



HAL
open science

Exploration d'approches hybrides pour la lisibilité : expériences sur la complémentarité entre les traits linguistiques et les transformers

Rodrigo Wilkens, Patrick Watrin, Rémi Cardon, Alice Pintard, Isabelle
Gribomont, Thomas François

► To cite this version:

Rodrigo Wilkens, Patrick Watrin, Rémi Cardon, Alice Pintard, Isabelle Gribomont, et al.. Exploration d'approches hybrides pour la lisibilité : expériences sur la complémentarité entre les traits linguistiques et les transformers. 35èmes Journées d'Études sur la Parole (JEP 2024) 31ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2024) 26ème Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL 2024), Jul 2024, Toulouse, France. pp.8-8. hal-04622996

HAL Id: hal-04622996

<https://inria.hal.science/hal-04622996v1>

Submitted on 28 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Exploration d’approches hybrides pour la lisibilité : expériences sur la complémentarité entre les traits linguistiques et les transformers.

Rodrigo Wilkens¹ Patrick Watrin¹ Rémi Cardon¹
Alice Pintard¹ Isabelle Gribomont^{1,2} Thomas François¹

(1) CENTAL, IL&C, Université catholique de Louvain, Belgique

(2) Royal Library of Belgium (KBR)

{rodrigo.wilkens, patrick.watrin, remi.cardon, alice.pintard,
isabelle.gribomont, thomas.francois}@uclouvain.be

RÉSUMÉ

Les architectures d’apprentissage automatique reposant sur la définition de traits linguistiques ont connu un succès important dans le domaine de l’évaluation automatique de la lisibilité des textes (ARA) et ont permis de faire se rencontrer informatique et théorie psycholinguistique. Toutefois, les récents développements se sont tournés vers l’apprentissage profond et les réseaux de neurones. Dans cet article, nous cherchons à réconcilier les deux approches. Nous présentons une comparaison systématique de 6 architectures hybrides (appliquées à plusieurs langues et publics) que nous comparons à ces deux approches concurrentes. Les diverses expériences réalisées ont clairement mis en évidence deux méthodes d’hybridation : *Soft-Labeling* et concaténation simple. Ces deux architectures sont également plus efficaces lorsque les données d’entraînement sont réduites. Cette étude est la première à comparer systématiquement différentes architectures hybrides et à étudier leurs performances dans plusieurs tâches de lisibilité.

ABSTRACT

Exploring hybrid approaches to readability : experiments on the complementarity between linguistic features and transformers¹

Linguistic features have a strong contribution in the context of the automatic assessment of text readability (ARA). They have been one of the anchors between the computational and theoretical models. With the development in the ARA field, the research moved to Deep Learning (DL). In an attempt to reconcile the mixed results reported in this context, we present a systematic comparison of 6 hybrid approaches along with standard Machine Learning and DL approaches, on 4 corpora (different languages and target audiences). The various experiments clearly highlighted two rather simple hybridization methods (soft label and simple concatenation). They also appear to be the most robust on smaller datasets and across various tasks and languages. This study stands out as the first to systematically compare different architectures and approaches to feature hybridization in DL, as well as comparing performance in terms of two languages and two target audiences of the text, which leads to a clearer pattern of results.

MOTS-CLÉS : évaluation de la lisibilité, modèles hybrides, soft-labeling.

KEYWORDS: readability assessment, hybrid models, soft-labeling.

1. Référence de la publication : Wilkens, R., Watrin, P., Cardon, R., Pintard, A., Gribomont, I., & François, T. (2024, March). Exploring hybrid approaches to readability : experiments on the complementarity between linguistic features and transformers. In *Findings of the Association for Computational Linguistics : EACL 2024* (pp. 2316-2331).