

Towards automatic essay scoring of Basque language texts from a rule-based approach based on curriculum-aware systems

Text classification of writing and reading materials is laborious and sometimes hard to do manually. Teachers that do not have a linguistic background do not feel confident in this task, but in some languages, researchers can use automatic text classification tools to point safely to some measures (TTR, PoS-based measures...). However, this task is difficult to address for low resourced-languages. Using these measures could be difficult for teachers as far as these measures do not follow the language curriculum or the learning stage of the student. If automated systems can describe the curriculum or the learning stage of the student in a way that the teachers can understand or employ, this would be very useful, and teachers will have an additional source of information where they could offer more adapted materials and teaching.

This work aims to explore rule-based models (Bick & Didriksen, 2015) to grade written learner texts. The objective is, in principle, to lighten the burden on teachers in the correction task. Previous work (Zupanc & Bosnic, 2016) emphasises the role of automatic systems to help teachers: “Automated essay evaluation represents a practical solution to a time-consuming, labour-intensive and expensive activity of manual grading of student's essays”. Furthermore, this approach could help to define grading criteria that is more standard and less evaluator-dependent.

In particular, we focused on all the *Common European Framework of Reference for Languages* (CEFR) from B1 to C2. To do that, we followed a curriculum-based and language-based approach. We adapted the adult Basque learner curriculum (HEOC, HABE 2015), the Basque curriculum based on the CEFR, to create a rule-based grammar for Basque. Using this grammar, we labelled all the linguistic features corresponding to each of the aforementioned levels.

There have been other studies of automatic classification for the Basque language but from different approaches (Castro et al., 2008; Zipitria et al., 2010; Zipitria et al., 2011; Olaizola, 2022). We expand on these works and study how automatic text classifiers can benefit from a *linguagrammar*, the formalisation of the linguistic expressions described in the HEOC.

Other similar works include a system for the Arabian language (Alqahtani & Alsaif, 2019) as well as feature-based machine learning approaches for Estonian (Vajjala & Loo, 2014) and monolingual, cross-lingual, and multilingual classification with three languages: German, Czech and Italian (Volodina et al. 2016). For Estonian, the best model reported by Vajjala & Loo (2014) reaches a prediction accuracy of 79%.

Regarding the automated essay-scoring task, Lin et al. (2021) conducted an automatic assessment using Automatic Essay Scoring systems (AES). Gaillat et al. (2022), in their work, showed that early approaches were rule-based, but latter systems relied on probabilistic models based on Natural Language Processing (NLP) methods that exploit the corpus of learners. In their proposal, they presented a machine learning approach.

In our approach, the Basque CG3 grammar contains 296 rules that *add* language level information. These rules are based on the linguistic indicators described for each level in the HEOC. The grammar is cumulative, meaning that each level encompasses all the features and linguistic phenomena of the levels below it. This entails the need to find indicators that are determinant when classifying a text at the appropriate level.

We applied this grammar to a student corpus, annotating each instance of the linguistic forms collected in HEOC for each level. We found which rule is relevant to distinguish each CEFR level using standard feature selection algorithms (that evaluate the weight of an attribute by measuring the information gained concerning to the class) and machine learning techniques. Results show that for C1 level, the most important linguistic features are related to the verbal system,

morphological cases, connectors and some subordinate clauses, whereas in B2 level, the most relevant features correspond to the use of a series of particles.

We plan to combine this curricular knowledge about linguistic features with a deep learning approach using large language models. The aim is to provide more information on the output, and point to the factors that contributed to the automatic scoring, creating a more explainable system and provides some type of feedback that the teachers and students can use in their language development.

As for the future, we plan to complete the grammar with two modules: a) the mal-rule module: to capture the incorrect use of the linguistic features and b) the correct rule module: to determine the correctness of linguistic constructions for each level. The rule-based module will be integrated into the Common Text Analysis Platform (Chen & Meurers, 2016), as another curriculum-based input to general complexity analysis.

References

- Alqahtani, A., & Alsaif, A. (2019, December). Automatic evaluation for Arabic essays: a rule-based system. In *2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)* (pp. 1-7). IEEE.
- Bick, E., & Didriksen, T. (2015, May). Cg-3—beyond classical constraint grammar. In *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA 2015)* (pp. 31-39).
- Castro-Castro, D., Lannes-Losada, R., Maritxalar, M., Niebla, I., Pérez-Marqués, C., Álamo-Suárez, N. C., & Pons-Porrata, A. (2008). A multilingual application for automated essay scoring. In *Advances in Artificial Intelligence-IBERAMIA 2008: 11th Ibero-American Conference on AI, Lisbon, Portugal, October 14-17, 2008. Proceedings 11* (pp. 243-251). Springer Berlin Heidelberg.
- Council of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Chen, X., & Meurers, D. (2016). CTAP: A web-based tool supporting automatic complexity analysis.
- Gaillat, T., Simpkin, A., Ballier, N., Stearns, B., Sousa, A., Bouyé, M., & Zarrouk, M. (2022). Predicting CEFR levels in learners of English: the use of microsystem criterial features in a machine learning approach. *ReCALL*, 34(2), 130-146.
- HABE (2015). *Helduen Euskalduntzearen Oinarrizko Curriculum*. Donostia: HABE.
- Lim, C. T., Bong, C. H., Wong, W. S., & Lee, N. K. (2021). A comprehensive review of automated essay scoring (AES) research and development. *Pertanika Journal of Science & Technology*, 29(3), 1875-1899.
- Olaizola, X. A. (2022). Euskarazko testuen komunikagaitasun-maila automatikoki sailkatzeko lehendabiziko urratsak. *Hizpide: helduen euskalduntzearen aldizkaria*, 40(99), 3.
- Quintian, M. I., & Agirre, M. O. (2022). Hizkuntza-ikasleen testu-corpus etiketatuaren analisia eta interpretazioa B2 eta C1 mailetan. *Hizpide: helduen euskalduntzearen aldizkaria*, 40(100), 6.
- Vajjala, S., & Loo, K. (2014, November). Automatic CEFR level prediction for Estonian learner text. In *Proceedings of the third workshop on NLP for computer-assisted language learning* (pp. 113-127).
- Volodina, E., Pilán, I., & Alfter, D. (2016). Classification of Swedish learner essays by CEFR levels. *CALL communities and culture—short papers from EUROCALL, 2016*, 456-461.
- Zipitria, I., Elorriaga, J. A., & Arruarte, A. (2011). Corpus-based performance analysis for automatically grading use of language in essays. In *Artificial Intelligence in Education: 15th International*

Conference, AIED 2011, Auckland, New Zealand, June 28–July 2011 15 (pp. 591-593). Springer Berlin Heidelberg.

Zipitria, I., Arruarte, A., & Elorriaga, J. A. (2010). Automatically grading the use of language in learner summaries. In *Proceedings of the 18th international conference on computers in education, Putrajaya, Malaysia* (pp. 46-50).

Zupanc, K., & Bosnic, Z. (2016). Advances in the field of automated essay evaluation. *Informatica*, 39(4).