Investigating L2 errors in a quasi-longitudinal learner English corpus, with particular reference to word order and the position of *also*

Amanda Murphy and Francesca Poli (Università Cattolica del Sacro Cuore, Italy)

The motivation for this study comes from university language teachers' observations of Italian EFL students' 'non-target use' of certain elements or structures (i.e. errors) in written English texts. On the basis of the experience of correcting scripts by EFL students, teachers identified certain errors as persistent, particularly, incorrect choice of prepositions, usage of the present perfect, past simple tense, conditionals, and word order. The intuition of the presence of the Second Language Acquisition concept of fossilisation, understood by Han (2004: 13) both as a process (the premature cessation of development in defiance of optimal learning conditions) and as a product (structures that persist over time, against any environmental influences) began to trigger some research questions. The teachers began to wonder when these errors arise, and devised a project on learner language as it develops through time.

In an era when the debate around English as a lingua franca continues to grow (Di Scala, 2016), and new varieties of English are regularly coded and described by international scholars (starting from the ICE project – Greenbaum, 1991), it may seem curious to insist on correct or incorrect usage of English. However, as Granger robustly argues (2015: 495), it is early days to specify new norms in English, and in a school/university learning context it is not unreasonable to investigate and point out incorrect usage, even more so in a faculty of foreign languages where teaching staff and students alike would like to strive for excellence.

The first broad research question asks whether these five errors - incorrect choice of prepositions, usage of the present perfect, past simple tense, conditionals, and word order - occur continuously throughout written learner production from school to university, or whether they diminish in frequency. The second, stemming from the first, investigates the specific case of the additive adverb *also*, seen in Italian native speaker written learner production to be often misplaced in the clause.

To trace the history of such errors, it was decided to compile what Granger (2008: 262) calls a pseudo or quasi-longitudinal corpus, which collects data from different learners with different levels. A learner corpus of written EFL texts was thus collected between February 2013 and October 2014 and divided into four sub-corpora, corresponding to texts from middle school, secondary school, undergraduate and postgraduate university students. Texts in the Undergraduate Corpus were written in the first of the three BA years, and the texts in the Postgraduate Corpus were written in the first year of a two-year MA degree.

No judgement was made as to the students' level of proficiency; the texts were grouped by external criteria. The corpus was compiled 'as it was possible', i.e. local teachers from the different school levels were contacted, and asked for examples of student texts to include in the corpus. The only requirement given, following Granger (2002: 8), was that "the language sample must consist of continuous stretches of discourse, not isolated sentences or words". No short exercises or translations were accepted. As most of the texts were handwritten, they were digitalised and coded according to school level (Middle School, Secondary School, etc.). The Middle School sub-corpus comprises 35 texts for a total of 31,480 words, the Secondary School sub-corpus 258 texts for a total of 53,678 words, the University sub-corpora consist of 47 texts for a total of 13,939 words; subdivided into 33 Undergraduate texts totalling 3,675 words and 14 Postgraduate texts, totalling 10,264 words. The resulting corpus is clearly not balanced, but nevertheless provides data that can be used for broad comparative research.

To cross-check whether the five errors to be investigated were indeed relevant, a sample of the corpus (10% of the 340 texts in the corpus, which amounts to 100,000) was

submitted to four other colleagues, two non-native speaker teachers of English in school and two native speaker teachers of English at university) with the request to categorise errors in the corpus.

The raters identified two key patterns, one positive and one negative: the former demonstrates that peak incorrect usage of the two past tenses and of the conditionals occurs at high school and middle school level respectively, followed by gradual improvement in the more advanced stages of learning. The negative pattern identified that incorrect usage of the prepositions and word order errors peaked at high school and postgraduate level. Thus, these findings offered the grounds for a more in-depth analysis of the corpus in order to verify whether there is indeed improvement in the use of tenses and conditionals, and a decline in the use of prepositions and word order.
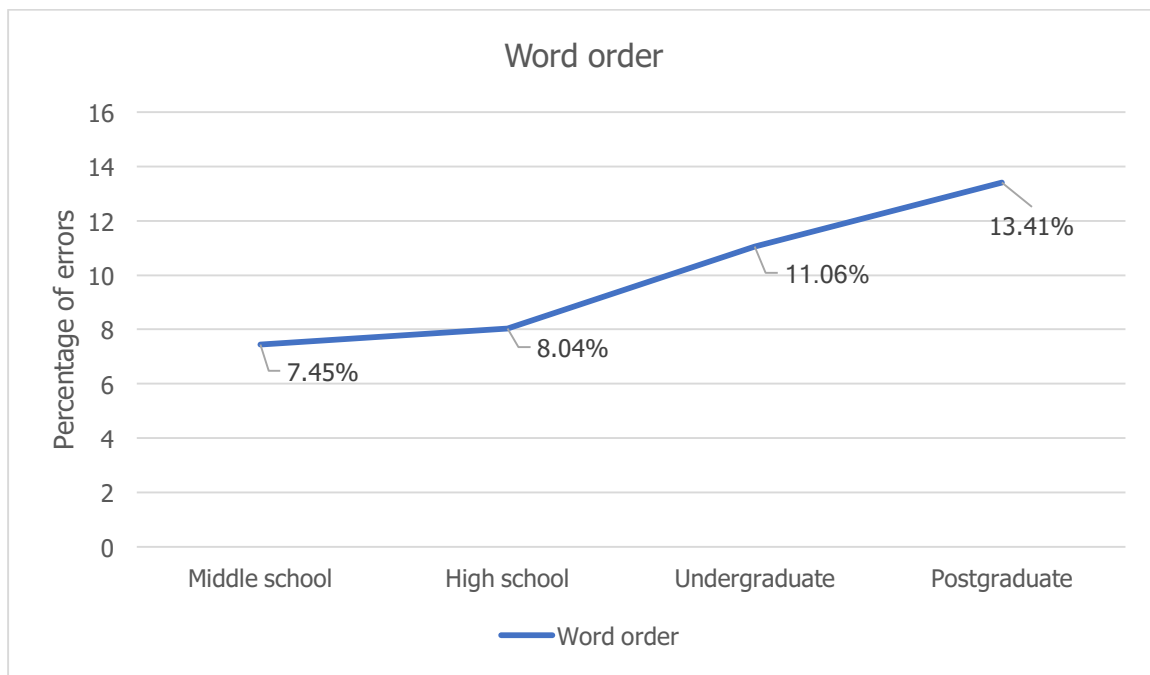
Each text in the corpus was then manually read for the five errors under investigation, firstly by the principal researcher, then by one of two raters, both non-native teachers of English, who categorised errors. A score was assigned to each text for the total number and type of errors, and these scores were subsequently compared across the sub-corpora. Where their scores differed, an average was made of the errors they identified.

The error proportion was calculated with two variables acting as the main operands in the formula: the number of errors (E) and the total number of sentences (S):

$$E/S \times 100$$

Thus, the number of errors (e.g. 1) was divided by the total number of sentences (e.g. 2), obtaining the values of 0.5, which was subsequently multiplied by 100 resulting in the final error proportion of 50%. This simple formula allows for easy and immediate comprehension of the data. The error proportion was calculated for the texts and the percentages were added up and divided by the total number of texts of that specific sub-corpus, thus obtaining an average error proportion for each educational level. The single values for each level for each of the structures analysed in the research were then used to carry out a comparison (in the form of a graph) between the texts.

A comparison of the results revealed that except for word order, the proportionate frequency of the other four categories of errors diminished as the level of education rose, i.e. middle school students incorrectly used verbs in the past simple, present perfect, and the conditional, and prepositions more than postgraduate students. The only category which did not improve was that of word order, as Graph 1 illustrates.

Word order

Percentage of errors

Middle school — 7.45%
High school — 8.04%
Undergraduate — 11.06%
Postgraduate — 13.41%

Word order

Overall, in the Middle School Corpus, 18/35 texts (51.42%) contained word order mistakes; in the Secondary School Corpus 112/231 texts (48.18%); in the Undergraduate Corpus, 18 out of 33 texts (54.54%), and all 12 Postgraduate texts (100%).

The hypothesised reason for this finding was that the Postgraduate texts under analysis were more complex, with long sentences. This hypothesis finds some evidence in the texts, since the Postgraduate texts were short academic essays, requiring some argumentation. Perhaps Selinker's 1972 hypothesis that as students concentrate on content, their attention to form diminishes is also correct, although this cannot be proved.

The second investigation into the learner language in the corpus regarded the use of the additive adverb *also*, which was noted in the reading of the texts as often being misplaced in the clause, especially in the Postgraduate Corpus.

This may not be surprising, since reference grammars such as the *Longman Grammar of Spoken and Written English* warns readers that the meaning of additive adverbs can change, according to its position in the clause (Biber et al. 1999: 781)

Investigations of the adverb *also* using the concordance software AntConc (Anthony, 2011) produced examples that showed the common tendency for *also* to be placed directly after the verb in the Postgraduate Corpus, as in:

(1)   'the Fall of the Berlin Wall had *also* (WO) negative consequences';
(2)   'A widespread system of communication improves *also* (WO) our possibilities to know the world'

Quasi-longitudinal learner corpora can provide a bird's-eye view of the overall broad progression of learner language, although detailed comparisons between levels need to be made with caution. Obviously, the larger the corpus, the more representative the results would be, and the unbalanced nature of the present corpus is a limitation of the present research. Nevertheless, the investigation opens up interesting questions as to possible areas of fossilisation among learners, particularly at a higher level; according to Selinker's early study (1972: 215), fossilised structures tend to remain as potential performance, reappearing in learners' production even when thought to have been eradicated. Findings such as those from the present study could direct teachers of English to areas of language use, such as word

order, and the placing of particular adverbs, such as *also*, whose difficulty is perhaps under-recognised and deserves attention. A hypothesis for the frequent incorrect use of *also* could be negative transfer (Selinker and Gass, 2008) from the students' L1, in this case Italian, but that would require research into contrastive analysis hypotheses and learner corpora, and another paper.

References

Anthony, L. (2011). *AntConc* (Version 3.2.2) [Computer Software]. Tokyo, Japan: Waseda University.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999). *Longman Grammar of Spoken and Written English.* Harlow: Longman.

Di Scala, R. (2016). At the Crossroads of the Global and the Local: Suggested Future Directions for 21st-century ELT. Anglistiki. *An International Journal of English Studies*, 27(1), 175-184.

Garside, R. (1987). The CLAWS Word-Tagging System. In Garside, R., Leech, G. and Sampson, G. (Eds.), *The Computational Analysis of English: A Corpus-based Approach*. London: Longman.

Granger, S., Hung, J. and Petch-Tyson, S. (Eds.) (2002). *Computer Learner Corpora, Second Language Acquisition, and Foreign Language Teaching*. Amsterdam: John Benjamins Publishing.

Granger, S. (2008). Learner Corpora. In Lüdeling, A., *Corpus Linguistics. An International Handbook.* Berlin and New York: Walter de Gruyter.

Granger, S. (2015). The Contribution of Learner Corpora to Second Language Acquisition and Foreign Language Teaching. In Granger, S., Gilquin, G. and Meunier, F. (Eds.), *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press.

Greenbaum, S. (1991). ICE The International Corpus of English. *English Today*, 7(04), 3-7.

Han, Z. (2004). *Fossilization in Adult Second Language Acquisition*. Clevedon, UK: Multilingual Matters.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10(3), 209-232.

Selinker, L. and Gass, S. M. (2008). *Second Language Acquisition: An Introductory Course*. London: Routledge.