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

SPOT : un jeu de données français annoté pour la détection d'interventions critiques dans les conversations en ligne

Manon Berriche[*,1] Célia Nouri[*, 1, 2] Chloé Clavel[2,3] Jean-Philippe Cointet[1]
(1) Sciences Po médialab, 1 Place Saint-Thomas d'Aquin, 75007 Paris, France
(2) Inria ALMAnaCH, 48 Rue Barrault, 75013 Paris, France
(3) Télécom Paris, 19 Place Marguerite Perey, 91120 Palaiseau, France
(*) Contributions égales
{manon.berriche, celia.nouri}@sciencespo.fr

RÉSUMÉ

Nous présentons SPOT (*Stopping Points in Online Threads*), le premier corpus annoté traduisant le concept de sociologie pragmatique de *point d'arrêt* en une tâche reproductible de traitement automatique du langage. Les points d'arrêt sont des interventions critiques qui interrompent ou redirigent les discussions en ligne. Nous opérationnalisons ce concept en tâche de classification binaire et proposons un guide d'annotation conduisant à un fort accord inter-annotateurs. Le corpus rassemble 43 305 commentaires Facebook en français, issus de publications partageant des URLs signalées comme fausses par des utilisateurs. Chaque commentaire est annoté manuellement et enrichi de métadonnées (article, publication, commentaire parent, page/groupe). Nous comparons des encodeurs CamemBERT fine-tunés à des LLM promptés. Les encodeurs surpassent les LLM de plus de 10 points de F_1, confirmant l'importance de l'apprentissage supervisé pour des tâches sociales francophones. L'intégration du contexte améliore encore les performances (F_1 : 0,75 $\rightarrow$ 0,78).

ABSTRACT

SPOT : An Annotated French Corpus and Benchmark for Detecting Critical Interventions in Online Conversations

We introduce SPOT (*Stopping Points in Online Threads*), the first annotated corpus translating the sociological concept of *stopping point* into a reproducible NLP task. Stopping points are ordinary critical interventions that pause or redirect online discussions. We operationalize this concept as a binary classification task and provide annotation guidelines achieving strong inter-annotator agreement. The corpus contains 43,305 French Facebook comments labeled as stopping point or not, enriched with contextual metadata (shared article, post, parent comment, page/group, media source). Comments are drawn from posts sharing URLs flagged as false by users on public Facebook pages or groups. We compare fine-tuned CamemBERT encoders with prompted large language models (LLMs). Fine-tuned encoders outperform prompted LLMs by over 10 F1 points, highlighting the importance of supervised learning for subtle, non-English tasks. Incorporating contextual metadata further improves encoder F_1 scores from 0.75 to 0.78. The anonymized dataset, guidelines, and code are publicly released.

MOTS-CLÉS : annotation, classification, contexte de publication, conversations en ligne, Facebook, français, interventions critiques, jeu de données, modération, point d'arrêt, réseaux sociaux, sociologie pragmatique, traitement automatique du langage naturel.

KEYWORDS: annotation, classification, context-aware NLP, critical interventions, dataset, Face-

Actes des 33ème Conférence sur le Traitement Automatique des Langues Naturelles.
Volume 2 : articles déjà publiés, pages 43–44.

book, French, online conversations, online moderation, social media, stopping point, pragmatics, publication context.

ARTICLE ACCEPTÉ À : International Conference on Language Resources and Evaluation (LREC) 2026.
URL : https://arxiv.org/abs/2511.07405
