



The Contribution of Selected Linguistic Markers for Unsupervised Arabic Verb Sense Disambiguation

ASMA DJAIDRI, Laboratory of Research in Artificial Intelligence (LRIA), USTHB, Algeria and Research Center for Scientific and Technical Information CERIST, Algeria

HASSINA ALIANE, Research Center for Scientific and Technical Information CERIST, Algeria

HAMID AZZOUNE, Laboratory of Research in Artificial Intelligence (LRIA), USTHB, Algeria

Word sense disambiguation (WSD) is the task of automatically determining the meaning of a polysemous word in a specific context. Word sense induction is the unsupervised clustering of word usages in a different context to distinguish senses and perform unsupervised WSD. Most studies consider function words as stop words and delete them in the pre-processing step. However, function words can encode meaningful information that can help to improve the performance of WSD approaches. We propose in this work a novel approach to solve Arabic verb sense disambiguation that is based on a preposition-based classification that is used in an automatic word sense induction step to build sense inventories to disambiguate Arabic verbs.

However, in the wake of the success of neural language models, recent works obtained encouraging results using BERT pre-trained models for English-language WSD approaches. Hence, we use contextualized word embeddings for an unsupervised Arabic WSD that is based on linguistic markers and uses sentence-BERT Transformer pre-trained models, which yields encouraging results that outperform other existing unsupervised neural AWSW approaches.

CCS Concepts: • **Computing methodologies** → **Artificial intelligence**; **Natural language processing**; **Lexical semantics**;

Additional Key Words and Phrases: Natural language processing, word sense disambiguation, word sense induction, Arabic language, linguistic markers, clustering, contextualized word embeddings, SBERT

ACM Reference format:

Asma Djaidri, Hassina Aliane, and Hamid Azzoune. 2023. The Contribution of Selected Linguistic Markers for Unsupervised Arabic Verb Sense Disambiguation. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 22, 8, Article 208 (August 2023), 23 pages.

<https://doi.org/10.1145/3605777>

This work has been supported by the Research Center for Scientific and Technical Information (CERIST), the Laboratory of Research in Artificial Intelligence (LRIA), and the Algerian Directorate General for Scientific Research and Technology Development (DGRSTD), Algeria.

Authors' addresses: A. Djaidri (corresponding author), Laboratory of Research in Artificial Intelligence (LRIA), University of Science and Technology Houari Boumediene, BP 32 Bab Ezzouar, Algiers, Algeria, 16111 and Research Center for Scientific and Technical Information CERIST, 05, Rue Des 3 Freres Aissou, Ben Aknoun, Algiers, Algeria; email: adjaidri@usthb.dz; H. Aliane, Research Center for Scientific and Technical Information CERIST, 05, Rue Des 3 Freres Aissou, Ben Aknoun, Algiers, Algeria; email: haliane@cerist.dz; H. Azzoune, Laboratory of Research in Artificial Intelligence (LRIA), USTHB, BP 32 Bab Ezzouar, Algiers, Algeria; email: hazzoune@usthb.dz.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2375-4699/2023/08-ART208 \$15.00

<https://doi.org/10.1145/3605777>