

Measuring diachronic language distance using perplexity: Application to English, Portuguese, and Spanish

Título Measuring diachronic language distance using perplexity: Application to English, Portuguese, and Spanish

Autores José Ramom Pichel Campos, [Pablo Gamallo Otero](#), Iñaki Alegria Loinaz

Tipo Artículo de revista

Fonte  [Natural Language Engineering](#), Cambridge Journals Online, Vol. 26, No. 4, pp. 433-454 , 2020.

Rank  [Provisionally ranked Q1 in Language and Linguistics by SJR 2019](#)

ISSN 1351-3249

DOI [10.1017/S1351324919000378](https://doi.org/10.1017/S1351324919000378)


Abstract The objective of this work is to set a corpus-driven methodology to quantify automatically diachronic language distance between chronological periods of several languages. We apply a perplexity-based measure to written text representing different historical periods of three languages: European English, European Portuguese, and European Spanish. For this purpose, we have built historical corpora for each period, which have been compiled from different open corpus sources containing texts as close as possible to its original spelling. The results of our experiments show that a diachronic language distance based on perplexity detects the linguistic evolution that had already been explained by the historians of the three languages. It is remarkable to underline that it is an unsupervised multilingual method which only needs a raw corpora organized by periods.

Palabras clave natural language processing, language distance, perplexity, historical linguistics

LIGAZÓNS

 [Versión da editorial](#)

DESCARGAS

 [Referencia BibTex](#)

PROXECTOS DE INVESTIGACIÓN

eRISK: Tecnoloxías para a predición temperá de signos relacionados con trastornos psicolóxicos

PROGRAMAS CIENTÍFICOS

Tecnoloxías da Linguaxe Natural