From texts to tasks: Text features and cognition

When comparing writing tasks, the choice of the source texts is crucial. A widespread practice is to resort to readability formulae because of their simplicity of operation, since any text or excerpt can very quickly be scanned into a computer and numerically classified through readability formulae (Wray & Janan 2014). However, they are atheoretical and might be too approximate and possibly poor to yield a comprehensive text profile. Source-text choice needs to be careful to reduce the impact of source-text features and let task-related cognitive processes emerge. The goal of the present study is to find an alternative to readability formulae to select and relevantly describe source texts for cognitive translation and writing studies.

The chosen text analysis parameters will not define text complexity in absolute terms; the attempt is to make sure that the texts of a certain study, and even between studies, are comparable. The following criteria were chosen to identify the suitable indexes: (1) not overlapping; (2) adhering to a universalist approach; (3) easy to compute; (4) covering different language levels. The indexes were divided into three main categories: lexical, syntactic and textual.

Lexical indexes finally resulted in *vocabulary frequency, nouns-modifiers ratio* and *verbs-modifiers ratio*. Syntactic indexes include *average sentence length, average nouns per sentence, average verbs per sentence, dependency distance* and *subordination*. Finally, textual indexes correspond to *type-token ratio* and *lexical density*.

The indexes' rationales are the following:

- *vocabulary frequency* rationale: correlation between the informants' cognitive effort and the frequency of the lexical words of the text;
- *nouns-modifiers ratio* rationale: the more the modifiers (usually adjectives), the higher the cognitive effort needed to process a nouns;
- *verbs-modifiers ratio* rationale: same as nouns-modifiers ratio, but with verbs and adverbs;
- *average sentence length* rationale: longer sentences are generally related to higher cognitive effort;
- average nouns per sentence rationale: it corresponds to the number of entities in a text that the reader has to keep in memory to understand the text;
- average verbs per sentence rationale: it defines the properties of and relations between concepts/entities;
- dependency distance rationale: the focus of attention is capacity-limited to about four chunks in normal adult humans. Longer dependency distance is associated with higher cognitive effort;
- subordination index rationale: subordinate clauses tend to add a further degree of complexity to texts.

The resulting texts are available in a shared sheet[1].

Keywords: cognitive translation studies, writing studies, cognitive linguistics, text comparability, text profiling

Reference

Wray, David and Dahlia Janan. 2014. "Reassessing the Accuracy and Use of Readability Formulae." In *Malaysian Journal of Learning and Instruction.* 127–145.

[1] The texts can be found here: https://docs.google.com/spreadsheets/d/1EUl-zS6b6lrjTkCxAjVBPq4D5mcQ37gYipR2aobfvco/edit#gid=1349428814