

What's in a comma: Corpus study of punctuation errors and L1 interference.

The observations for the presented research were carried out in the learner corpus consisting of essays written in English in the second-year English examination by university students with Russian as their first language. The essays have been collected in REALEC - learner corpus consisting of about 3.5 million words. The research was carried out over a data set of about 900 sentences total with consistent errors in the use of punctuation. The errors are categorised into the following 11 classes:

- confusion of punctuation in different types of relative clauses (cf. *groups prefer going to the cinema, 70% of inhabitants, 65% of migrants from England and 50% of those, who came from other countries*);
- a redundant comma in front of the conjunctions introducing a subordinate clause (such as *It's a big pleasure to see the face of your friend, when you are far away from each other*);
- a redundant comma separating participial construction in the function of attributive or adverbial modifier in the position after the head noun (as in *On the other hand, there are arguments, supporting the opposite view*);
- a redundant comma in front of *than* in comparative constructions (cases like *However there is always a particular woman, which will be more capable of mathematic, than most men.*);
- a redundant comma after *while* in the initial position in a sentence (though not a frequent mistake, such occurrences nevertheless take place in the written production of Russian learners- compare: *While, tracking the behavior of pronouns, a researcher will receive examples ...*);
- a redundant colon after the subject clause in front of the form of the verb BE (cf. *What is worth doing, is worth doing well, it is unwise to forbid it*);
- absence of a comma in front of *too* at the end of the clause (cf. *But all the gold medals have two sides, so there is a bad side of this too.*);
- absence of Oxford comma (see, for example: *... elegant and graceful Meg, shy and timid Beth, who adores music and her doll collection and, finally, Amy, a little artist and «the most important person»* - compare it with the comments about Russian punctuation in <https://blogs.transparent.com/russian/comma-abuse-in-russian/>);
- a redundant dash, comma or colon between the verb BE and the complement (*In conclusion, it is not so easy, to deal with two factors: good salary or enjoyable job.*);
- use of a colon instead of a dash – as in *In conclusion, it is not so easy, to deal with two factors: good salary or enjoyable job.*);
- use of a comma instead of a semi-colon (run-in sentence) - the example is the same sentence as above: *What is worth doing, is worth doing well, it is unwise to forbid it.*

It can be seen from the examples in the classes listed above that students with lower sensitivity towards punctuation conventions may commit more than one mistake in a sentence – like: ...*there are a lot of opponents of this theory, who believe, that at an early age child ...* Russian learners of English are rarely taught punctuation conventions systematically, and this is consistent with the attitude common for learners of English with other native languages (Hirvela, Nussbaum & Pierson 2012).

In each class in our dataset possible interference with the native language (Russian) is analysed as a possible cause of mistaken uses of punctuation, just as it was done in the previous research (Markov, I., Nastase, & Strapparava 2018). The additional research in the Russian National Corpus provides the necessary confirmation with the plethora of its data. On top of the thorough look at the standards (Huddleston & Pullum 2002), we analyse the information structure of the erroneous sentences collected in line with the previous research in (Moore 2016).

References

- Hirvela, A., Nussbaum, A. & Pierson, H. (2012). ESL students' attitudes toward punctuation. *In System*, 40(1).
- Huddleston, R. & Pullum, G.K. (2002) *The Cambridge Grammar of the English Language*. Cambridge University Press.
- Markov, Nastase, V. & Strapparava, C. (2018). Punctuation as Native Language Interference In *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA, pp. 3456–3466.
- Moore, N. (2016). What's the point? the role of punctuation in realising information structure in written English. *In Functional Linguistics*, 3(1).
- Russian National Corpus accessed at <http://www.ruscorpora.ru/>