



HAL
open science

Expanding Methodological Approaches in DDL Research

Alex Boulton, Nina Vyatkina

► **To cite this version:**

Alex Boulton, Nina Vyatkina. Expanding Methodological Approaches in DDL Research. TESOL Quarterly, 2023, 10.1002/tesq.3269 . hal-04478651

HAL Id: hal-04478651

<https://hal.science/hal-04478651>

Submitted on 26 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

This is a pre-publication version; where possible, please refer to the definitive text:
Boulton, A., & Vyatkina, N. (2024). Expanding methodological approaches in DDL research. *TESOL Quarterly*, 1–12. <https://doi.org/10.1002/tesq.3269>

Expanding methodological approaches in DDL research

ABSTRACT.

This paper analyses the methodologies in 148 empirical data-driven learning studies for L2 English in prestige journals to examine best practice. Manual coding and corpus analysis of key words and n-grams from the past five years (2018-22) explore the field as a whole and how methodologies have evolved, suggesting improvements and future avenues for research.

INTRODUCTION

Data-Driven Learning (DDL) involves the use of corpus tools and techniques for foreign/second (L2) language learning and use. Hundreds of empirical studies have led to several research syntheses, from qualitative narrative reviews (e.g., Chambers, 2007) to quantitative meta-analyses (e.g., Boulton & Cobb, 2017), bibliometric analyses (e.g., Dong et al., 2023) and other systematic reviews (e.g., Pérez-Paredes, 2022). In the most comprehensive to date, Boulton and Vyatkina (2021) focused on how proposals for future research directions were taken up for exploration in subsequent studies: several methodological practices remained remarkably constant over thirty years despite repeated calls for change from DDL researchers themselves. This study builds on the methods and results of earlier DDL syntheses, with a particular focus on methodology in empirical DDL studies targeting English, since research results depend on the methodologies employed (cf. Chong & Plonsky, 2023). Specifically, manual coding alongside corpus analysis of the methodology sections allows a two-pronged overview of the field to identify methodologies used, especially in the last five years (2018-2022), and to suggest avenues for future research.

METHOD

The *article sample* reflects our primary interest in empirical DDL research, defined as studies that evaluate some aspect of the explicit use of corpus tools and techniques for L2 learning or use. Scientific “quality” is clearly a major methodological issue here: while all researchers know intuitively what quality means, it is difficult to establish specific criteria to measure it empirically: relevance, research questions, rigorous design with appropriate materials and instruments, analysis and interpretation, etc. – all of these entail an element of subjective appreciation. Many studies of quality in applied linguistics, such as Xu et al. (in press), have combined quantifiable aspects (e.g., journal impact factor) with qualitative judgements (researchers’ ratings of various journals), finding correlation between them (see also Choubsaz et al., 2024). Drawing on this, we limited our survey to research articles (RAs) published in journals ranked in the latest Journal Citation Reports (2023) for Linguistics and Education. While journal ranking is no guarantee of the scientific quality of any individual paper, taken as a group, papers published in high-ranking journals are likely to be better received and, by definition, have more impact overall. It additionally allows confidence in a near-exhaustive collection up to and including 2022. Keyword searches (*DDL*, *data-driven*, *corpus/corpora*, *concordanc**) resulted in a total of 148 papers for English L2. The

timeline with RA counts by year is presented in Figure 1: minor fluctuations in output are inevitable with relatively small numbers; the minor decrease between 2019 (13 RAs) and 2022 (7 RAs) can be compared with the fact that, at the time of writing, we are aware of 14 RAs *in press* in their final form, complete with DOI. Table 1 gives details of journals with at least three entries including one in each period, followed by less frequent sources.

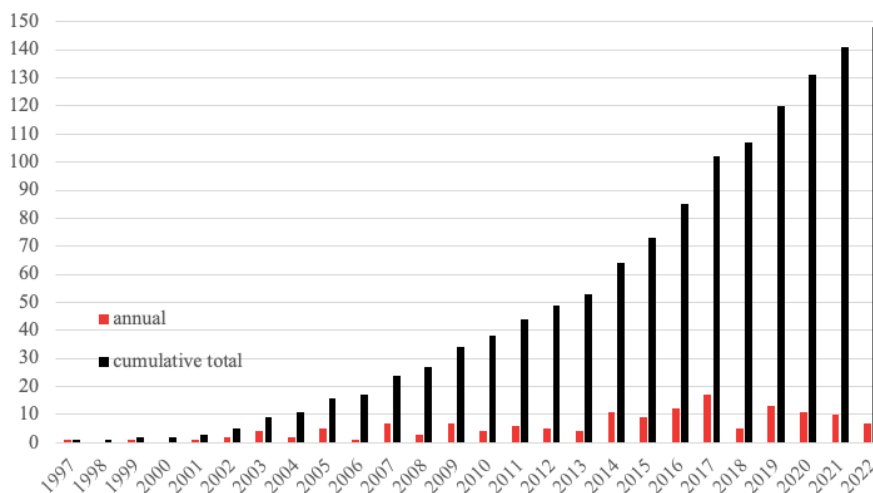


FIGURE 1. Annual and cumulative totals for included RAs

**TABLE 1
Journals, Journal Impact Factor (JIF) and RA counts**

JIF 2022		Journal	1997-2017	2018-2022	TOTAL
LING	EDU				
4.5	4.5	ReCALL	20	6	26
7.0	7.0	Computer Assisted Language Learning	22	3	25
3.8	3.8	Language Learning & Technology	17	2	19
6.0	6.0	System	8	6	14
3.0	3.0	Journal of English for Academic Purposes	3	6	9
2.5		English for Specific Purposes	4	3	7
2.5	2.5	English Language Teaching Journal	5	1	6
0.5		International Journal of Lexicography	4	1	5
6.1		Journal of Second Language Writing	2	2	4
1.6	1.6	International Journal of Applied Linguistics	1	2	3
OTHERS: Australasian Journal of Educational Technology, British Journal of Educational Technology, Education and Information Technologies, Educational Technology & Society, English in Education, Ibérica, IEEE Transactions on Learning Technologies, International Journal of Corpus Linguistics, International Review of Applied Linguistics in Language Teaching, Interpreter and Translator Trainer, Journal of Computer Assisted Learning, Journal of Computing in Higher Education, Language Awareness, Language Learning, Language Teaching Research, Lingua, Modern Language Journal, Perspectives: Studies in Translation Theory and Practice, RELC Journal, TESOL Quarterly					
TOTAL			102	46	148

The manual stage of *data analysis* involved reading new papers to update the coding sheet¹ in Boulton and Vyatkina (2021). In the semi-automated stage, the 148 RAs included were converted to txt format and ‘cleaned’ to retain the authors’ own text; a subcorpus of only the Methodology

¹ Available in the IRIS repository; the full text corpus cannot be made public due to copyright restrictions.

sections totals 252,326 words. The corpus was lemmatized via TagAnt, and analysis conducted using AntConc, with a stoplist to exclude the 200 most frequent items (e.g., *the, of*) from the BNC. Keyword and key n-gram lists were used for analysis of recurring methodological themes. This dual approach allowed us to triangulate the results and pinpoint specificities which one alone might miss.

CODING RESULTS

This section outlines the main methodological characteristics of DDL research to date and, where substantial differences are noted, specific to the last five years (2018-2022). First, an overview of the background will provide minimum context. Overall, Asia accounted for 44% of all studies and is on the rise; Europe remained stable at around 27% of studies, with contributions from 14 countries; the Middle East rose from 7% to 13% (total 9%); North America represented 17% of all studies, mainly from the USA (16), but dropped from 22% to 7%. In total, ‘inner-circle’ English-speaking countries represented 24% of RAs (down from 28% to 15%); these accounted for most second-language contexts, though the majority (72%) were in countries where English is not spoken as an everyday or official language. The participants are predominantly at upper-intermediate or advanced levels of English proficiency: 57% compared to 20% lower-intermediate or below (there were no ‘beginners’); the others are labeled ‘intermediate’ with no further qualification. 90% of studies were conducted at university, with just 11 in earlier education and 2 in other contexts. Where specified, nearly half of all courses were language related (45%); most were for English for general purposes (EGP, 60%), though in the final period, academic purposes (EAP) rose from 25% to 45%, while specific purposes (ESP) halved to just 5%. Vocabulary and lexicogrammar have dominated the linguistic goals with around a third of studies each, the others consisting of grammar, discourse and error-correction. Writing was already the main language skill in early papers but has risen further, from 58% to 69%, to the detriment in recent years of spoken (17%) or receptive skills (10%), and translation (down from 18% to just 3%).

The duration of intervention has been reported variously, making accurate comparison all but impossible. However, for most studies (100) reporting in hours/minutes, the mean duration dropped by a third, from 16 hours 48 minutes to 11 hours 09 minutes. A wide variety of corpora has been used hands-on, with a preference for large national corpora with built-in tools: COCA rose from 12% to 48% while the BNC dropped from 29% to 15%; between them they feature in 46% of all studies, followed by custom-designed corpora at 38%. Variety is decreasing, with no instances in 2018-22 of the web-as-corpus, graded texts, literary corpora, multimodal or parallel corpora. Overall, corpus size is increasing, from a median of 26 million to 100 million tokens. Concordancers remain the tool of choice for learner-corpus interaction in 65% of the studies, compared to prepared materials or corpora integrated to a wider CALL package.

Sample sizes vary considerably; the median has increased from 29 to 55 participants, although only 31% included a non-DDL control group. The main focus has been on the exploration of language gains (70% of the studies), using corpora either as a learning aid (39%) or a reference resource (32%). Gains are generally measured via tests (43%) or learner productions (27%), though questionnaires are the most popular data-collection instrument overall, rising from 57% to 67% of the studies. This is mainly for learner attitudes (54%); other emic tools include interviews (20%) and diaries (18%), though these are sometimes for assessing what the participants do with

the tools (28%) – computerized logs to track behavior are relatively rare (16%). 64% used more than one instrument; most employed quantitative designs with only 10% being purely qualitative, 32% using raw numbers, percentages or at most descriptive statistics, and 58% at least some inferential statistics; these figures remain largely unchanged.

CORPUS RESULTS

To triangulate these results, AntConc was used for a semi-automated analysis of our lemmatized Methodology subcorpus. Table 2 shows keywords (items that are statistically significantly more frequent in the recent vs early RAs); all appear in at least 20% of papers (10/46), ranked by log likelihood. All items were analyzed in context to identify evolutions in terminology and themes related to methodology compared to the earlier publications.

One group becoming more frequent in the recent period points to a greater focus on research: *intervention*, *instruction* and *participants*. Other instances include *test*, which generally co-occurs with *pre* and *post* or *t-*; *min* and *score*, *condition* (group comparisons), *indicate* (reporting results), *change* (following instruction), *author* (citations), *question* (either research questions, or in questionnaires or interviews) and *response*, and *open* (questions). Another group reflects specific technology or aspects of corpora: *COCA* and *AntConc*; *sub* (corpus), *file* (in corpus creation) and (concordance) *line*. *Target* (word, collocation, etc.) shows different framing; *meaning* also highlights a move away from form alone, with which it is often contrasted. Other keywords suggest lexical preferences (e.g., *learner* vs *student*), a range of meanings (e.g., *interactions* between students, or with the software or corpus, or between different data sets, etc.), or quite general concepts (e.g., *follow*), and are harder to interpret.

TABLE 2
Keywords in 2018-2022

Rank	Type	<i>f</i> Tar	<i>f</i> Ref	<i>R</i> Tar	<i>R</i> Ref	Keyness (LL)	Rank	Type	<i>f</i> Tar	<i>f</i> Ref	<i>R</i> Tar	<i>R</i> Ref	Keyness (LL)
1	post	154	122	30	31	68.28	15	response	109	108	21	43	32.15
2	intervention	58	18	11	11	63.92	16	indicate	78	66	21	40	31.05
3	min	62	25	15	11	57.30	17	pre	133	146	27	39	30.94
4	coca	69	34	20	14	53.96	18	score	140	164	26	49	27.38
5	teacher	274	328	35	63	50.39	19	author	76	72	21	34	24.58
6	learner	404	549	38	72	48.45	20	meaning	98	110	23	45	21.52
7	file	47	19	16	12	43.35	21	question	212	301	39	80	21.20
8	participant	467	675	41	80	43.24	22	follow	198	277	44	87	21.08
9	t	55	27	12	15	43.16	23	change	49	40	15	26	20.70
10	test	452	652	34	66	42.23	24	open	52	44	24	27	20.70
11	condition	75	54	12	24	38.24	25	sub	42	31	10	10	20.68
12	theme	25	5	12	5	34.28	26	target	174	238	29	56	20.29
13	interaction	42	21	12	13	32.40	27	instruction	111	134	27	44	19.88
14	line	141	156	28	53	32.16	28	antconc	31	19	10	10	19.37

Note: *f*=frequency; *R*=range; *Tar*=target corpus (2018-22); *Ref*=reference corpus (1997-2017)

These results are largely corroborated by key n-grams – clusters that are significantly more frequent in the recent studies. The cut-off point was set at 10% of RAs (5 of the 46 in 2018-22), resulting in 100 key 2-grams, 77 3-grams, 18 4-grams and 5 5-grams; the top 5 of each are presented in Table 3. As with keywords, some are not particularly revealing (*the teacher*, *the*

participants were), but most of the others relate to corpus tools (*Corpus of Contemporary American English*) and features (*the concordance lines*), and in particular to research design: *the comparison (group), by the researchers, of the target (word, etc.), before and after the (instruction, etc.), in the pre- and / post- (tests)*. Continuing down the lists supports these general findings.

TABLE 3
Key n-grams in 2018-2022

	Rank	Frequency	Range	Keyness	Effect	Key n-gram
2-grams	1	110	24	88.97	0.0089	concordance lines
	2	83	18	73.14	0.0067	post test
	3	116	23	67.07	0.0093	the teacher
	4	22	7	60.36	0.0018	the comparison
	5	201	33	54.17	0.0160	the participants
3-grams	1	12	5	36.63	0.0058	by the researchers
	2	25	13	33.90	0.0120	the concordance lines
	3	29	14	32.11	0.0139	of the target
	4	53	21	30.74	0.0250	the participants were
	5	10	6	30.52	0.0048	english language teaching
4-grams	1	20	17	29.23	0.0749	corpus of contemporary american
	2	20	17	29.23	0.0749	of contemporary american english
	3	8	6	27.19	0.0314	before and after the
	4	7	5	23.78	0.0276	in the pre and
	5	7	7	23.78	0.0276	pre test and post
5-grams	1	20	17	22.67	0.2759	corpus of contemporary american english
	2	6	6	17.67	0.1017	pre test and post test
	3	11	6	13.66	0.1705	the pre and post tests
	4	14	12	13.04	0.2044	the corpus of contemporary american
	5	12	12	11.57	0.1805	of contemporary american english coca

Note: Range refers to the latest period

DISCUSSION AND CONCLUSION

No synthesis should be treated as the last word: each has its own strengths and limitations in terms of scope, inclusion, research questions and analysis. Our definition of DDL may be too broad for some, too narrow for others; non-empirical studies are excluded, as are non-ranked journals and other publications, and texts in languages other than English; different coding or analysis might uncover additional methodological items. However, these criteria do provide greater focus on a near-exhaustive collection of generally high-quality papers that meet transparent criteria.

On the positive side, there is a substantial body of empirical DDL research in prestige journals, especially over the last 10 years. Further, quantitative meta-analyses have shown medium or large effect sizes for DDL (e.g., Boulton & Cobb, 2017). Our analysis originally divided the collection into several periods, but showed remarkably little difference in proportions of methodological features at that level of granularity. This is an interesting finding in itself: either the research has always been solid right from the early days or, alternatively, there is a strong degree of inertia – not learning from past difficulties or adopting better methodological practices.

In general terms of background, several comparisons can be drawn with Plonsky's (2023) review of applied linguistics. We too find a dominance of L2 English (see Forti, 2023, and Vyatkina, 2024, for DDL of other languages) and regional hubs, but our East Asia hub is more prominent

and North America less so. Most DDL research has featured relatively advanced-level learners compared to applied linguistics more widely, though numbers should be interpreted as approximate since “proficiency is notoriously difficult to define and measure” (Plonsky, 2023, p. 11), and is reported inconsistently. Nonetheless, the overall higher proficiency level of participants in DDL research may be an indication of perceived challenges and, thus, reluctance to explore it more widely with lower-proficiency learners. University contexts also dominate DDL research, though at a staggering 90% compared to 39% in Plonsky. Clearly DDL has similar methodological strengths and weaknesses as other applied linguistics research. As elsewhere, there is an acute need for more research in other contexts and with other profiles. More specifically, DDL may have already established its default niche, painting itself into a corner with relatively advanced learners in language-related courses at university in certain regions of the world. It may of course be that this is the optimum setting for DDL, or simply that researchers default to the convention of existing studies (Larsson et al., 2023); work is under way to examine this.

The aim is generally EGP, with vocabulary and lexicogrammar for writing including error-correction. One possibility would be to consider domain-specific corpora to help with both ESP and EAP, and with content: even a small corpus of relevant texts (e.g., research articles or textbooks) could help learners explore terminology in their field and how it is put together in academic writing, and to find information about their own area of specialization. Second, traditional corpus tools favor lexical searches: new tools might help at other levels, for example through texts annotated for discourse features. Third, there is no obvious reason why a DDL approach cannot apply to skills such as translation, which is all but invisible in our recent collection.

Such moves would involve rethinking DDL not as *corpus + concordancer* but, as we defined it, as using the tools and techniques of corpus linguistics: harnessing software to explore large quantities of text, discovering patterns in a constructivist approach to language learning. This shift involves creative imagining in using familiar tools, even Google or Linguee or ChatGPT, to bring DDL to the learners rather than asking them to make the leap to corpus linguistics – a kind of “DDL-lite”. For spoken skills, free tools like YouGlish allow users to query a particular word or phrase and then to watch short chunks of videos where the item occurs, hearing it in context and reading it in the subtitles. Despite the qualitative difference (target items can only be accessed sequentially), the principles are essentially the same. These points all highlight the need for greater variety in the texts used. Large corpora dominate much DDL for good reason, but smaller ones potentially have a significant role to play: if learners are shown how to create their own collections for their individual interests, they might look at news articles about their favorite sport, a novel they are reading, scripts from a much-loved TV series, etc. Personal involvement would provide greater relevance and potentially familiarity with the contents. In addition to traditional concordancers, software such as SKELL and Linggle are simpler and more user-friendly, and with English-Corpora.org it is possible to input text for immediate frequency analysis (an indication of difficulty, or items worth learning), visualization of register distribution, topics, collocates by part of speech, clusters and concordance lines – all on the same page.

Returning to more mainstream methodological issues, one surprising find is that the duration of the target instruction has decreased; this however may be due in part to poor reporting, an area in need of improvement in much applied linguistics. At the same time, sample size has increased,

though this is partly due to the use of more than one experimental group – the intact group is still the norm. An incidental problem of convenience sampling is that it encourages pre/post-test designs: only a third of our studies involved a control group, though some did compare two groups doing DDL in different ways. Increased collaborative research would provide larger samples and greater statistical power (cf. Plonsky, 2023), and could provide more options for comparative designs with different learner profiles (e.g., L1s, disciplines, proficiency levels), corpora, software, types of scaffolding, etc. Direct comparison would provide greater confidence than by comparing different studies. A final problem concerns the instruments used: questionnaires are becoming ubiquitous, though they often feel as if they are tacked on at the end in lip-service to some idea of triangulation or mixed methods research. This can be valuable in some cases, but other instruments (e.g., trackers) should be considered for monitoring learners' behavior. Another pressing need is for more research on learners' use of DDL outside class and, indeed, after the end of the course: since DDL has been claimed to foster autonomy and life-long learning, we need to actually research this and not just take it as given. The same goes for other underpinnings of the approach, inferred but rarely explored first-hand as the topic of a specific research question: whether DDL really does foster noticing, language awareness, greater cognitive depth, etc. There is certainly room for DDL to take greater inspiration from SLA theories (O'Keeffe, 2021), and greater consideration of best practice. Lack of research here no doubt reflects the difficulty of implementing such studies – we need greater creativity and imagination in designing studies to test the foundations of DDL.

To finish on a positive note: there is considerable empirical research in DDL for L2 English, much of it of relatively high quality. It is always possible to improve on research designs, but many of the criticisms apply to applied linguistics in general, and it is encouraging that the latest period shows an increase in research-related terms. More fundamentally, it is important for methodologies to go hand-in-hand with research questions right from the start of study design. The question is not 'what can we do?', or 'what is easy to do?', but 'what do we want to find out?' and, following that, 'what is the best way to go about it?'

REFERENCES

- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393. <https://doi.org/10.1111/lang.12224>
- Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions. *Language Learning & Technology*, 25(3), 66–89. <https://www.lltjournal.org/item/10125-73450>
- Chambers, A. (2007). Popularising corpus consultation by language learners and teachers. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 3–16). Rodopi. <https://doi.org/10.1163/9789401203906>
- Chong, S. W., & Plonsky, L. (2023). A typology of secondary research in Applied Linguistics. *Applied Linguistics Review*. Advance Online Publication. <https://doi.org/10.1515/applirev-2022-0189>
- Choubsaz, Y., Jalilifar, A., & Boulton, A. (2024). A longitudinal analysis of highly cited papers in four CALL journals. *ReCALL*. Advance Online Publication. <https://doi.org/10.1017/S0958344023000137>

- Dong, J., Zhao, Y., & Buckingham, L. (2023). The intellectual structure of the data-driven learning field: A bibliometric analysis. *ReCALL*, 35(3), 339-355. <https://doi.org/10.1017/S0958344022000222>
- Forti, L. (2023). *Corpus use in Italian language pedagogy: Exploring the effects of data-driven learning*. Routledge. <https://doi.org/10.4324/9781003137320>
- Larsson, T., Plonsky, L., & Hancock, G.R. (2022). On learner characteristics and why we should model them as latent variables. *International Journal of Learner Corpus Research*, 8(2), 23 - 260. <https://doi.org/10.1075/ijlcr.21007.lar>
- O’Keeffe, A. (2021). Data-driven learning – a call for a broader research gaze. *Language Teaching*, 54, 259-272. <https://doi.org/10.1017/S0261444820000245>
- Pérez-Paredes (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011-2015. *Computer Assisted Language Learning*, 35(1-2), 36-61. <https://doi.org/10.1080/09588221.2019.1667832>
- Plonsky, L. (2023). Sampling and generalizability in Lx research: A second-order synthesis. *Languages*, 8(1), 75. <https://doi.org/10.3390/languages8010075>
- Vyatkina, N. (2024). *Corpus-based applications in language teaching and research: The case of data-driven learning of German*. Routledge.
- Xu, Y., Zhuang, J., Blair, B., Kim, A., Li, F., Thorson Hernández, R., & Plonsky, L. (in press). Modeling quality and prestige in applied linguistics journals: A bibliometric and synthetic analysis. *Studies in Second Language Learning and Teaching*.