Learner language as a diversified, genre-dependent concept: from corpus analysis to data-driven learning.

Catherine Thewissen and Liesbet Heyvaert (KU Leuven – Campus Kortrijk, Belgium; FWO-Flanders)

The field of learner corpus research hasn’t stopped growing since the 1990s and has come to include studies on all major areas of language structure whether it be syntax (e.g. Granger 1999 on the use of tenses by learners), lexis (e.g. De Cock 2005 on phrasal verbs and learners), phraseology (e.g. Altenberg and Granger 2001 on the patterning of ‘make’, Granger 1998 on prefabricated patterns in advanced EFL writing) and discourse (e.g. Gilquin 2008 on hesitation markers among EFL learners). The majority of these studies have been based on corpora of written learner English consisting mainly of general essays or academic text types (as Nesselhauf (2004:132) points out, corpora “are often restricted to one medium, to one or only a few text types, and to one level of proficiency”). The contrastive type of analysis that underlies them typically involves quantitative or qualitative comparisons with native corpora or with non-native data produced by learners with a different L1 background (e.g. Granger and Tyson 1996). A desirable development in the field of learner language would be, according to Nesselhauf (2004:132), to create corpora “containing data from learners of different proficiency levels and of corpora of different media and text types [italics CT and LH]”. Numerous studies comment on the fact that thus far most studies have focused on one register, written English, and one text type, the academic essay. While studies strongly advocate greater variety in text type in learner corpora (e.g. Granger 2002), development in that field has been slow.

In our paper, we seek to redress this imbalance by considering learner language not as a monolithic but as a diversified entity that performs differently depending on text type. The contrastive type of Interlanguage Analysis that we present therefore involves the qualitative and quantitative comparison of a group of learners’ production of English in two different text types: we compiled a corpus of learner language consisting of academic (i.e. essay) writing and one that was made up of narratives, both of which are produced by the same group of advanced learners of English with Dutch as their mother tongue. To map out the differences between both corpora, we kept our focus of interest broad and considered lexicogrammatical phenomena ranging from, for instance, the expression of the possessive case in English and the choice between the three variants available in English (the of-phrase, the possessive ‘s and compounding), to collocational patterning and cohesion phenomena. In our paper, we report on these differences, as well as on the similarities that we found between the two corpora. On the basis of the results that we obtained, we developed data-driven activities to be integrated in our own language teaching. A number of these activities will be presented as well.


