expériences menées sur des benchmarks juridiques, médicaux et scientifiques montrent des améliorations régulières par rapport à de fortes bases de référence, avec des gains d'environ 4 points de Macro-F1 sur les rôles peu fréquents. Nous analysons en outre les implications de ces résultats à l'ère des grands modèles de langage et complétons nos observations par une évaluation experte.

ABSTRACT

Coupling Local Context and Global Semantic Prototypes via a Hierarchical Architecture for Rhetorical Roles Labeling

Rhetorical Role Labeling (RRL) identifies the functional role of each sentence in a document, a key task for discourse understanding in domains such as law and medicine. While hierarchical models capture local dependencies effectively, they are limited in modeling global, corpus-level features. To address this limitation, we propose two prototype-based methods that integrate local context with global representations. **Prototype-Based Regularization (PBR)** learns soft prototypes through a distance-based auxiliary loss to structure the latent space, while **Prototype-Conditioned Modulation (PCM)** constructs corpus-level prototypes and injects them during training and inference. Given the scarcity of RRL resources, we introduce SCOTUS-LAW, the first dataset of U.S. Supreme Court opinions annotated with rhetorical roles at three levels of granularity : *category*, *rhetorical function*,

Actes des 33ème Conférence sur le Traitement Automatique des Langues Naturelles.
Volume 2 : articles déjà publiés, pages 9–10.
Cette œuvre est mise à disposition sous licence Attribution 4.0 International.

and *step*. Experiments on legal, medical, and scientific benchmarks show consistent improvements over strong baselines, with ~ 4 Macro-F1 gains on low-frequency roles. We further analyze the implications in the era of Large Language Models and complement our findings with expert evaluation.

MOTS-CLÉS : Étiquetage des rôles rhétoriques, Architecture hiérarchique, Prototypes sémantiques, Domaine juridique.

KEYWORDS: Rhetorical Role Labeling, Hierarchical Architecture, Semantic Prototypes, Legal Domain.

ARTICLE ACCEPTÉ À : The 19th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2026), Main Conference.
URL : https://arxiv.org/abs/2603.03856