Broadening the spectrum of syntactic complexity measures in relation to the typological specificities of the target language: A corpus-based study of word order diversity in L2 German

Mathieu Lecouvet

University of Louvain, Belgium

mathieu.lecouvet@uclouvain.be

Linguistic complexity has become an established construct in SLA research, mainly due to the essential role of complexity measures in the assessment of learners' written performance. However, research on syntactic complexity was criticized for promoting a reductionist approach, with the consequence that an entire dimension of the construct was left unexplored. In fact, while several definitions of syntactic complexity emphasize concepts such as variety and diversity, most research on L2 writing has confined its calculation to a handful of indices, mainly measures of syntactic elaboration (length-based metrics and frequency of clause-linking devices). There are several reasons for extending the operationalization of syntactic complexity to diversity measures. First, proficient language use does not necessarily imply an increase in the use of weighty structures (longer sentences with more subordinate clauses). Second, the complexification of learners' interlanguage is largely dependent on the typological specificities of the target-language system since languages may rely on different grammar features to encode similar propositional content.

In this study, we make ours this objective to broaden the spectrum of complexity measures and propose to examine the potential of syntactic diversity indicators to distinguish between proficiency groups in advanced L2 German. Concretely, we argue for the use of two types of non-canonical structures, argument inversions and passives, as diversity indices and apply them alongside traditional measures of elaboration on 75 texts written by upper-intermediate to nearnative L2 learners of German. Using data from the FALKO corpus (Humboldt University Berlin), we compiled three sub-corpora of 25 texts, following the corpus guidelines regarding the mapping between learners' score on a C-test and the three superior CEFR proficiency levels (B2, C1, C2). All texts were segmented into clauses and all clauses coded for (a) instances of passives and argument inversions and (b) five indices of syntactic elaboration targeting various syntactic levels (sentence, T-unit and clause): T-units/sentence, T-unit length, clauses/T-unit, NP length, and length of clause midfield. Regarding the identification of non-canonical structures, it should be noted that our coding scheme included a further distinction between the topicalization of non-subject arguments in main clauses and instances of scrambling involving argument reordering in the midfield of main and embedded clauses.

To examine the relationship between complexity measures and proficiency, a KruskalWallis test with pairwise comparisons was performed. Results only revealed statistically significant differences between groups for argument inversions. Specifically, data pointed to a significant increase in the frequency of such structures as from C1 level. The fact that no significant differences between groups could be found for all measures of syntactic elaboration confirms the necessity to supplement large-grained measures (length and clause linking) with more fine-grained indices of diversity which help locate areas of interlanguage complexification that may otherwise stay under the radar.