LED-A: a web app for measuring distances in the sound components among local dialects

Wilbert Heeringa, Vincent van Heuven, Hans Van de Velde

In order to explore a dialect landscape and to reveal spatial patterns dialectologists measure linguistic distances among local dialects. Kessler (1995) found the Levenshtein distance very suitable for measuring linguistic distances among Irish Gaelic dialects using phonetic transcriptions. Other scholars followed him by applying the method to dialects from other language families. The Levenshtein distance is a numerical value of the cost of the least expensive set of insertions, deletions or substitutions that would be needed to transform one string into another (Kruskal 1999). This distance measure is available in the online web app Gabmap (Nerbonne et al. 2011, Leinonen et al. 2015).

In this poster we present a new web app – LED-A - that shares the features of Gabmap that are often used by dialectologists such as cluster analysis, multidimensional scaling, beam maps, area maps and RGB maps (maps that visualize the dialect landscape as a continuum). Our web app, however, differs in the design of the user interface, and the app includes features that are not found in Gabmap.

As for the user interface, we aimed to maximize flexibility, user-friendliness and intuitiveness. As for the features, four different variants of the Levenshtein distance are readily available without the need of coding a configuration file. Among others, PMI Levenshtein is included, which learns segment distances based on the alignments that are generated by the algorithm (see Wieling et al. 2009, Wieling 2012). Both aggregated and individual word distances can be obtained on the basis of whole words or only on the basis of vowel or consonant substitutions or indels.

When creating maps, it is sufficient to upload the coordinates of the places, i.e. no coordinates that constitute the outline are required. Multiple map backgrounds can be chosen from.

References

Kessler, B. (1995). Computational dialectology in Irish Gaelic. In *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics*, pages 60–67, Dublin. EACL.

Kruskal, J. B. 1999. "An overview of sequence comparison." In *Time Warps, String edits, and Macromolecules. The Theory and Practice of Sequence Comparison* edited by D. Sankoff, and J. Kruskal, 2nd ed., 1–44. Stanford: Center for the Study of Language and Information. 1st edition appeared in 1983.

Leinonen, T., Çöltekin, Ç., & Nerbonne, J. (2016). Using Gabmap. *Lingua,* 178, 71-83.

Nerbonne, J., Colen, R., Gooskens, C., Kleiweg, P., & Leinonen, T. (2011). Gabmap-a web application for dialectology. *Dialectologia: revista electrònica*, 65-89.

Wieling, M., Prokić, J., and Nerbonne, J. (2009). Evaluating the pairwise alignment of pronunciations. In Borin, L. and Lendvai, P., editors, Proceedings of the EACL 2009 Workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education, (LaTeCH - SHELT&R 2009). Workshop at the 12th Meeting of the European Chapter of the Association for Computational Linguistics. Athens, 30 March 2009, pages 26—34.

Wieling, M. B. (2012). A quantitative approach to social and geographical dialect variation. PhD thesis University of Groningen.