

LEXICAL DIFFERENTIATION OF FLAMENCO GENRES USING MACHINE LEARNING

Pablo Rosillo-Rodes,* David Sánchez, and Maxi San Miguel

Institute for Cross-Disciplinary Physics and Complex Systems IFISC (UIB-CSIC),

Campus Universitat de les Illes Balears, E-07122, Palma, Spain

(Dated: June 7, 2023)

Music is a cultural practice showing a broad variety of forms. Despite this heterogeneity, there exist patterns that help define genres with proper characteristics. How can musical genres be distinguished and how can we identify the different forms in an objective manner? Here, we argue that an automatic analysis of lexical variation within song lyrics allows us to distinguish among flamenco genres (*palos*). Using machine learning techniques, we train a model using a corpus of about 3400 lyrics with detailed metadata (palo, authorship, origin, year). In addition to traditional elements such as rhythm, tonality, and melodies, our study demonstrates that lexical features describe quite well flamenco genres. By analyzing the weights assigned by the model to each word and word combinations within each palo, we investigate the distinctive lexicons associated with each flamenco style. The results obtained thus far indicate that the model successfully identifies the palos of numerous flamenco song lyrics. To enhance result interpretation and gain deeper insights into the significance and relevance of lexicon in flamenco differentiation, we determine which topics and lexical fields characterize each genre the most. For instance, we find that while *soleares* are characterized by deep and existential concerns, *alegrías* mainly refer to Cádiz and its places and traditions. Thus, our approach sheds light on the intricate relationship between lexical choices and flamenco forms and shows the strong capability of automatic approaches to objectively analyze music genres in general.

* prosillo@ifisc.uib-csic.es