

A review of the usefulness of data-driven second language learning

What this research was about and why it is important

Teachers often make decisions about *what* language to present to learners and *how* to do so. For example, should language be simplified or 'authentic'? Should it be formally presented or should learners discover it for themselves? Data-driven learning (DDL) involves the use of electronic corpora (databases) of spoken or written language as a resource for second language (L2) learning. This review systematically combined the results of 64 studies on DDL to investigate its usefulness in L2 learning. The results suggested that DDL has positive effects in a range of contexts.

What the researchers did

- After a systematic search for studies on DDL, the authors identified 64 that were suitable for inclusion in their review. These studies had two different types of design:
 - (1) The results of a single participant group were compared before and after DDL. These studies investigated how much DDL increases knowledge and skills (i.e., effectiveness).
 - (2) An experimental group, who used a language corpus, was compared with a control group who covered the material in a different way. These looked at how well DDL works compared to other types of language learning (i.e., efficiency).
- Effect sizes (a way to measure the size of difference between two groups) were calculated for each study. They were also averaged across studies, and separately for the two types of study design above.
- The results were also broken down for further analyses that might explain variation in the overall results. These included analyses for different types of language aspect (e.g., vocabulary vs. grammar), participant proficiency level (e.g., intermediate vs. advanced) and type of corpus use (e.g., actual corpus use vs. printed materials).

What the researchers found

- **Overall, DDL was found to have a large effect for both effectiveness and efficiency**
- The largest effects were found when learners themselves interacted with a language corpus, rather than using print-outs and paper-based materials.
- DDL worked equally well in different environments (e.g., classroom vs. labs) and over different time periods
- DDL seems to be broadly applicable across a range of pedagogical contexts and learners, whether for
 - undergraduates or graduates;
 - intermediates or advanced learners;
 - general and specific-purpose L2 learning;
 - vocabulary or lexicogrammar

Things to consider

- In their analyses of the overall results, the authors didn't find any conditions or learner profiles that predicted where DDL did *not* work. This suggests that DDL is *potentially* useful for a very wide range of learners and teaching contexts.
- Nearly all participants in the studies included were learners at universities. It remains less well understood how effective DDL is for younger age groups.
- DDL was found to have positive effects on L2 writing, but far fewer studies investigated listening and reading, and none looked at speaking. More research would be needed to know if DDL affects the development of these language skills in the same way and to the same extent.
- The studies included used post-tests (measuring the effects of DDL straight afterwards), but very few studies used a delayed post-test. This would have been useful in assessing the effectiveness of DDL over a longer time.
- The largest effects were found in Asian and Middle Eastern countries. This goes against what might be expected in such teaching contexts, which often favour teacher-centered approaches to learning.
- **In your context:** Do you use DDL? How do/would your learners react to using corpora for language learning?

Materials and data available from <https://osf.io/jkktw/> and <https://www.iris-database.org/>

How to cite this summary: Avery, N., Plonsky, L., & Boulton, A. (2018). Corpus use in language learning: A meta-analysis. OASIS Summary of Boulton & Cobb (2017) in *Language Learning*. <https://oasis-database.org>

This summary has a [CC BY-NC-SA license](https://creativecommons.org/licenses/by-nc-sa/4.0/).