



Experimenting with corpus-driven approaches: On paper or on computer

LILIT AVETISYAN^a  

MARINE AVETISYAN^a  

^aEuropean University of Armenia, Armenia

Abstract

Language teachers who have received training in corpus linguistics can resort to Data-Driven Learning (DDL) as a supplement to their conventional teaching in two ways: direct or indirect. This means that learners can use concordances indirectly through corpus-based materials designed by teachers as handouts or they can have direct computer-based experience with corpora (Chambers, 2022). The aim of this study is to provide useful skills and processes in the use of language corpora as a resource for grammar instruction and to experiment with paper-based and computer-based approaches of DDL. The mixed-method design combined pre- and post-tests with semi-structured interviews to evaluate the effects of paper-based (“hands-off”) and computer-based (“hands-on”) DDL over a 10-week instructional period. The study recruited 32 Armenian adult learners studying English at elementary and pre-intermediate levels outside an academic context. The analysis of the evaluation data collected from pre- and post-tests reveals that while the elementary group improved their grammar competency more from the paper-based work, the pre-intermediate learners benefited from both the paper-based and hands-on training. Additionally, the semi-structured interviews sought to enhance insights into the language learning process and brought to light some of the benefits and challenges of using DDL. These findings demonstrate that DDL can be meaningfully adapted for lower-proficiency learners through appropriate scaffolding and may inform future curriculum design and teacher development in comparable EFL contexts.

Keywords: data-driven learning (DDL); direct/indirect DDL; grammar teaching; low-level language learners; inductive learning

Introduction

Much of the research in the area of Data-Driven Learning (DDL) focuses on the operationalization of DDL, and the key topics that arise include the direct and indirect uses of DDL. This means that learners can use concordances indirectly through corpus-based materials designed by teachers as handouts or they can have direct computer-based experience with corpora (Chambers, 2022). The direct and indirect approaches are also termed “hard” and “soft” (Gabrielatos, 2005), or “hands-on” and “hands-off” (Boulton, 2010a) approaches, respectively. While both methods share the goal of promoting inductive, learner-centered discovery of language patterns, they differ substantially in the degree of learner autonomy, the authenticity of input, and the extent of teacher mediation.

For readers who may be unfamiliar with DDL, it is important to note that these approaches lie on a continuum rather than forming a binary opposition. At one end of this continuum are teacher-led, paper-based activities with simplified input, and at the other are learner-driven explorations of authentic corpus data on computers. A detailed model of this continuum, adapted from Gabrielatos (2005), is presented in the Literature Review section (see Figure 1), where its dimensions are discussed more fully in connection with previous research and theoretical frameworks.

Despite a growing body of literature exploring the pedagogical applications of DDL, its use and effectiveness remain underexplored in the Armenian EFL context. English instruction in Armenia is predominantly textbook-oriented and deductive, leaving little room for learner discovery or technology-enhanced approaches (Mkrtychyan, 2021). This creates a valuable opportunity to examine how DDL, particularly at lower proficiency levels and in professional rather than academic settings, can promote engagement, noticing, and autonomous learning. Theoretically, the study is grounded in constructivist and socio-cultural perspectives that emphasize active hypothesis formation and scaffolding. Pedagogically, it seeks to test whether corpus-based input can bridge the gap between explicit instruction and inductive learning practices in under-researched educational contexts. Empirically, existing literature has not sufficiently explored (a) how DDL functions within workplace-oriented learning environments, (b) how elementary and pre-intermediate learners respond to “soft” versus “hard” DDL, or (c) how learners perceive these approaches when used sequentially within the same instructional cycle. Moreover, literature reports operational uncertainties related to the use of DDL with low-level learners. These gaps provide the rationale for the present study.

Accordingly, this mixed-method study was designed to experiment with direct and indirect approaches of DDL to gain primary proof of the concept of the operationalization of DDL, as well as to bring to light some of the fears, challenges, and benefits of using DDL at lower levels of language proficiency. It particularly aimed to answer the following research questions:

- 1) Is there a difference in the learning outcomes between paper-based (hands-off) and computer-based (hands-on) DDL at lower levels of language proficiency?
- 2) Is there a difference in the learning outcomes between different levels of language proficiency?
- 3) What are the learners’ preferences and attitudes towards the use of DDL on paper and on computer?

The findings revealed that elementary learners benefited more from paper-based DDL, while pre-intermediate learners showed improvement through both paper- and computer-based activities. These outcomes suggest that DDL can be effectively adapted to diverse proficiency

levels through appropriate scaffolding. Moreover, learner interviews highlighted both enthusiasm for DDL and practical challenges in engaging with corpora, offering pedagogical insights for integrating DDL into similar EFL contexts.

Literature Review

Within the field of second language pedagogy, the emphasis has shifted from deductive to inductive learning (Ellis, 2008; Lam, 2024; Lee, 2025; Nassaji & Fotos, 2011; Schmidt, 2022). The inductive approach prioritizes the concept of “noticing” (Schmidt, 1990) language patterns (or forms). This type of student-centered self-discovery of lexico-grammatical patterns can be greatly aided by “DDL” (Johns, 1986; Liu & Chen, 2024; O’Keeffe et al., 2007; Pérez-Paredes & Boulton, 2025), which involves learners investigating real language use through corpora, in contrast to Large Language Models (LLM), such as ChatGPT, which provide users with answers to queries (Flowerdew, 2024). Since the teacher’s role is to mediate language learning needs, they have the responsibility to critically evaluate and recontextualize the corpus-based descriptions. One distinction they should know is between paper-based “hands-off” and computer-based “hands-on” methods. According to Johns (1991), hands-off corpus-driven activities can be introduced at lower levels of language proficiency for immediate results. They require minimal or no corpus training, which can be an advantage for learners who are reluctant to work with software or are not well aware of how to work with it or how to interpret the results (Boulton, 2010a). However, the mere fact that learners work with a corpus-based handout does not guarantee successful learning unless the teacher is able to use them judiciously (Frankenberg-Garcia, 2014). The soft type of DDL can be a solution in contexts where computers are not available on a regular basis, and both teachers and learners are overwhelmed by the use of “new material (the corpora), new technology (the software), and new approach (DDL) all at once” (Boulton, 2010a, p. 539). Despite the overstated motivating factor of technology in education, computers can be unappealing for many teachers (O’Keeffe & Farr, 2003), as well as learners, and, therefore, become an obstacle for wider uptake of DDL (Yoon & Hirvela, 2004).

The use of prepared materials allows the teacher to tailor activities to learners’ needs and abilities (Boulton, 2012) and avoid the indiscriminate use of concordances (Frankenberg-Garcia, 2014). This way the teacher can edit the language by removing overly complex or sensitive content, thus sheltering learners from many of the challenges associated with raw corpus data (Frankenberg-Garcia, 2014). Furthermore, printed materials can provide a gentle lead-in to hands-on experience (Gabrielatos, 2005), and “scaffolding can be gradually reduced until students can be presented with concordance output to investigate independently and unaided” (Johns et al., 2008, p. 495).

The difference between direct and indirect DDL is not merely the medium of delivery. Hands-off concordance does not have the full potential of hands-on corpus work. The latter can be achieved through extensive training, though it is often difficult to implement in an already established syllabus. The benefits that learners can extract from hands-on corpus consultation include flexibility, autonomy, lifelong learning, and long-term recall (Boulton, 2012; Flowerdew, 2024). Recent research in technology-enhanced language learning further supports the link between learner autonomy and digital contexts. For example, Teng (2024) showed that learners’ self-efficacy, a key aspect of autonomy, mediates how they engage with online English learning, highlighting the importance of designing tasks that foster independent exploration in digital environments. Direct corpus use can also provide learners with an experience of a linguist. However, a suggestion is given that hands-on corpus activities, like handouts, should be immediately applicable to learners’ language learning interests, needs, and goals (Frankenberg-Garcia, 2014). Through hands-on experience, learners have more opportunity to find

answers to their individual questions, to select data relevant to them, and to see more contexts, which are selectively printed on handouts. In the direct treatment of corpora, the roles of teachers and learners are changed: the teacher is no longer the sole source of knowledge but rather a facilitator of the learning process, and the learner can become more independent and capable of searching the corpus effectively, analyzing the data, and interpreting the outcome (Chambers, 2022).

To clarify this continuum, Figure 1 below shows how DDL implementations differ across key pedagogical dimensions: authenticity of input, learner control, teacher mediation, and locus of learning, as illustrated by Gabrielatos (2005). The continuum highlights that the distinction between these approaches is conceptual rather than purely technological. The “soft” (hands-off) approach allows the teacher to tailor corpus examples to learners’ needs and proficiency, while the “hard” (hands-on) approach maximizes learner agency and engagement with authentic data. More specifically, Gabrielatos (2005) identifies seven key dimensions along which DDL tasks can be situated. First, the size of the corpus ranges from small, manageable datasets in soft DDL to large, authentic corpora in hard DDL. Second, the authenticity of input spans from simplified or pre-selected examples to naturally occurring language, allowing learners to engage with real-world usage. Third, the medium of delivery varies from paper-based materials to computer-based concordance tools, reflecting the technological integration of the task. Fourth, the learning approach differs along a deductive–inductive continuum, where soft DDL often guides learners toward established rules, while hard DDL encourages discovery and pattern recognition. Fifth, teacher mediation ranges from highly teacher-led instruction to learner-centered facilitation, emphasizing autonomy in harder DDL approaches. Sixth, the locus of learning shifts from in-class guided activities to independent, outside-classroom exploration, providing flexibility in how learners interact with the data. Finally, the interaction format moves from pair or group work in soft DDL to individual tasks in hard DDL, reflecting the varying degrees of collaboration and independent analysis required. Together, these seven dimensions illustrate that DDL is not a single method but a versatile pedagogical framework, allowing educators to design activities that balance guidance, authenticity, and learner engagement according to instructional goals.

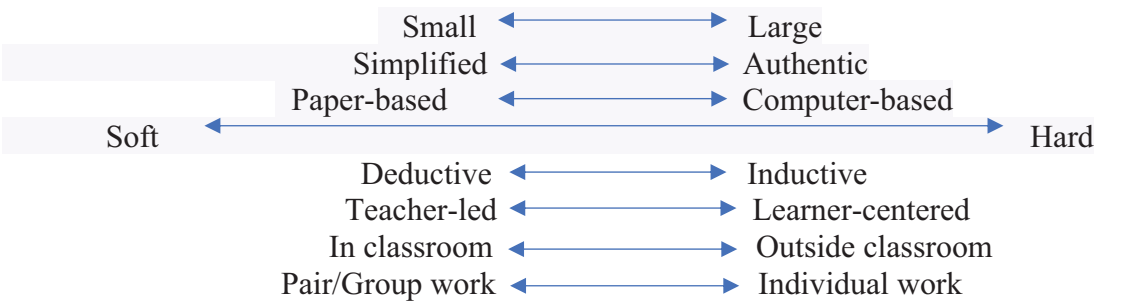


Figure 1 Continuum of possible dimensions of ‘Soft DDL’ and ‘Hard DDL’.

The dimensions that this study employed from the continuum of possible dimensions of DDL included “soft” or paper-based DDL with simplified input and “hard” or computer-based DDL with authentic input, both of which promote inductive, learner-centered approaches of learning through individual and pair/group work. The teaching framework established through these dimensions provided compatibility with explicit language instruction. It reserved a central role for noticing or consciousness-raising and placed the learners in an active role to induce the meaning and use of new language forms and build upon their existing knowledge.

A growing number of studies have examined the effects of direct and indirect DDL approaches, but their results remain inconclusive. While studies report various findings on the learners' gains from DDL, Boulton (2012) suggests that they should be treated with caution, meaning that several factors need to be considered: practical, cultural, individual, and pedagogical. Johns (1991) argued that paper-based DDL can introduce beginners to corpus-based thinking without overwhelming them with software complexity. Boulton (2012) confirmed this, noting that teacher-prepared handouts can enhance comprehension by reducing cognitive load and filtering inappropriate data. Conversely, Boulton and Cobb (2017) conducted a large-scale meta-analysis and found stronger learning gains in studies using computer-based DDL, particularly among higher-proficiency and tertiary-level learners. Meanwhile, Frankenberg-Garcia (2014) demonstrates that guided, paper-based DDL activities foster immediate comprehension and short-term learning, though the long-term retention benefits of authentic corpus interaction are less clear. For lower-proficiency or less technologically confident learners, research such as Yoon and Hirvela (2004) and Granath (2009) highlighted motivational and logistical challenges, including difficulty while interpreting raw corpus data and dependence on teacher mediation. Yet, subsequent studies (e.g., Boulton, 2010b; Lee et al., 2019) showed that even limited exposure to corpus-based materials can enhance language awareness, indicating that appropriately scaffolded DDL can be viable beyond advanced university contexts. Recent empirical and synthesis studies corroborate and nuance these conclusions: for example, intervention and review studies published since 2020 report that (a) scaffolded, teacher-mediated DDL is particularly effective for lower-proficiency learners, (b) hands-on corpus use tends to yield larger gains for higher-proficiency or tertiary learners, and (c) task design, corpus suitability, and implementation context matter substantially for transfer and retention. Representative recent contributions include a 2023 review/meta-analytic update addressing corpus-use effectiveness (Ueno, 2023), empirical studies showing sustainable effects of DDL on collocation learning in pre-tertiary contexts (He, 2024), and intervention research investigating DDL form-focused tasks and student engagement (Zare, 2024). Finally, recent conceptual and methodological discussions (e.g., Flowerdew, 2024) emphasize how advances in tools and interfaces and the integration of AI-supported resources have broadened both the possibilities and the operational challenges of implementing DDL in diverse educational settings.

Both hands-on and hands-off DDL have benefits and limitations, and therefore, each might be appropriate for certain learners, teachers, and contexts. In support of this view, Gilquin and Granger (2022) note that the decision on the type of DDL treatment is crucial. Teachers need to be able to make informed decisions as to whether learners should query the corpus by themselves or use the handouts prepared by the teacher in advance. The choice depends on several factors, including the availability of the necessary hardware and software, the level of learners, and institutional constraints. Thus, DDL researchers and specialists point to various affordances of DDL that language learners can avail of. However, it is still far from "normalization" (a point where a particular technology reaches its fullest possible effectiveness and becomes a "valuable element in the language learning process" (Boulton & Pérez-Paredes, 2024, p. 216).

Despite these advances, several gaps remain that motivate the current study and link directly to its research questions. First, few studies have explored DDL in non-academic or workplace settings, leaving questions about its adaptability to professional learning environments (Research Question 1). Second, there is insufficient evidence on how different proficiency levels respond to soft versus hard DDL treatments, particularly among elementary and pre-intermediate learners (Research Question 2). Third, existing research has rarely investigated learners' attitudes and preferences when both approaches are applied sequentially within the same instructional context (Research Question 3).

In the Armenian EFL context, where grammar instruction traditionally relies on deductive teaching and limited technological integration, investigating these issues carries both theoretical and pedagogical significance. Theoretically, it tests whether the principles of noticing, scaffolding, and inductive hypothesis-building can be effectively implemented through corpus-based materials. Pedagogically, it examines whether DDL can promote learner engagement and independence in contexts where access to computers is limited and teacher mediation remains essential. By addressing these gaps, the present study contributes to the broader discussion on the operationalization of DDL. It provides empirical evidence on how both paper-based and computer-based approaches can be adapted to lower proficiency levels and alternative learning environments, thereby expanding our understanding of DDL's scope and flexibility.

Methodology

Research Design and Corpus Treatment

This study adopted an explanatory sequential mixed-methods research design (Creswell & Plano Clark, 2018), integrating quantitative and qualitative approaches to achieve a comprehensive understanding of the effects of paper-based (“soft”) and computer-based (“hard”) DDL on learners’ grammar development and attitudes. The quantitative phase involved pre-tests, post-tests, and final tests administered across two proficiency levels to measure learning gains following each instructional treatment. This was followed by a qualitative phase consisting of semi-structured interviews designed to explain and contextualize the quantitative results by exploring learners’ perceptions, challenges, and overall engagement with DDL. The integration of these two strands enabled methodological triangulation, thereby enhancing the credibility and validity of the findings.

To address the research questions above, DDL was integrated into the participants’ conventional learning for 10 weeks: 5 weeks paper-based and 5 weeks computer-based instruction. Practically, both groups first received paper-based DDL, followed by computer-based DDL, for which the Corpus of Contemporary American English (COCA) was selected. Both approaches included studying the grammar points presented in their coursebooks and promoted induction of grammar rules on the part of the learner instead of conventional explicit explanation of rules on the part of the teacher, as described in more detail below. In terms of the rationale behind which items were studied using paper versus those that used computer, we moved along the sequence of the coursebooks so that the first five weeks’ grammar points were on paper and the second five weeks’ points were on computer. It should also be acknowledged that this type of random division could be considered a limitation. Since the difference in the level of difficulty of the target items affecting their potential of being amenable to corpus consultation was not controlled, it could lead to varied learning outcomes. However, this was mitigated by the interview data, which will be discussed further.

The corpus-derived handouts developed by the teacher included simplified corpus evidence (e.g. selecting sentences that appeared to be more suited to the level) and encouraged induction of grammar rules, which was carried out in the classroom in pairs or groups. The examples were presented in the KWIC (Key Word in Context) format, that is in the form of a concordance, where the item under study is in the middle, which makes it easier to discern patterns. During the hands-on DDL, the students had access to the authentic data on the computer, which aimed to promote discovery of lexico-grammatical patterns through individual work outside the classroom and discussion of feedback in the classroom. The transition to the hands-on experience was scaffolded with tasks devised and guided by the teacher.

Both types of DDL instruction were based on the four-step teaching framework proposed by Chujo and Oghigian (2008). This study added a fifth step between the first and the third,

and the adopted guided procedure included: (1) forming hypotheses through inductive corpus-driven tasks; (2) sharing hypotheses in groups; (3) verifying the validity of hypotheses with the teacher; (4) practicing the language point in a follow-up controlled exercise; (5) producing the language item through follow-up activities. The rationale behind this choice can be found in the paradigmatic stance undertaken by this study, that is in the shared constructivist and socio-cultural epistemological ground between theory and DDL, where the key pedagogical concepts of noticing, discovery, usage-based learning, cognitive stimulation, and social interaction can be operationalized through the above 5-step model of knowledge construction. As Boulton and Vyatkina (2024) argue, DDL should function as a technological aid, allowing learners to explore and uncover language patterns within a constructivist approach to learning. Firstly, the learners were provided with instructions on corpus work for a particular grammar item and encouraged to induce and form their own hypotheses about the use, form, meaning, and function of the language point by exploring the related concordance lines. At the second stage, they were given opportunity to engage in social dialogue and share their own generalizations about the grammar point in pairs or in groups. To verify the validity of the hypotheses, the teacher provided guidance by eliciting the learners' answers and giving feedback, which allowed the learners to reorganize or confirm their knowledge and arrive at a right conclusion. As following tasks, the practice of the discovered pattern was realized through controlled exercises, followed by its production through free writing tasks.

The learners received training in how to work with a corpus and navigate through concordances, which was a crucial part of the treatment. As mentioned above, COCA was selected to carry out DDL. This introductory training was held at the Training Center of ANPP and lasted for three hours. The participants were introduced to the corpus training tools and their functions on a large screen via a computer and a projector, after which they got registered on COCA by their mobile phones (the center is not equipped with computers sufficient for all learners) and had some practice in operating the tools. Considering the groups' low levels of English language proficiency, the learners were provided with a handout listing the corpus tools and functions, both in English and Armenian. To mediate the challenge of the foreign language, as well as that of acquiring a new skill, the training session was delivered in their L1 (Armenian). This ensured that they were able to gain the necessary skills and knowledge about the direct corpus work. The training proved to be encouraging as it was accepted with enthusiasm and willingness to acquire new skills and knowledge about a new technological learning tool. The corpus functions that they were trained in and subsequently used in grammar tasks are provided below:

- List – Provides frequency information and allows learners to: retrieve multiple examples of a word in context; observe the words that occur before and after the keyword; examine the parts of speech preceding and following the keyword; identify words with the same root; view different inflected forms; and access synonyms.
- Chart – Displays patterns of usage across different registers and over time.
- Collocates – Identifies the words that most frequently co-occur with the keyword, enabling exploration of collocational patterns.
- Compare – Compares how two lexical items are used (e.g., *beautiful* vs. *handsome*).
- KWIC (Key Word in Context) – Presents concordance lines that show the grammatical and lexical patterns in which a word appears.

Participants

The participants of this study were 32 Armenian learners studying English at elementary (A2) and pre-intermediate (B1) levels, according to the Common European Framework of

Reference for Languages (CEFR), as determined by a standardized proficiency test. There were two classes, A2 and B1, with 16 participants in each group. They were employees of the Armenian Nuclear Power Plant (ANPP), aged 30–50, with the gender proportion of 30% female and 70% male in each group. They had been selected by the management of the plant to study English. The aim was to get access to job-related reference materials that are only available in English, to participate in international seminars and conferences, to make good presentations and write reports in English. It should be mentioned that this cohort of participants are not typical of DDL studies where more often the participants are university students with intermediate and higher proficiency levels of English, as noted by the meta-studies on DDL (e.g. Boulton & Cobb, 2017; Lee et al., 2019; Pérez-Paredes, 2019). Each group had classes twice a week, for three hours each session, and the textbooks they used were *Interchange 1* and *Interchange 2* authored by Jack C. Richards. The classes were held at the Training Center of ANPP and delivered by the author. To minimize potential bias arising from the dual role of instructor and researcher, structured lesson plans were strictly followed, all DDL activities were standardized across sessions, and learner performance data were collected and analyzed systematically without subjective interpretation. Additionally, the study focused on objective measures of learner engagement and grammar development rather than personal evaluations, ensuring that the research findings remain reliable and valid despite the instructor-researcher role.

The study used the principle of convenience sampling where the researcher selects the research sample based on ease and proximity to the researcher: the students are easy to reach and easy for the researcher to contact and, therefore, convenient in her research (Creswell & Creswell, 2018; Dornyei, 2007). The participants for this research were the author's student cohorts at that period, and there were a few reasons that necessitated their recruitment. There were no changes in the learners' academic timetable, and the new pedagogical treatment of direct and indirect DDL was possible to be implemented in a real class time integrated into the conventional EFL classroom. And no less importantly, the genuine relationship, having been established between the author and her students, allowed for more authentic and reliable responses. In the meantime, to avoid the issue of the conflict of interest, student interviews were administered by a third party. Participation in the interviews was based on the participants' consent, and all the 16 learners in each group agreed to be interviewed.

Evaluation Data

The study pursued a mixed-method design. The quantitative data were obtained from pre-tests, post-tests, and final tests, whilst the qualitative data were collected from the semi-structured student interviews.

To assess the intervention effect, post-tests were administered after each new treatment. Thus, the groups took a separate pre-test before the DDL instruction, corresponding to their proficiency levels (A1, B2). After the first month of paper-based DDL instruction, both groups did a post-test, and after the second month of computer-based instruction, they sat a post-test and, in the end, a final test relevant to each level of proficiency (A1, B2). All the tests were based on the grammar points covered throughout the 2-month instruction and included completion or gap-filling activities each comprising several points to measure the performance on each grammar item. The post-tests and final test designed for the elementary group included grammar tasks related to: *simple present*; *time expressions*; *demonstratives*; *present continuous*; *quantifiers*; *adverbs of frequency*; *there is/there are*; *simple past*; *present perfect*; *modal verbs for suggestions*; *modal verbs for requests*; *expressing agreement*; *comparative and superlative forms of adjectives*; and *future with present continuous and be going to*. The

tests developed for the pre-intermediate group assessed the learners' grammar knowledge on: *simple past; used to for habitual actions, quantifiers with count and non-count nouns, evaluations and comparisons with adjectives and nouns, simple past vs. present perfect; future with be going to and will, modals for necessity and suggestion; two-part verbs; requests with modals and Would you mind...?; infinitives and gerunds for uses and purposes; imperatives and infinitives for giving suggestions; conditional sentences with if clauses; passive (simple present, simple past); present perfect continuous; participles as adjectives; relative pronouns for people and things.*

The evaluation data from the pre- and post-test results were assessed through Wilcoxon signed-rank test (the non-parametric equivalent of a paired t-test). This non-parametric test was selected because the sample size in each group was small ($n = 16$), and the distributional assumptions required for a parametric paired t-test could not be reasonably established. Given that learner test scores at lower proficiency levels often display non-normal patterns and are based on discrete, bounded scales, the Wilcoxon test offered a more appropriate and robust procedure for detecting changes in performance across the two instructional phases. Despite the issue of the difference in the level of difficulty of hands-on and hands-off grammar points, which the study failed to control, the analysis of the qualitative data confirmed the findings from the quantitative data. The statistical analysis of the test results, conducted for each group, revealed the impact of both the computer-based and paper-based methods on the learners' grammar competency. This provided answers to the first research question that aimed to find out if there was a difference in the learning outcomes between paper-based handouts and computer-based hands-on DDL for each proficiency level. The comparison of the test results between the groups revealed the difference in the outcomes between different levels of language proficiency, thus answering the second research question.

The evaluation data also included semi-structured interview data, which explored the learners' engagement in the learning process and their attitudes and preferences towards the different approaches of DDL, addressing the third research question. More specifically, the questions sought answers on their attitudes towards studying grammar, the contribution of DDL, both on paper and on computer, to their grammar knowledge, their preferences for discovery of rules on their own and direct explanation of rules, the benefits they gained and the challenges encountered when working with concordance lines, and their motivation related to further use of language corpora. The interviews were administered in focus groups at the end of the DDL treatment. They were conducted in Armenian to ensure that the interviewees were able to fully and accurately convey their ideas and feelings. Participation in the interviews was based on the participants' consent. The interview data were translated from Armenian into English, transcribed manually with color-coded tags to facilitate the systematic identification of patterns under each pre-defined theme and analyzed according to these themes. To enhance the trustworthiness of the analysis, coding decisions and emerging themes were reviewed and discussed with a peer researcher experienced in qualitative data analysis. Any ambiguities in interpretation were refined through iterative reading and cross-checking of transcripts. This process ensured that the themes accurately represented the learners' perspectives while maintaining transparency and consistency throughout the analysis.

Results and Discussion

The discussion will be elaborated in the manner to address the research questions of the study in light of the findings observed through the evaluation of the quantitative and qualitative data.

Research Question 1: Is there a Difference in the Learning Outcomes Between Paper-based Handouts and Computer-based Hands-on DDL at Lower Levels of Language Proficiency?

The assumption that paper-based and computer-based DDL might differently affect the learning outcomes of the same student cohort was verified by the assessment of learner performance both after the “soft” and after the “hard” DDL instruction. The analysis of the pre- and post-test results for the elementary group, as illustrated by Table 1, documented an insignificant effect of the real, unmodified language, but the substantial contribution of adapted handouts, on the learners’ performance. This difference was also observed between the two post-test results obtained after the ‘soft’ and ‘hard’ instruction.

Table 1 Wilcoxon test for elementary group between pre-test and post-tests (N = 16).

	<i>Descriptive statistics</i>					
	<i>M</i>	<i>SD</i>	<i>W obs.</i>	<i>df</i>	<i>W critical</i>	Sig. (2-tailed)
Pre-Test	68.4	12		15		0.05
Post-Test (after paper-based DDL)	82.5	11	0		5	
Post-Test (after computer-based DDL)	72	11	4		3	
Final Test	85.2	10	0		5	

This finding suggests that the elementary level of learners can benefit more from the paper-based approach, which is likely to be due to the input being simplified or tailored to the readability and comprehensibility of the level. Although text adjustments have been criticized by some researchers giving central premise to the authenticity of data in DDL (Daskalovska, 2015; Smart, 2014), it can be stated that corpus-based handouts secure better learner comprehension, when the teacher selects the most suitable instances or adapts them to learners’ needs. It is also helpful when it does not seem worth reading hundreds of concordance lines for learning one element. Saying this, we also acknowledge that the representativeness of the language pattern under exploration is affected. This limitation is elaborated by Gabrielatos (2005) explaining that these putative corpus-informed texts sound as inauthentic as the traditional text, since they are packed with an unnatural number of certain linguistic features, which affects the other elements of discourse. At the same time, he warns against “frequency worship”, which can deprive learners of alternative or idiosyncratic choices, and suggests that “less frequent” does not mean “less acceptable”, and that “infrequent” does not mean “wrong”, and that frequencies change according to context of use” (Gabrielatos, 2005, p. 21). The primary evidence above can suggest that even though authenticity of data is sacrificed for its comprehensibility, more increased performance is observed due to the paper-based rather than computer-based DDL at elementary level. In this respect, we can refer to Boulton (2011, p. 575) claiming that “boundaries are fuzzy, and any identifiable cut-off point will necessarily be arbitrary”. While Cresswell (2007) and Gabrielatos (2005) propose that experience in “soft” or “deductive” DDL can bring learners to a level of competence which will allow them to effectively use “hard” or “inductive” DDL (Cresswell, 2007; Gabrielatos, 2005), this study demonstrated that the elementary level may as well begin with the combination of “soft” but inductive dimensions of DDL. The

challenge does not appear to be with the inductive means of learning but with the comprehensibility of input learners are exposed to.

Similar analysis of the pre-intermediate group’s performance, as provided in Table 2, showed that the latter benefited from direct and indirect uses of DDL almost equally. This highlights that the differences in paper-based and computer-based DDL instruction are expected to bring about varying degrees of success for learners with different proficiency levels. For the pre-intermediate group, hands-on corpus experience was beneficial as they were able to take advantage of the richness of the input and, consequently, to take more control over their own learning.

Table 2 Wilcoxon test for pre-intermediate group between pre-test and post-tests (N = 16).

	<i>Descriptive statistics</i>					
	<i>M</i>	<i>SD</i>	W obs.	<i>df</i>	W critical	Sig. (2-tailed)
Pre-Test	60	7.42		15		0.05
Post-Test (after paper-based DDL)	67	8.5	3		5	
Post-Test (after computer-based DDL)	70	9	0		5	
Final Test	74.5	11	0		5	

The results give confidence to state that the quality of the output is largely determined by the quality of input, which enabled the pre-intermediate group to turn the possibility of corpus work into reality. The relevance of the corpus input allowed them to investigate real language use and more detailed patterns, which gives a more accurate and representative picture of a target item, to expose themselves to diverse contextual uses of the target forms, to raise their consciousness of language beyond its monolithic representation, and therefore, gain more clues for inducing the meaning and use of lexico-grammatical patterns. Direct corpus consultation also prepares learners for independence, as a long-term asset. In this respect, it is worth mentioning Johns’ (1991) description that hands-on use of corpora is the epitome of induction. It is more learner-centered and allows access to more factual linguistic performance. The discovery of patterns by learners leads to the construction of their own linguistic competence and gives more control over their own learning.

Research Question 2: Is there a Difference in the Learning Outcomes Between Different Levels of Language Proficiency?

To provide an answer to this research question, a comparative evaluation of the post-test results of the paper-based instruction was conducted between the elementary and pre-intermediate groups, as presented in Table 3. The computation revealed that the gains from the paper-based approach were twice as much for the elementary group as for the pre-intermediate group. However, given the differences in the studied language items, which could have affected the learning outcome, it did not provide enough confidence to suggest that the elementary benefited more than the pre-intermediate group. The comparison between the post-test results of the “soft” instruction and those of the “hard” instruction increased the reliability of the findings that the

elementary group achieved almost four times as many gains from the paper-based DDL as from the computer-based. This suggests that the former approach is more compatible with the elementary proficiency level. One explanation that should be noted in relation to this outcome is the above, acknowledged as a limitation, variable of the difference in the difficulty level of grammar items. Easier items were presented for hands-off work, which could increase the likelihood of more learning gains. The reason for this outcome could also be the influence of the (un)suitability of grammatical items for corpus consultation, which was not minimized. However, another possible explanation lies in the nature of the hands-off treatment. During paper-based activities, the teacher had greater control over key variables such as the readability of input, the quantity of examples, the level of cognitive demand, and the degree of mediation provided. This control resulted in more comprehensible input for lower-proficiency learners. The qualitative interview data supported this interpretation: participants from the elementary group consistently reported that the paper-based materials were “clearer” and “less overwhelming” than the computer-based tasks. Consequently, the structured guidance and reduced cognitive burden appear to have led to better learner performance during the paper-based phase. To elaborate on how hands-off input works better for low-level learners, it can be stated that the input is simplified on the paper leaving out the difficult or irrelevant language in the corpus. Activities are tailored according to learners’ needs and abilities, which is more motivating. The amount of input is reduced through hands-off work, and learners do not have to deal with countless concordances, which is often overwhelming. Thus, the hands-off work shelters learners from many problems of working with raw corpus data (Frankenberg-Garcia, 2014), increasing the comprehensibility and learnability of the target language.

Table 3 Gains on post-tests and final test for elementary group and pre-intermediate group.

Group	Pre-Test Means	Post-Test (after paper-based DDL) Means (M) and Gains (G)	Post-Test (after computer-based DDL) Means (M) and Gains (G)	Final Test Means (M) and Gains (G)
Elementary	$M = 68.4$	$M = 82.5$ $G = 14.1\%$	$M = 72$ $G = 3.6\%$	$M = 85.2$ $G = 16.8\%$
Pre-Intermediate	$M = 60$	$M = 67$ $G = 7\%$	$M = 70$ $G = 10\%$	$M = 74.5$ $G = 14.5\%$

A similar comparative analysis was conducted between the elementary and pre-intermediate post-test results of the computer-based treatment. It showed that the pre-intermediate learners were able to benefit from the hands-on corpus experience three times as much as the elementary group. This finding suggests that the elementary group did not succeed in taking advantage of the computer-based work as the challenges here were not mitigated, but made more gains from the tailored input. The challenges resulting from unmodified input seem to have been overcome by the pre-intermediate level leading to increased performance. The calculations of the post-test results of the pre-intermediate group between these two approaches indicated almost similar performance, which can mean that handouts could be as effective as hands-on corpus work at least for low levels of language proficiency and for short-term learning outcomes. The analysis of the final test results between the two student cohorts, administered at the end of the whole DDL treatment, recorded a very similar increase in learner performance. Both the elementary group and the pre-intermediate group benefited from the combination

of the direct and indirect uses of DDL. The findings of this cross-analysis appear to be in line with the perspective that DDL can be beneficial for low-level learners (e.g. Boulton, 2012; Yoon & Hirvela, 2004), thus contradicting the belief that only advanced learners can be entrusted with direct computer-based work (e.g. Granath, 2009). The findings can also serve as primary evidence for the statement that paper-based work with highly controlled activities can provide a gentle lead-in to more open-ended individual work and prepare learners well for further computer-based work (Gabrielatos, 2005).

Research Question 3: What are the Learners' Preferences and Attitudes Towards the use of DDL on Paper and on Computer?

The analysis of the semi-structured student interviews enhanced or confirmed perceptions derived from the quantitative data about a number of issues related to the direct and indirect uses of DDL. The questions sought answers on their attitudes towards studying grammar, the contribution of DDL, both on paper and on computer, to their grammar knowledge, their preferences for discovery of rules on their own and direct explanation of rules, the benefits they gained and the challenges they experienced from concordance, and their motivation related to further use of language corpora. The analysis of the semi-structured interviews yielded two broad thematic categories that reflected learners' experiences and attitudes toward DDL: (1) perceived benefits of DDL and (2) challenges encountered. While these categories were predetermined, they also emerged inductively during the iterative coding process. Initial open coding of the translated transcripts produced descriptive codes related to motivation, comprehension, difficulty level, autonomy, and attitudes toward technology. These codes were then grouped into conceptual categories through axial coding based on thematic similarity and frequency of occurrence. The discussion of this qualitative data will be grouped around two broad themes the study was concerned with. It will focus on the benefits and challenges of DDL, as reflected in the learners' attitudes towards the new pedagogical intervention, and go on to provide the teacher's reflection on certain issues related to this teaching practice.

Benefits of DDL

At the beginning of the interview, the participants stressed the importance of studying grammar for their professional needs. They mentioned that the inductive approach of DDL, both on paper and on the computer, was a new learning experience for them. In Armenian EFL classrooms, English grammar instruction is commonly deductive and accompanied with direct explanation of rules. In terms of preferences, the elementary group favored the paper-based approach primarily due to the simplified input (i.e. mediated by the teacher to suit their level, as discussed above). This mitigated the feeling of irrelevancy of authentic language and facilitated the discovery of linguistic rules.

Interviewee 6 (Elementary): *Well, the method itself seems very useful, because it encourages you to engage with the language more, but working with COCA was too challenging – many of the sentences were completely incomprehensible for me. I liked more working with handouts; the method was the same, but the sentences were simple and discovering rules was easier and more interesting.*

This result reinforces the finding from the quantitative data, discussed above, that corpus-based handouts are more effective for elementary levels, as the use of prepared materials allows the teacher to tailor activities to learners' needs and abilities (Boulton, 2012) and to

avoid the indiscriminate use of concordances (Frankenberg-Garcia, 2014). As further analysis recorded, the majority of the pre-intermediate group gave preference to the computer-based corpus work, highlighting the new experience, the new knowledge and the skills of new learning tools, the richness of language data, the ease of finding the needed example, the opportunity to engage with real language, as benefits of DDL instruction. Another point experienced in the study was the feeling of confidence that there is a reliable source of language use to resort to in the absence of a native speaker.

Interviewee 9 (Pre-Intermediate): *In my opinion, direct corpus experience was more useful. Yes, that's true, there were many examples I was not able to understand, but it was impressive to see so many examples. In fact, you can meet all the possible examples that are used in the language. Handouts were not so attractive – it was like usual work, but the method was new. Now, when I know how to use the corpus, I feel more confident. And, in general, any new skill or knowledge can give more confidence.*

In Gabrielato's (2005) words, corpus-based introspection, as an alternative to native speaker introspection, provides additional insights into language structure and use. However, all this makes sense if learners are proficient enough to avail of these opportunities, and this availability was manifested not only in their positive attitudes towards the direct corpus work, but also in their improved learning outcomes, as presented above. Learners' preference for direct hands-on concordance is encouraging since it provides more opportunity for securing long-term benefits such as noticing, language awareness, implicitness, and autonomy. Although the 5-week hands-on instruction was not sufficient to record this advantage, it can be suggested that through minimal support and mediation, termed "scaffolding" by Vygotsky (a term coined by Psychologist Jerome Bruner (1966)), which can include strategic questioning by a teacher, structured collaboration, and dialogues within and between groups, learners can achieve self-regulation – an ultimate goal for any education.

Challenges of DDL

The study brought to light some of the fears and challenges of using DDL. One of them is in line with Boulton's (2012) finding that the majority are generally favorable to the DDL approach as a whole, but after some time, they become less motivated, which could be explained by the idea that any novelty appeal diminishes over time. With regard to the preference between the discovery of language rules and the direct explanation of rules, although most preference was given to the second, this was merely due to the opportunity to spend short time on learning rules.

Interviewee 4 (Elementary): *I like studying foreign languages, but my time is so limited that I can't spend a long time on it. Especially working with COCA required a long time – it has a lot of unnecessary sentences, and most of them are difficult to understand.*

Meanwhile, they felt motivated and encouraged by discovery learning, which, as reported by them, facilitated the retention of knowledge.

Interviewee 9 (Pre-Intermediate): *In the past, when I needed to correct my mistakes or find information, I did Google search, but now I can do it with*

COCA as well. I like new learning programs and new learning processes, so my attitude towards the corpus is very positive. With the help of the corpus, my learning process is becoming more interesting.

Thus, the reservations expressed towards DDL are not as much about the inductive mode of learning, nor about being sidetracked by technology, as about the exposure to indiscriminate language input.

Interviewee 4 (Pre-Intermediate): Sometimes I have difficulty choosing good examples or sentences – they can be from difficult and boring texts, which I can't translate. So, I have to read all the examples. But there are so many of them that you can find at least a few relevant sentences and form an understanding about that grammar point.

This created a number of challenges – the learners became less motivated after a while as they had to deal with information in concordances that were difficult to understand because of insufficient knowledge of the target language. As a result, they had to read numerous examples to be able to select good samples and, therefore, invested a large amount of time. For the elementary group, a lot of information in concordances was viewed as unnecessary, as the search queries resulted in too much, too little, or too complex data which they were not able to comprehend, which is why they gave preference to simplified handouts. This finding was also mirrored in the elementary group's performance, which demonstrated higher results on the post-test of the paper-based instruction than that of the computer-based instruction. As the meta-studies on DDL (e.g. Boulton & Cobb, 2017; Lee et al., 2019) note, most studies use DDL for intermediate and higher proficiency levels in academic settings as a writing resource, claiming that DDL works best for these variables. However, this study found that both the elementary group and the pre-intermediate group benefited from paper-based and computer-based work, respectively. Moreover, while in most studies the participants are full-time students, the cohorts of this study are taking their English classes in the workplace. This suggests that the use of DDL can be extended beyond the confines of academia to professional settings, beyond higher levels to lower levels of English, and beyond teaching lexis or writing to lexico-grammatical constructions. These lead to certain considerations on the part of the teacher, which will be the focus of the following discussion.

Teacher's Perspective on the Study

The DDL instruction for both groups was delivered by me, and insights were drawn from this experience as a teacher. It is apparently true that the knowledge and skills of using corpus tools equips EFL teachers with confidence that the language they bring into classrooms is “the real world of language use” (Reppen, 2022, p. 15); that features about language that had eluded our intuition are brought to light through corpora (O’Keeffe et al., 2007); that they know the source of descriptive insights for language teaching/learning, which can directly influence the learning process (Bernardini, 2000).

The literature on DDL reports several concerns for teachers with the use of language corpora for pedagogical purposes. The concerns are related to the lack of computers, insufficient DDL training, the distortion of authenticity of language input, and the overwhelming experience of designing corpus-driven materials (Reppen, 2022; Zareva, 2017). The experience in the study as a teacher showed certain ways to address these concerns. Related to technological considerations, it can be suggested that both the lack of computers and the lack of class time can be

mitigated when corpus tasks are assigned for homework and class time allocated to soliciting and refining learners' hypotheses through feedback discussion in the classroom.

Reflecting as the teacher in the experiment, it became apparent that introducing DDL is a "disrupter". This needed to be mitigated and mediated in my role as the teacher. Using DDL challenges the teacher as well as the student, it forces changes in action and mindset. It also requires confidence and conviction on the part of the teacher. On my part, I had completed a doctoral module on Corpus Linguistics and Language Teaching within my structured program and so I felt equipped for the challenge of the undertaking and confident that it was worthwhile. From the teacher's perspective, I also recognize that learners have to be trained in the use of corpora and data-driven procedures. In essence, this means that the teacher, using DDL, is not just trying to promote inductive learning of the rule, but learning *how* to discover the rule. This follows Zaki's (2016) view that it is no less important for them to understand the benefits and know-how of corpus use, which will lead to more engaging and cognitively conscious language learning process. It should also be recognized that over-reliance on DDL cannot serve as a standalone approach to language instruction. Corpora provide evidence of actual reality of language use, serving as a source for checking out intuitions. Thus, DDL should be treated as an enhancement, rather than substitution for the present teaching methodologies. Learners' training in corpus use can be provided through a three-hour workshop, which was accepted by the participants of the study with enthusiasm. While direct computer-based work can increase time investment on corpus training, and learners' readiness to take up hands-on concordance, it appears to be more effective in terms of securing long-term learning habits. Therefore, for the main study, this was seen as a crucial component, as we discuss below.

The study also showed that the idea of being sidetracked by technology seems to be irrelevant, since the participants demonstrated interest in obtaining new knowledge and new skills that would contribute to their learning. The issue, as discussed earlier, was not with the reluctance to work with computers or handheld devices but with the unmediated direct approach of concordance that caused difficulties in comprehending the raw data. This leads to the consideration of another concern expressed in literature in relation with contrived examples of language, which are often perceived as distorted representations of language. However, it should be noted that authenticity is not necessarily a predictor for efficiency at low levels of language proficiency, as the study revealed. If we are to ensure that learning takes place, we should contextualize the input to learners' "zone of proximal development" (Vygotsky, 1978, p. 86). This can be done through teacher-designed materials based on careful evaluation of corpus results in light of the pre-set pedagogical goal. In this respect, it is worth quoting O'Keeffe and Farr (2003), "The more teachers know about corpora and how to use them, the more they will be empowered to evaluate corpus-based materials objectively" (p. 412). This experience confirmed the concern that the development of corpus-based tasks is time-consuming; nevertheless, there seems to be one possible way to address this issue – as teachers, we will need to sacrifice time if we are called to contribute to long-term learning assets. Thus, to be able to overcome all these challenges, in-service teachers need to receive training in corpus instruction (a long-standing call as noted in O'Keeffe and Farr (2003)).

Insights and Conclusion

The underlying rationale of the study was to gain primary evidence for the concept of operationalization of DDL, which is reported inconclusively in literature on DDL. The following issues and insights were drawn:

- It can be agreed that DDL should be accessible to learners with different needs and different preferences, but it should also be noted that different language proficiency levels will require different dimensions of the DDL continuum with different degrees of scaffolding and teacher mediation.
- DDL has to be treated as a supplementary rather than substitution for the current teaching practices – a corpus provides evidence of real language use, serving as a source for validating intuitions. However, it does not directly, but most rather indirectly, enable us to exploit the virtual reality of language.
- The cross-analysis of the quantitative data demonstrated that the pre-intermediate (B1) group succeeded in the hands-on DDL experience, while the elementary (A2) group benefited more from the handouts and struggled with direct corpus work. This suggests that the computer-based DDL can be applied with the pre-intermediate level if we are to secure long-term learning assets, such as language awareness, implicitness, and self-regulation.
- The issue of ‘technophobia’ was excluded through the interview analysis, when the learners expressed interest in gaining new technological skills for learning purposes.
- The study confirmed the efficiency of assigning DDL tasks as homework in terms of saving time and access to computers.
- Even though the pre-intermediate group demonstrated high gain rates due to the direct corpus work, there was still the reported difficulty of working with authentic data. This suggested that the corpus-related tasks should be mediated by the inclusion of instructions guiding learners towards a certain ‘selective’ but authentic input.
- The research helped to audit the scale of the work involved for the teacher in the development of the corpus materials. While the time is both crucial and worthwhile, it is important to be aware of the work that it entails when scaling up to the main practice.
- Another consideration emerging from the study was that DDL cannot be equally beneficial for all language teaching points. Hence, the need for the identification of those language areas that can particularly benefit from DDL treatment – an aspect which was not dealt with by the study and would inform a more conscious use of language corpora for future practices.

As an explanatory sequential mixed-methods study with 32 participants divided into two groups, the research was limited in sample size, which constrains the generalizability of the findings. Although the Wilcoxon signed-rank test, a non-parametric equivalent of the paired *t*-test, was appropriate for small samples and non-normally distributed data, the limited number of participants may reduce statistical power and the ability to detect small effect sizes. Additionally, the random assignment of grammar items to paper-based and computer-based phases may have introduced variability in task difficulty. The qualitative data, while rich and contextualized, relied on self-reported perceptions, which may be subject to recall bias. Future studies with larger, randomly assigned samples and controlled linguistic variables could yield more robust and generalizable results.

This study investigated the differences in the use of direct and indirect approaches of DDL in grammar instruction with lower-level learners (at A2 and B1 levels of proficiency), thus challenging the dominant view in the literature that DDL is only beneficial for advanced learners and gaining primary evidence for minimizing its operational uncertainties at low levels. We also note that the study was carried out in an Armenian EFL context where the integration of DDL into conventional language learning was the first attempt. Moreover, the study was conducted in a workplace with adults in a work-based training context rather than in the more often used university setting where full-time students are the participants of DDL studies. It has offered

insights into the benefits and challenges of both approaches in this context while also adding to the body of existing research on DDL.

References

- Bernardini, S. (2000). *Competence, capacity, corpora: A study in corpus-aided language learning*. CLUEB.
- Boulton, A. (2010a). Data-driven learning: On paper, in practice. In T. Harris & M. Moreno Jaén (Eds.), *Corpus linguistics in language teaching* (pp. 17–36). Peter Lang.
- Boulton, A. (2010b). Learning outcomes from corpus consultation. In M. Moreno Jaén, F. S. Valverde, & M. C. Pirez (Eds.), *Exploring new paths in language pedagogy: Lexis and corpus-based language teaching* (pp. 129–144). Equinox.
- Boulton, A. (2011). Data-driven learning: The perpetual enigma. In S. Goźdź-Roszkowski (Ed.), *Explorations across languages and corpora* (pp. 563–580). Peter Lang.
- Boulton, A. (2012). Hands-on/hands-off: Alternative approaches to data-driven learning. In J. Thomas & A. Boulton (Eds.), *Input, process and product: Developments in teaching and language corpora* (pp. 152–168). Masaryk University Press.
- 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. (2024). Expanding methodological approaches in DDL research. *TESOL Quarterly*, 58(3), 1193–1204. <https://doi.org/10.1002/tesq.3269>
- Boulton, A., & Pérez-Paredes, P. (2024). Data-driven learning: Pedagogy and technology. In R. Hampel & U. Stickler (Eds.), *Bloomsbury handbook of language learning and technology* (pp. 212–225). Bloomsbury.
- Bruner, J. S. (1966). *Toward a theory of instruction*. Belknap Press of Harvard University.
- Chambers, A. (2022). What is data-driven learning? In A. O’Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (2nd ed., pp. 416–430). Routledge.
- Chujo, K., & Oghigian, K. (2008). A DDL approach to learning noun and verb phrases in the beginner-level EFL classroom. In A. Frankenberg-Garcia, I. Rkibi, M. R. Cruz, R. Carvalho, C. Direito, & D. Santos-Rosa (Eds.), *Proceedings of the 8th Teaching and Language Corpora Conference (TaLC 8)*, Lisbon (pp. 65–71) Associação de Estudos e de Investigação Científica do ISLA-Lisboa.
- Cresswell, A. (2007). Getting to “know” connectors? Evaluating data-driven learning in a writing skills course. *Language and Computers*, 61, 267–287.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.
- Daskalovska, N. (2015). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning*, 28(2), 130–144. <https://doi.org/10.1080/09588221.2013.803982>
- Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford University Press.
- Flowerdew, J. (2024). Data-driven learning: From Collins COBUILD dictionary to ChatGPT. *Language Teaching*. Advance online publication. <https://doi.org/10.1017/S0261444824000144>
- Frankenberg-Garcia, A. (2014). The use of corpus examples for language comprehension and production. *ReCALL*, 26(2), 128–146. <https://doi.org/10.1017/S0958344014000093>
- Gabrielatos, C. (2005). Corpora and language teaching: Just a fling, or wedding bells? *TESL-EJ*, 8(4), 1–37. <https://tesl-ej.org/wordpress/issues/volume8/ej32/ej32a1/>
- Gilquin, G., & Granger, S. (2022). Using data-driven learning in language teaching. In A. O’Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (2nd ed., pp. 430–443). Routledge. <https://doi.org/10.4324/9780367076399-30>
- Granath, S. (2009). Who benefits from learning how to use corpora? In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 47–65). John Benjamins.
- He, M. (2024). The sustainable impact of data-driven learning on noun collocations (pre-tertiary study). *Humanities & Social Sciences Communications*, 11(1). <https://doi.org/10.1057/s41599-024-04038-6>

- Johns, T. (1986). Micro-Concord: A language learner's research tool. *System*, 14(2), 151–162. [https://doi.org/10.1016/0346-251X\(86\)90004-7](https://doi.org/10.1016/0346-251X(86)90004-7)
- Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. *English Language Research Journal*, 4(1), 1–16.
- Johns, T., Hsingchin, L., & Lixun, W. (2008). Integrating corpus-based CALL programs and teaching English through children's literature. *Computer Assisted Language Learning*, 21(5), 483–506. <https://doi.org/10.1080/09588220802448006>
- Lam, K. N. (2024). An examination of teaching methods for grammar instruction in young learners: A comprehensive review. *International Journal of Science and Management Studies*, 7(1), 127–133. <https://doi.org/10.51386/25815946/ijsms-v7i1p119>
- Lee, J. F. K. (2025). Innovative grammar instruction: Employing corpus-based pedagogy and the 4P model for teaching wh-questions. *Journal of China Computer-Assisted Language Learning*. <https://doi.org/10.1515/jccall-2025-0012>
- Lee, H., Warschauer, M., & Lee, J.-H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, 40(5), 721–753. <https://doi.org/10.1093/applin/amy012>
- Liu, T., & Chen, M. (2024). An investigation into learners' cognitive processes in data-driven learning: Case studies of six learners of Chinese. *Chinese Journal of Applied Linguistics*, 46(4), 544–561. <https://doi.org/10.1515/CJAL-2023-0404>
- Mkrtchyan, A. (2021). *Implementation of group work activities in an Armenian public high school to improve communicative skills* (Master's thesis). American University of Armenia. <https://dspace.aua.am/xmlui/handle/123456789/1931>
- Nassaji, H., & Fotos, S. (2011). *Teaching grammar in second language classrooms: Integrating form-focused instruction in communicative context*. Routledge.
- O'Keeffe, A., & Farr, F. (2003). Using language corpora in language teacher education: Pedagogic, linguistic and cultural insights. *TESOL Quarterly*, 37(3), 389–418. <https://doi.org/10.2307/3588397>
- O'Keeffe, A., McCarthy, M. J., & Carter, R. A. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.
- Pérez-Paredes, P., & Boulton, A. (2025). *Data-driven learning in and out of the language classroom*. Cambridge University Press.
- Pérez-Paredes, P. (2019). 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(5–6), 1–26. <https://doi.org/10.1080/09588221.2019.1667832>
- Reppen, R. (2022). Building a corpus: What are key considerations? In A. O'Keeffe & M. J. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (2nd ed., pp. 13–21). Routledge. <https://doi.org/10.4324/9780367076399-2>
- Schmidt, N. (2022). Unpacking second language writing teacher knowledge through corpus-based pedagogy training. *ReCALL*, 35(1), 40–57. <https://doi.org/10.1017/S0958344022000106>
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Smart, J. (2014). The role of guided induction in paper-based data-driven learning. *ReCALL*, 26, 184–201. <https://doi.org/10.1017/S0958344014000081>
- Teng, M. F. (2024). Do self-efficacy belief and emotional adjustment matter for social support and anxiety in online English learning in the digital era? *Digital Applied Linguistics*, 1, Article 2227. <https://doi.org/10.29140/dal.v1.2227>
- Ueno, S., & Takeuchi, O. (2023). Effective corpus use in second language learning: A meta-analytic approach. *Applied Corpus Linguistics*, 3(3), Article 100076. <https://doi.org/10.1016/j.acorp.2023.100076>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2. *Journal of Second Language Writing*, 13(4), 257–283. <https://doi.org/10.1016/j.jslw.2004.09.001>
- Zaki, M. (2016). Corpus-based teaching in the Arabic classroom: Theoretical and practical perspectives. *International Journal of Applied Linguistics*, 27(2), 514–541. <https://doi.org/10.1111/ijal.12159>

- Zareva, A. (2017). Incorporating corpus literacy skills into TESOL teacher training. *ELT Journal*, 71(1), 69–79. <https://doi.org/10.1093/elt/ccw045>
- Zare, J., Azari Noughabi, M., & Al-Issa, A. (2024). The impact of data-driven learning form-focused tasks on learners' task engagement: An intervention study. *ReCALL*, 36(3), 306–323. <https://doi.org/10.1017/S0958344024000120>

Author Bio

Lilit Avetisyan, PhD, is an English language instructor who has worked with several institutions, including the American University of Armenia, the British Council Armenia, the Polytechnic University of Armenia, and the European University of Armenia. She received her PhD in Applied Linguistics from Mary Immaculate College, University of Limerick, Ireland, where she was a recipient of the MIC Doctoral Award. Her research interests lie in applied linguistics, with a particular focus on corpus linguistics, data-driven learning, grammar instruction, and learner autonomy.

Marine Avetisyan is an EUA Associate Professor, currently holding the position of Dean of the Faculty of Linguistics, Communication and Pedagogy at the European University of Armenia. She has over 28 years of experience in TEFL and language pedagogy. Throughout her academic career, she has been actively engaged in curriculum design, language policy implementation, digital language learning, and cross-cultural communication. Her research interests include second language acquisition, corpus linguistics, and education management.