Tracking the Evolution of the English Language as a Network with Temporal Clustering

Oliver Blanthorn[1] and Eva Navarro-López[2]

[1]School of Computer Science, The University of Manchester, Manchester, United Kingdom,
oliver.blanthorn@manchester.ac.uk
[2]School of Environment, Education and Development, The University of Manchester, Manchester, United Kingdom, eva.navarro@manchester.ac.uk

The English language has existed for approximately 1500 years and has therefore been exposed to massive societal and technological change. Previous work has shown that this exogenous change is visible in the language [1, 2]. We hope to understand this change more clearly through representing the language as a dynamic network, and performing temporal clustering on it.

We will present a subset of the language as a multi-layer dynamic network with adjacency tensor

$$A^{\alpha t}_{i \leftarrow j} \in \{0, 1\},$$

where 0 will denote the absence of a link and 1 will denote the presence of a link; $i \leftarrow j$ will specify a link from the word j to the word i, $1000 < t < 2020$ will denote the year and $\alpha \in \{\text{synonymy}, \text{hierarchical}\}$ will specify the type of relationship between two words. Note that in this work, each distinct meaning of a word with a single spelling will have a separate node; i.e. homonyms will be spread out across distinct nodes. For example, "cat" and "moggie" would have two directed links between them in the synonymy layer; both would have a directed link to "felis domesticus" in the hierarchical layer. Each word will have a range in t for which it is in use; it will have no incoming and outgoing links outside of this time. The words and links in the network will be generated from a database created by domain experts.

We will perform non-negative tensor factorisation on this network [3] to extract temporal communities of words. We hope to be able to see the rise and fall of terminology related to specific fields, such as those used in sailing, the automotive industry, and computing; we also hope to see changes from exposure to foreign influence, such as the Columbian exchange. We believe this work will improve our understanding of how groups of words enter and leave the English language, and how the rate at which that evolution occurs has changed over time.

Acknowledgements

References

[1] Cuskley, C. F., Pugliese, M., Castellano, C., Colaiori, F., Loreto, V., & Tria, F. (2014). Internal and external dynamics in language: Evidence from verb regularity in a historical corpus of English. *PLoS ONE*, 9(8).

[2] Perc, M. (2012). Evolution of the most common English words and phrases over the centuries. *Journal of the Royal Society Interface*, 9(77), 3323–3328.

[3] Gauvin, L., Panisson, A., & Cattuto, C. (2014). Detecting the community structure and activity patterns of temporal networks: A non-negative tensor factorization approach. *PLoS ONE*, 9(1).