Maltese morphophonology – a merger of two worlds

Jessica Nieder & Fabian Tomaschek

Heinrich-Heine-Universität Düsseldorf, Eberhard Karls University of Tübingen

From a linguistic perspective, the Semitic language Maltese is situated between two worlds: Romance, especially Sicilian Italian, vs. Semitic, especially Maghrebi Arabic varieties (Comrie, 2007). The "hybrid" morphophonological system (Borg & Gatt, 2017; Mayer et al., 2013) is, for example, visible in the Maltese plural formation: A variety of concatenative *sound* plurals, such as *omm-ommiijiet* 'mothers', characterised by additonal phonological material at the end of the word form contrasts with a variety of non-concatenative *broken* plurals, such as *kelb-klieb* 'dogs', characterised by a stem alternation.

In this study we present a computational modeling approach of this hybrid morphophonological system by using Naive Discriminative Learning (NDL), an error-driven learning algorithm predicting so-called outcomes based on form-based cues (e.g. phones), to classify Maltese nouns as belonging to the language families Semitic vs. Non-Semitic. NDL was trained on a data set of 2377 Tunisian nouns from Gugliotta and Dinarelli (2020), 2377 randomly selected Italian nouns from the `subtlex-it` corpus[1] and 2377 randomly selected German nouns from the `CELEX` corpus (Baayen et al., 1996). The model was then tested on 6511 Maltese singular and sound and broken plural nouns from Nieder et al. (2021). We used the same quasi-phonetic transcription based on SAMPA for all three languages. 2-phones were used as phonological cues, the language families `semitic` vs. `non-semitic` were used as outcome.

If Maltese is indeed a merger the two linguistic worlds `Romance & Semitic` and the knowledge of native speakers of this merged system is based on error-driven learning, we expect NDL to show high classification accuracies for a training based on Tunisian and Italian data and lower accuracies for a training based on Tunisian and German data. Table 1 shows the accuracies for the tested models. The evaluation on the training data (first row) showed an excellent overall accuracy of $>$97%, indicating that NDL is able to capture the phonological differences between Semitic (Tunisian) and Non-Semitic (Italian, German) languages and to accurately classify nouns accordingly.

For the Maltese test data (second row), we found that the network predicts the origin of the word forms with a high accuracy of 72% for a training based on Tunisian and Italian. For a training on Tunisian and German, the accuracy drops to 59% and thus confirms our initial hypothesis.

Interestingly, when having a closer look at the Maltese test nouns that were to be classified, we discovered that within the group of nouns with a Semitic origin, NDL showed overall high classification accuracies with $>$80% across all number categories (singular, broken, sound). For nouns with a Non-Semitic origin, NDL was unsure how to classify singulars and broken plural nouns. This uncertainty is a direct reflection of the hybrid morphological system with nouns such as *denfil* (sg.) - *dniefel* (broken pl.), etymologically related to Italian *delfino* 'dolphin', following a non-concatenative plural pattern despite having a Non-Semitic origin.

Our results show that it is possible to classify Maltese nouns using NDL with a training based on Tunisian and Italian nouns only. The model successfully applied the language-specific knowledge gained during the training to Maltese nouns, thus confirming the hybrid status of the Maltese morphophonological system.

[1] downloaded from http://crr.ugent.be/programs-data/subtitle-frequencies

doi:10.11576/pundp2022-1018

	Italian+Tunisian	German + Tunisian
accuracy training	97.35	97.23
accuracy test	71.59	58.96

Table 1: Accuracies of NDL models trained on Italian + Tunisian data and trained on German + Tunisian data.

References

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1996). The celex lexical database (cd-rom).

Borg, C., & Gatt, A. (2017). Morphological analysis for the Maltese language: The challenges of a hybrid system. *Proceedings of the Third Arabic Natural Language Processing Workshop*, 25–34. https://doi.org/10.18653/v1/W17-1304

Comrie, B. (2007). Maltese and the World Atlas of Language Structures. *Introducing Maltese Linguistics: Selected Papers from the 1st International Conference on Maltese Linguistics* (pp. 3–15). John Benjamins Publishing.

Gugliotta, E., & Dinarelli, M. (2020). TArC: Incrementally and Semi-Automatically Collecting a Tunisian Arabish Corpus. *LREC*.

Mayer, T., Spagnol, M., & Schönhuber, F. (2013). Fixing the broken plural in Maltese. *Perspectives on Maltese Linguistics* (pp. 129–158). De Gruyter.

Nieder, J., Chuang, Y.-Y., van de Vijver, R., & Baayen, R. (2021). Comprehension, Production and Processing of Maltese Plurals in the Discriminative Lexicon. https://doi.org/https://doi.org/10.31234/osf.io/rkath

doi:10.11576/pundp2022-1018