

Adaptation of a complexity tool for grading student essays in Basque / Konplexutasuna neurtzen duen tresna baten egokitzapena ikasleen euskarazko testuak sailkatzeko

Proposer(s) / Proposatzailea(k):

Mikel Iruskieta

Contact / Kontaktua:

mikel.iruskieta@ehu.eus

Description / Deskribapena

A hot topic in the Basque Country is that students (%25-%35 personal estimations) don't achieve linguistic proficiency in the Basque language. For example, B2 when they start at the university and C1 at their graduation, but they get a certificate of C1 by law. That is causing a big preoccupation at different levels (university departments, teachers, some sections of the university administration). An interesting research question can be to analyze a corpus we had compiled and test if we are capable to grade automatically or to detect problems or typical errors of these students.

The corpus is composed of:

- Readings, writings and speaking corpora of learners of teaching schools at the entrance of the university and the same tests at the end of their university studies (graduation).

The student can try to analyse writings and speaking tests of students to detect if they have the proper language proficiency (classify their writings or speakings as B1, B2 or C1) and detect the typical errors of these learners. This can be done with CTAP. The student can create or adapt CTAP (Chen and Meurers (2016)): <https://www.aclweb.org/anthology/W16-4113>.

After or before the student could adapt VIEW (Reynolds et al. 2014) for Basque and propose some exercises for students.

Goals / Helburuak

Adapt a complexity tool developed for English or German to grade student essays in Basque

Requirements / Betebeharrak

Computer scientists

Basque language: B2

Framework / Esparrua

In collaboration with Detmar Meurers (<http://www.sfs.uni-tuebingen.de/~dm/>) University of Tübingen

If the student decides to follow with this study, she or he will joint in an Innovative Training Networks (ITN) and have financial to finish the PhD.

Tasks and plan / Atazak eta plana

1. Study and understand the following two systems:

VIEW: <http://sifnos.sfs.uni-tuebingen.de/VIEW/index.jsp?content=home>

CTAP: <http://samos.sfs.uni-tuebingen.de:8080/ctapweb/#signin>

2. Design of an automatic system for Basque.

3. Evaluate the system.

References:

Chen, X. Meurers, D. *CTAP: A Web-Based Tool Supporting Automatic Complexity Analysis* Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CL4LC), COLING 2016, Osaka, Japan.

González Dios, I. (2016). Euskarazko egitura sintaktiko konplexuen analisirako eta testuen sinplifikazio automatikorako proposamena. EHU doktore-tesia.

Reynolds, R., Schaf, E., & Meurers, D. (2014). A VIEW of Russian: Visual Input Enhancement and adaptive feedback. In *Proceedings of the third workshop on NLP for computer-assisted language learning* (pp. 98-112).

Vajjala, S., & Meurers, D. (2012). On improving the accuracy of readability classification using insights from second language acquisition. In *Proceedings of the seventh workshop on building educational applications using NLP* (pp. 163-173). Association for Computational Linguistics.