Drawing on Research on Explainability of Dialect Classifiers to Extract Greek Dialect Features

Dana Roemling[1,4] d.roemling@bham.ac.uk
Erofili Psaltaki[2,4] erofili.psaltaki@helsinki.fi
Yves Scherrer[3,4] yves.scherrer@ifi.uio.no

Aleksandra Miletić[4] aleksandra.miletic@helsinki.fi

[1]University of Birmingham, [2]University of Crete, [3]University of Oslo, [4]University of Helsinki

Although interest in computational dialectology is growing, studies are often relying on data collected through classic dialectological approaches, i.e. surveys and interviews, while approaches working with unstructured data frequently use pre-selected feature sets to analyse data (Grieve, 2015). Within Natural Language Processing unstructured data is used, but even if approaches work on classifying dialect data, the features underlying these predictions are not known given the nature of machine learning methods (see Aepli et al., 2023).

Xie et al. (2024) were the first to propose two approaches that make use of NLP methods to interpret which features are used for dialect classification. One of these is a post-hoc method in which they use a leave-one-out (LOO) approach to understand which features are most impactful for classification. The impact score is calculated by using the change in probability for classification if a token is left out of the input sentence. We employ the same approach and apply it to a corpus of Greek dialect data.

The corpus we use is the GRDD (Chatzikyriakidis et al., 2023), consisting of raw text data from four dialects of Modern Greek: Cretan, Pontic, Northern Greek and Cypriot Greek. Employing both `xlm-roberta-base` and `bert-base-greek-uncased` models and fine-tuning them on the dialect data, the dialect classifiers outperform the chance baseline by a large margin with accuracies of 91.39% and 91%. We find that the LOO approach extracts features that match the predicted dialect classes and that extracted features are indeed dialectal. For example Cretan *χέρα* (Engl. *hand*), which has a different grammatical gender (feminine) in comparison to its Modern Greek equivalent (neuter), a characteristic phenomenon of the Cretan dialect, and Pontic *επέμναν* (Engl. *(they) remained*), which shows a typical morphological phenomenon of the Pontic verbal system, where the -να suffix is used to indicate present perfect active.

Aepli, N., Çöltekin, C., Van Der Goot, R., Jauhiainen, T., Kazzaz, M., Ljubesic, N., North, K., Plank, B., Scherrer, Y., & Zampieri, M. (2023). Findings of the VarDial Evaluation Campaign 2023. *VarDial 2023*, 251–261.

Chatzikyriakidis, S., Qwaider, C., Kolokousis, I., Koula, C., Papadakis, D., & Sakellariou, E. (2023). GRDD: A Dataset for Greek Dialectal NLP. *arXiv:2308.00802*.

Grieve, J. (2015). Dialect variation. In D. Biber & R. Reppen (Eds.), *The Cambridge Handbook of English Corpus Linguistics* (pp. 362–380). CUP.

Xie, R., Ahia, O., Tsvetkov, Y., & Anastasopoulos, A. (2024). Extracting Lexical Features

from Dialects via Interpretable Dialect Classifiers. *arXiv:2402.17914*.