

The use of learner corpora in applied linguistics research is growing exponentially (LeBruyn & Paquot, accepted) due to the provision of representative data and to the potential for automatic analyses to address many research questions (Granger, Gilquin & Meunier, 2015). Learner corpora are created for specific and general research objectives (Gilquin, 2015) and, as such, understanding methodological decisions in their compilation is crucial to ensure the corpus is exploited accurately and research findings are correctly interpreted. For example, a strict time limit during the collection of written data affects how the resulting corpora should be interpreted in terms of spontaneous versus planned language use. However, little research has been conducted to understand the impact of different methodological decisions on findings generated automatically from learner corpora, and the focus is largely on oral data (Zechner, 2009).

To this end, the present study examined the impact of two different transcription protocols in terms of L1 versus L2 use on a written corpus containing 32,568 words from narrative and argumentative texts from 235 pupils. The objective protocol required all words that exist in the L2 to be transcribed as L2 words. The subjective protocol required the highlighting of words that were open to interpretation (e.g., "After the police SON parti@s:fra\$v"), which were then discussed by the three authors before being classified as either L1 French or L2 English.

The resulting corpora were automatically analysed using a number of popular analyses for language development (e.g., type/token ratio, VOC-D). The results demonstrated our transcription decisions had a later impact on results. These findings shed light on the importance of transparency when building a learner corpus, the need to reconsider the impact of transcription decisions on written corpora, and eventually, the importance of understanding whether objectivity is always the sine qua non in applied linguistics research.