

5 Conclusions and future work

We have presented the main tasks we carried out to obtain the data for an academic writing aid tool. Next, we will explore the transfer strategies for the automatic identification of ALCs in several languages. We start from the hypothesis that a cross-linguistic language model trained to identify the formulas in Spanish could recognise expressions with similar characteristics in Basque. If the results obtained with this strategy are adequate, we could, on the one hand, automatically obtain new formulas in both languages in other corpora and, on the other hand, identify formulas in Basque that could be mapped to those in Spanish. Pending the results of the experiments with distributional semantics techniques, we are making progress in the design of the tool, which must meet two requirements: 1) provide onomasiological access by discursive function; 2) include a field of warnings where examples will be provided as correction models.

Acknowledgments

This work has been supported by the Xunta de Galicia, through grant ED431C 2020/11, and the Spanish Ministry of Science and Innovation through projects PID2019-109683GB-C21 and PID2019-109683GB-C22. I would like to thank Olga Zamaraeva for her valuable and constructive suggestions.

References

- Alonso-Ramos, M., M. García-Salido, and M. Garcia. 2017. Exploiting a corpus to compile a lexical resource for academic writing: Spanish lexical combinations. In I. Kosem, et al. (eds.), *Electronic Lexicography in the 21st Century. Proceedings of eLex 2017 Conference*, pages 571–586. Lexical Computing Brno.
- Aranzabe, M.J., A. Gurrutxaga, and I. Zabala. 2022. Compilación del corpus académico de novelas en euskera HARTAvas y su explotación para el estudio de la fraseología académica. *Procesamiento del Lenguaje Natural*
- García, M., M. García Salido, and M. Alonso-Ramos. 2019. Weighted compositional vectors for translating collocations using monolingual corpora. In *Computational and Corpus-Based Phraseology (EUROPHRAS 2019)*. Lecture Notes in Artificial Intelligence, 11755. pages 113-128. Springer.
- García-Salido, M., M. Garcia, M. Villayandre, and M. Alonso-Ramos. 2018. A Lexical Tool for Academic Writing in Spanish based on Expert and Novice Corpora. In N. Calzolari et al. (eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pages 260–265. Miyazaki.
- García-Salido M., M. Garcia, and M. Alonso-Ramos. 2019. Identifying lexical bundles for an academic writing assistant in Spanish. In G. Corpas Pastor and R. Mitkov eds., *Computational and Corpus-Based Phraseology*. Europhras 2019, Proceedings, volume 11755 of Lecture Notes in Artificial Intelligence, pages 144-158. Springer.
- García-Salido, M. (2021). Compiling an Academic Vocabulary List of Spanish. Available at: <https://doi.org/10.13140/RG.2.2.27681.33123>
- Gurrutxaga, A., I. Alegria. 2011. Automatic extraction of NV expressions in Basque: basic issues on cooccurrence techniques. In *Proceedings of the Workshop on Multiword Expressions: from parsing and generation to the real world*, 2–7. Association for Computational Linguistics.
- Hyland, K., and P. Shaw, P. (eds.). 2016. *The Routledge Handbook of English for Academic Purposes*. London, Routledge.
- Zabala, I., Aranzabe, M^a J., and Aldezabal, I. 2021. Retos actuales del desarrollo y aprendizaje de los registros académicos orales y escritos del euskera. *Círculo de Lingüística Aplicada a la Comunicación* 88: 31-50.