Towards an infrastructure for the semi-automatic development of corpus-based language exercises.

Katrien Depuydt, Jesse de Does (Dutch Language Institute)
{katrien.depuydt,jesse.dedoes}@ivdnt.org

We report on the development of an environment for the semi-automatic creation of corpus-based language exercises using linguistically annotated corpora and lexical resources, both dedicated CEFR-lexica[1] and general-purpose lexical resources[2]. The aim is to reduce manual work, to produce more natural exercises "in context", and to allow for more efficient creation of exercise material adjusted to specific text types, e.g. "easy" material[3] or sports-related texts.

We have created a working environment, which combines the BlackLab[4] corpus exploration environment and the lexical resources in a Vue.js[5]-based frontend. A language learning specialist selects vocabulary and choses a type of exercise (cloze, scrambling, …), corresponding to a parametrized corpus query suitable for BlackLab. The exercises constructed automatically from the query results can be previewed, pruned and adjusted by the specialist.

The environment has been used to create exercises for the prototype of a mobile app developed in cooperation by Oefenen.nl[6] and Game Architect[7].

---

[1] notably NT2Lex, https://cental.uclouvain.be/cefrlex/nt2lex/
[2] Especially the morphosyntactic lexicon GiGaNT-Molex,
https://taalmaterialen.ivdnt.org/download/gigant-molex2-0/
[3] In our case: the Start!-krant, https://www.eenvoudigcommuniceren.nl/kranten/start-krant
[4] https://inl.github.io/BlackLab/
[5] https://vuejs.org/
[6] https://oefenen.nl/
[7] https://gamearchitect.eu/