The Effects of Corpus Use on Error Correction and Error Identification in L2 Writing

As there have been too few studies conducted on the effects of corpus use on error correction and identification in second language (L2) writing, the purpose of this study is to examine the effects of data-driven learning (DDL) on L2 error correction and error identification to explore how information from corpus linguistics influences learning in L2 writing.

The principle aim is to detect the error types for which DDL provides good error correction and identification in L2 writing in order to maximize the benefits of corpus use for L2 teaching and learning in classrooms.

Another aim is to compare the effects of teacher feedback and peer feedback, as they should have different effects on error correction with corpus use. The author examines the types of errors that learners can identify and correct with corpus use without teacher feedback, and the types of errors for which learners need teacher
feedback to identify and correct with corpus use, so that we can use the information from these results to make teacher instruction effective and contribute to learner autonomy.

The author hypothesizes that reading many examples of the target word(s) in a corpus helps learners correct and detect certain types of L2 errors in writing, as they can induce the correct pattern from the examples. Another hypothesis is that different types of feedback have different effects on error correction and error identification with corpus use. To confirm these hypotheses, the author used timed essay tasks. The participants were first given 25 minutes to write an essay based on a topic given by the author without access to reference materials such as the corpus or dictionaries. The participants were next given teacher or peer feedback on their errors. Based on the given feedback, they were given 15 minutes to perform revision tasks with use or non-use of reference materials. The participants’ essays were then collected. The author made learner corpora of the participants’ essays with error annotations containing information about parts of speech, error types, and feedback types. The author finally performed an error analysis based on the error annotations.

This study has two independent variables: reference materials and feedback. Concerning the reference materials, the variable has three levels: corpus, dictionaries,
and non-use of reference materials. The author compared the participants’ correction of errors with corpus use, dictionaries, and non-use of reference materials. The feedback variable also has three levels: teacher feedback, peer feedback, and no feedback. The participants’ corrections of errors were all based on one of the forms of feedback.

A summary of the main findings is as follows. First, corpus use contributed to accurate error correction in L2 writing. The advantages of corpus use were easy access to the exact target phrases and the frequency information of co-occurring words, which especially helped the participants to correct omission errors and word order errors.

Second, error correction with corpus use based on teacher feedback promoted more accurate correction than that based on peer feedback or no feedback, probably because the author tried to detect errors that are correctable with corpus use. A combination of teacher feedback and corpus use was particularly helpful in correcting omission errors and agreement errors. Since different feedback had different effects on correcting errors, appropriate feedback is needed for efficient error correction.

Third, corpus use promoted greater error identification than dictionary use, while corpus use promoted less error identification than no use of reference materials. Corpus use was particularly helpful in detecting collocational errors such as omission and addition errors and form-related errors such as number and agreement errors because
learners could make an inductive inference based on example sentences of the target phrases. As with error correction, the advantage of corpus use for error identification was easy access to the target phrases and information on frequency.

Judging from these findings, to make the most of the effects of corpus use, teachers must give enough instruction about the error types learners need to focus on and the number of errors they need to detect and correct during a limited time. Teachers especially need to offer learners instruction on appropriate error types for each feedback when learners correct their errors based on peer feedback or self-correct them based on no feedback, as peer feedback and no feedback contributed to less accurate correction than teacher feedback. To use a corpus to contribute to accurate error correction in L2 writing in classrooms, teachers need to consider various factors such as feedback, error types, learners’ language proficiency, and task types so that DDL can promote accuracy in L2 learning and serve as a practical option in L2 classes.

Among the advantages of this research is that its findings suggest how corpus use can promote practice in the SLA and English Language Teaching (ELT) areas. The findings should contribute to the spread of DDL in L2 classrooms and the application of corpus linguistics to SLA and ELT.

Limitations of this research include some unclear points: the long-term effects of
corpus use with new texts, the effects of peer feedback on accurate correction with corpus use, and the influence of learners’ proficiency on the effects of corrective feedback remain unclear. Further research is needed in these fields.

It seems reasonable to conclude that this research has provided practical suggestions for introducing DDL into L2 classes, and on the theoretical side has contributed to promoting the application of corpus linguistics in the SLA and ELT fields.