

International Journal of Learning, Teaching and Educational Research
Vol. 25, No. 1, pp. 374-399, January 2026
<https://doi.org/10.26803/ijlter.25.1.19>
Received Oct 26, 2025; Revised Dec 10, 2025; Accepted Dec 16, 2025

Exploring Data-Driven Learning Integration in ESL Grammar Teaching: Insights from Learners and Teachers in the Global South

Ferdinand Luneta Rellorosa* 

Southern Luzon State University and Ateneo De Manila University
Lucban, Quezon, the Philippines, and Quezon City, the Philippines

Abstract. Although Data-Driven Learning (DDL, thereafter) has demonstrated considerable pedagogical advancement since its inception in the early 1990s, its integration into ESL classrooms remains sporadic, particularly in Global South contexts. Despite its growing traction in ELT research and practice, DDL remains underutilized in Philippine secondary ESL classrooms. This study aims to address this gap by examining the efficacy and practical feasibility of DDL in teaching modal verbs to Filipino secondary learners. Employing a quasi-experimental, pretest-posttest design, two instructional groups were compared: one that followed conventional pedagogical methods and the other that utilized a DDL approach. Independent t-tests revealed significantly superior performance among learners exposed to DDL instruction, underscoring its pedagogical value. To enrich the findings, focus group discussions were conducted with five English teachers and five learner-participants from the experimental group. Responses indicated generally favorable perceptions, highlighting increased learner engagement and autonomy. Nevertheless, concerns were raised regarding technological accessibility and the inadequacy of teacher training, echoing issues scarcely addressed in the current literature. Using the lens of critical realism, the study interpreted not only surface-level outcomes but also the deeper mechanisms and structural conditions that shape the success or limitations of DDL in the Philippine context. Hence, findings suggest that while DDL offers a compelling alternative to traditional methodologies, its successful implementation in the Philippine context necessitates sustained investment in teachers guided by existing structural and cultural realities. This study contributes to the expanding discourse on DDL and signals new avenues for research and instructional innovation in underrepresented educational contexts.

Keywords: data-driven learning; corpus-based instruction; corpus linguistics; Philippine ESL classroom; Global South; learner perceptions; teacher perceptions; modal verbs; Grade 8 learners

*Corresponding author: Ferdinand Luneta Rellorosa; ikingrellorosa@gmail.com

1. Introduction

Despite decades of progress in language pedagogy, grammar instruction in many educational systems, including the Philippines, continues to rely heavily on traditional, teacher-centered approaches, leaving innovative, evidence-based methods like data-driven learning (DDL) largely unexplored. Since the development of linguistic corpora in the 1960s, Corpus Linguistics (CL, hereafter) has flourished into a well-established and influential field. It has spurred extensive theoretical and pedagogical discussions, particularly surrounding corpus-based, corpus-driven, and corpus-aided approaches, which have shaped both linguistic description and language education. A central point of debate concerns how CL should be classified: McEnery and Gabrielatos (2006) highlight that some scholars avoid defining it as a theory or subfield of linguistics, preferring instead to describe it as a methodology.

Others equate it with "empirical linguistics" (Samuel, 2001, as cited in McEnery & Gabrielatos, 2006, p. 34), emphasizing its focus on observable language patterns derived from authentic data. As corpora have grown in size and complexity, computer technology has become indispensable to corpus research. The use of specialized software for storing, retrieving, and analyzing massive language datasets has become standard practice, to the extent that corpus linguistics is now often associated with computational linguistics (Leech, 2006, as cited in McEnery & Gabrielatos, 2006, p. 3). Concordancing tools, in particular, enable users to investigate word frequencies, collocational patterns, and language use in context, capabilities that have become more sophisticated over time.

This computational advancement has also given rise to data-driven learning, a pedagogical approach that encourages learners to interact directly with corpora to discover linguistic patterns independently. Coined by Johns (1991), DDL aligns with inductive learning principles and promotes learner autonomy, making it a promising method in English Language Teaching (ELT). Despite its growing global application, DDL remains underutilized and under-researched in the Philippine context, particularly in secondary ESL classrooms.

To examine this issue, the present study employs critical realism as its theoretical lens. This perspective moves beyond merely recording observable improvements in learner performance, which current scholarship has chiefly focused on, to probing the structural, institutional, and contextual conditions that enable or constrain DDL in the Philippine secondary schools. By foregrounding the interplay between learner and teacher agency and systemic realities such as technological access and teacher training, critical realism provides a richer understanding of both the potentials and limitations of DDL in the Global South context – an area of inquiry largely absent from the current DDL debate.

Thus, this paper aims to explore the viability, effectiveness, and perceived challenges of implementing DDL in teaching modal verbs to Grade 8 learners in a Philippine public high school setting. Through a quasi-experimental design and qualitative insights from teachers and learners, this study contributes to the

growing discourse on corpus-based pedagogy and its potential integration in the ELT practices of the Global South.

2. Literature Review

2.1 Corpus Linguistics: Origins and Theoretical Debates

Corpus Linguistics (CL), once considered a marginal field, has matured into a central methodology in linguistics and applied linguistics. Emerging from the need to describe language empirically rather than intuitively, CL offers systematic, data-based methods for investigating real-world language use. The foundational shift began in the 1960s with corpora such as the Brown Corpus and London-Lund Corpus, which enabled linguists to analyze language across genres and registers. With the advent of computational tools, the scale and sophistication of corpus analysis grew exponentially. A key theoretical debate in CL involves the tension between corpus-based and corpus-driven approaches (Tognini-Bonelli, 2001).

The corpus-based paradigm uses corpus data to test pre-existing hypotheses and confirm theoretical frameworks. It is deductive in nature and often complements traditional linguistic theories. In contrast, the corpus-driven approach insists that patterns should emerge from the data itself without theoretical preconceptions, emphasizing inductive exploration. Scholars such as Sinclair (2004) and Stubbs (1996) argue that pure induction is implausible because no analysis is free of hypothesis. Yet, corpus-driven methods still emphasize minimizing theoretical bias. Closely related is the debate on annotation.

While annotation helps render implicit information explicit (McEnery & Wilson, 2001), critics warn that it imposes a theoretical lens that may obscure unexpected patterns. McEnery and Gabrielatos (2006) suggest a balanced perspective, noting that annotated and raw corpora each offer distinct insights and should be used in tandem depending on research goals. This duality reinforces that corpus linguistics is more than a tool; it is a methodological orientation that shapes how linguists think about evidence, hypothesis testing, and linguistic generalization.

2.2 Expanding Reach: Contributions of Corpus Linguistics

Corpus Linguistics has transformed various subfields of linguistics by enabling more empirically grounded, replicable research. Its most notable early contribution lies in language description. Comprehensive grammar references, such as *A Comprehensive Grammar of the English Language* (Quirk et al., 1985) and *Longman Grammar of Spoken and Written English* (Biber et al., 1999), owe much to corpus data. These works provide nuanced, register-sensitive descriptions that traditional grammar often overlook. Corpus methods have also enriched lexicography, resulting in dictionaries such as the COBUILD English Dictionary, which draws on authentic corpus examples. These resources improved learner comprehension by presenting language in use rather than in contrived examples.

Moreover, corpora have become indispensable in describing World Englishes, enabling comparative studies of native and non-native varieties. The International

Corpus of English (ICE) project, for instance, includes subcomponents from over 20 countries, including ICE-Philippines, enabling empirical studies of regional lexical and grammatical features (Bautista, 2011b; Dita, 2011). Specialized corpora have also flourished, catering to niche research needs. The MICASE corpus focuses on spoken academic English, while ICLE captures learner language across L1 backgrounds. These corpora have deepened insights into spoken discourse, interlanguage features, and genre variation, especially in English for Academic Purposes (EAP) and English for Specific Purposes (ESP).

2.3 Corpus Linguistics in Discourse and Genre Analysis

Corpus linguistics has significantly informed discourse analysis, enabling researchers to investigate language beyond the sentence level. Biber and Conrad's (1999) work on register variation and lexical bundles has clarified how different disciplines' structure meaning and argumentation. Similarly, corpus-supported Critical Discourse Analysis (CDA), as demonstrated in Baker et al.'s (2008) study of refugee and immigration discourse, reveals patterns of ideological framing in media language. In genre analysis, corpora help identify recurrent linguistic features in specific text types.

For instance, Hyland (1996, 1998) examined hedging and boosting in academic writing, highlighting disciplinary differences in rhetorical strategies. Harwood (2005) and Martinez (2005) extended this work by investigating the use of first-person pronouns in research articles across disciplines and among native and non-native speakers. These studies underscore the value of corpus-informed pedagogy. Lexical bundles, epistemic markers, and cohesive devices are essential features of academic and professional genres. By identifying these, corpus research informs materials design and EAP instruction, helping learners emulate disciplinary norms.

2.4 Corpus Linguistics in Language Pedagogy

Although most early corpus research was theoretical or descriptive, its pedagogical implications became increasingly evident. The 1990s marked a turning point as scholars began exploring how corpus findings could shape English Language Teaching (ELT). Corpus-informed materials include dictionaries, grammar books, and syllabi grounded in authentic language use. However, these materials still reflect a top-down transfer of knowledge from the corpus analyst to the learner. Learners themselves remain passive recipients of corpus-informed insights. This limitation prompted calls for more active learner engagement with corpora, leading to the emergence of Data-Driven Learning.

2.5 The Emergence and Principles of Data-Driven Learning

Introduced by Tim Johns (1991), Data-Driven Learning (DDL) proposes a radical shift: learners interact directly with corpora to observe, hypothesize, and derive linguistic rules. This inductive approach empowers learners to become "language detectives," uncovering patterns in authentic data. DDL draws on constructivist and discovery-learning theories. It promotes learner autonomy, critical thinking, and long-term retention through active problem-solving. Typically implemented via concordancing tools, DDL activities involve identifying collocations, examining usage patterns, and comparing grammatical structures. Johns and

King (1991) defined DDL as “the use in the classroom of computer-generated concordances to get the learners to explore regularities of patterning in the target language...” (p. iii). Since then, DDL has expanded to vocabulary acquisition, grammar instruction, writing development, and language awareness.

Early studies demonstrated DDL’s effectiveness in vocabulary learning. Cobb (1999) used a corpus of academic texts and a learner-friendly concordance to teach technical vocabulary to ESL learners, showing improved retention and use. Similarly, Chan and Liou (2005) found that Taiwanese learners enhanced their collocational competence through bilingual concordancing tools. Grammar-focused studies are fewer but promising. Vannestal and Lindquist (2007) showed that Swedish university learners engaged in concordancing developed better grammatical awareness, though beginners struggled without scaffolding. These findings highlight the role of learner proficiency and teacher mediation in DDL success.

In writing instruction, Hyland’s (1998) corpus-based research on hedging, modality, and metadiscourse has influenced EAP materials. Biber and Conrad’s (1999) work on lexical bundles has similarly informed curriculum design. A recurring theme is the importance of scaffolding. Studies by Chang and Sun (2009) and Lee and Liou (2003) revealed that learners performed better when corpus activities were guided by structured prompts or teacher feedback. These studies stress that DDL is not self-explanatory; it requires pedagogical design and teacher expertise. Moreover, Leńko-Szymańska (2017) emphasized the need for DDL-focused teacher training. Her study on pre-service teachers revealed gaps in corpus literacy, pedagogical strategy, and technical skills, reinforcing the idea that DDL’s promise cannot be fulfilled without professional development.

2.6 DDL in ELT and English Language Education: Possibilities and Barriers

Despite strong results at the tertiary level (e.g., Boontam & Phoocharoensil, 2024; Cotos et al., 2017; Crosthwaite, 2020; Sun & Wang, 2003; Yoon & Hirvela, 2004), DDL remains rare in secondary education settings. Supporting this observation, Boulton (2008) reviewed 50 DDL studies and found only four focused specifically on secondary learners. This scarcity is often attributed to the persistent assumption that DDL is too complex for younger or less proficient learners.

However, this assumption has increasingly been challenged by empirical evidence. For instance, in a more recent systematic review, Lusta et al. (2023) examined 89 peer-reviewed studies published between 1997 and 2022. They found that DDL is a promising method for language teaching and for improving language learning in secondary education settings, particularly when supported by scaffolding strategies such as customized tasks, peer collaboration, and guided instruction.

Adding to this, Lee and Liou (2003) demonstrated that high school learners can benefit from DDL, particularly when it is guided and when the materials align with their proficiency level. Similarly, Chang and Sun (2009) further showed that scaffolding prompts significantly improved learners’ proofreading skills in a

concordancing activity. Building on these earlier findings, Crosthwaite and Steeples (2024) confirmed that corpus-assisted instruction effectively enhanced secondary learners' use of passive voice in science writing. Despite these documented gains in receptive and productive skills, continued corpus use remained low, largely due to usability barriers and lack of curricular integration. The study reinforces calls for embedding corpus literacy into classroom practice, led by subject teachers using accessible tools.

Elsewhere, Zhong and Wakat (2023) highlight the success of integrating corpora in teaching grammatical constructions in a Chinese high school EFL setting. Their interviews suggest that inductive learning benefits mid-to-high proficiency learners more than deductive approaches. They advocate for collaboration among language experts and emphasize the need to embed corpus literacy in the teacher education curriculum, an argument also supported by other scholars in the field. (e.g., Farr, & Leńko-Szymańska, 2024; Leńko-Szymańska, 2022; Zareva, 2016).

Building on the theme of teacher preparedness, Hsu et al. (2024) explored the corpus literacy of Taiwanese senior high school English teachers and evaluated the impact of an online training workshop on their ability to integrate corpus tools into teaching. Survey results revealed that while teachers held favorable attitudes toward corpus use, many lacked practical knowledge and experience, prompting nationwide workshop using tools like SKELL and Sketch Engine. Post-workshop findings indicated significant improvements in teachers' corpus knowledge and teaching strategies, underscoring the importance of targeted training to support corpus-informed pedagogy.

Finally, in another area of DDL inquiry, Satake (2024) investigates how different types of feedback, teacher, peer, and self, affect the accuracy of error correction when learners use corpora in Data-Driven learning (DDL) settings. The study found that teacher feedback combined with corpus consultation was most effective, especially for correcting omission and agreement errors, whereas peer and self-feedback were less accurate and often misidentified correct expressions as errors. These findings highlight the importance of aligning feedback sources with error types to optimize DDL-based writing instruction and suggest the need for tailored classroom strategies to support corpus literacy and autonomous learning.

While Data-Driven learning (DDL) has long shown success in tertiary education, emerging research increasingly supports its relevance and effectiveness in secondary education, particularly when scaffolded through guided instruction, aligned materials, and collaborative learning. Empirical studies dispel the notion that DDL is too advanced for younger learners, demonstrating its capacity to enhance both receptive and productive language skills when usability and curriculum integration issues are addressed. The evidence also underscores the need to equip language teachers with corpus literacy through sustained professional development and accessible tools. Furthermore, the effectiveness of DDL-based instruction is amplified by strategic feedback mechanisms, particularly when teacher input is paired with corpus consultation. Altogether,

these findings call for a systemic shift that embeds corpus-informed pedagogy within secondary curricula and teacher education programs to fully harness the pedagogical potential of DDL.

Given these developments, there is a pressing need to extend DDL research to Global South contexts, particularly in countries like the Philippines, where secondary education systems face unique pedagogical, infrastructural, and resource-based challenges.

2.7 Corpus Linguistics and DDL in the Philippine Context

In the Philippines, corpus linguistics research has primarily focused on language description, especially of Philippine English (PhilEng). Studies by Bautista (2011a), Dita (2011), and Borlongan (2011) employed ICE-PH to investigate grammatical and lexical patterns in Philippine English. Comparative studies have also examined differences between PhilEng and other varieties such as British, American, and Singaporean English. However, applied corpus linguistics, especially DDL, has been largely absent. There are no known empirical studies that have implemented DDL in Philippine classrooms, particularly at the secondary level. This represents a significant gap in local ELT research and practice.

Despite global developments in corpus-based pedagogy, Philippine ELT remains reliant on traditional methods. Textbooks rarely reflect findings from the corpus, and teachers lack exposure to DDL methodologies. The use of concordancing tools, learner corpora, or hands-on corpus analysis is practically nonexistent in most classrooms. Mukherjee (2006) and Milizia (2013) note similar lags in other contexts, attributing them to insufficient training, technical constraints, and conservative teaching cultures. In the Philippine context, these issues are compounded by infrastructural limitations and a lack of institutional support for digital pedagogies.

2.8 The Present Study: Addressing the Research Gap

Given the demonstrated benefits of DDL in the global south contexts and the lack of local empirical studies, this study aims to integrate DDL into an English class at a public secondary school in the Philippines. Specifically, this study would like to answer the following research questions:

1. To what extent does the Data-Driven Learning (DDL) approach facilitate the acquisition and application of modal verbs in ESL instruction?
2. What perceptions do learners hold regarding the implementation and pedagogical value of DDL in their grammar lessons?
3. How do teachers perceive the integration and instructional effectiveness of DDL within the context of grammar teaching?

2.9 Theoretical Framework

Considering the studies on DDL reviewed above, this study operates on crucial assumptions and considerations for a successful DDL experiment. Nevertheless, before dealing with these assumptions, it is important at this point to establish first the advantages of using corpus data in this study; hence, the central question perhaps in this study is 'why use corpora and corpus linguistic methodologies in

this study' Svartvik (1992, as cited in Lindquist, 2009, p. 9) forwards the following advantages of CL:

1. Corpus data is more objective than data based on introspection.
2. Corpus data can easily be verified by other researchers, and researchers can share the same data instead of always compiling their own.
3. Corpus data are needed for studies of variation between dialects, registers, and styles.
4. Corpus data provides the frequency of occurrence of linguistic items.
5. Corpus data not only provide illustrative examples. But it is a theoretical resource.
6. Corpus data give essential information for a number of applied areas, like language teaching and language technology (machine translation, speech synthesis, etc.)
7. Corpora provide the possibility of total accountability of linguistic features – the analyst should account for everything in the data, not just selected features.
8. Computerized corpora give researchers all over the world access to data.
9. Corpus data is ideal for non-native speakers of the language.

These advantages make CL and corpus methodology particularly enticing for researchers, as they offer promising possibilities for their use in actual ELT practice. Furthermore, there are several pedagogical advantages that an ideal teaching method should normally provide. Below is a summary of the advantages noted by several DDL scholars in their studies. Gilquin and Granger (2010) identify four promising advantages of the DDL approach that no ordinary existing method can provide. These advantages can arguably be considered the prime characteristics of the DDL method in ELT.

First, DDL brings authenticity to an English classroom by using authentic data from the corpus, which comes from naturally occurring use of language in different contexts. This particular advantage enables learner to examine the data independently and discover some idiosyncratic features of a language that are not covered by commercially available textbooks, which are often products of the prescriptivist tradition of language description (i.e., intuition-based). Second, DDL “has an important corrective function” that is not usually present in a typical classroom setup (p. 1).

Learners' firsthand use of the corpora can enable them to verify whether their current knowledge (and perhaps competence) of the language accurately reflects what is expected of them. This is chiefly helpful to L2 learners in appropriating their language use to that of a target language (native corpus). In other words, a corpus can help learners identify and address their own interlanguage issues (i.e., misuse, overuse, and underuse), and teachers can draw learners' attention to these difficulties and place greater emphasis on them in their classes.

Another valuable benefit that DDL offers is the leeway it gives to the learner for discovery learning, broadly defined as an approach “designed to engage learners' inquiry through which, guided by the teacher and materials, they 'discover' the

intended content" (Hammer, 1997, p. 485). Because corpora are readily available and many of the concordancing tools (software) are likewise available free of charge, learners can freely explore the corpus by drawing their hypotheses about the language, carry out investigations in proving or disclaiming these hypotheses, and from there draw empirical conclusions from the analyses. Under this condition, learners are, in effect, considered "researchers" and independent learners (Johns, 1997, p. 101). This technique (inductive learning) promotes learner autonomy by placing a premium on learners' capacity to regulate their own learning, thereby developing their metacognitive skills. However, since the inductive approach has largely been considered suitable only for learners with high proficiency, as explained earlier, it may be inappropriate to apply this approach in an English class dealing with learners at a low proficiency level.

Although this is the case, DDL can still function under a deductive condition with learners of low proficiency levels. Teachers can lead learners to compare the grammatical rules provided in their textbooks with their corpus examples/extracts and see for themselves whether these rules hold. These two approaches in DDL have been classified by Cresswell (2007) as the "inductive DDL" and the "deductive DDL" (p. 270). Furthermore, O'Sullivan (2007) posits that DDL develops fundamental thinking skills, namely "predicting, observing, noticing, thinking, reasoning, analyzing, interpreting, reflecting, exploring, making inferences (inductively or deductively), focusing, guessing, comparing, differentiating, theorizing, hypothesizing, and verifying" (p. 277).

Among these thinking skills, this research study would like to highlight the importance of "noticing," which is particularly relevant to this research, as it focuses on ESL learners. Noticing has been extensively explored in the context of SLA research, as forwarded by Schmidt (1990). On the affective side, DDL also promotes greater motivation among learners. Mair (2002) describes learners as empowered individuals whose confidence and self-esteem are boosted by the autonomous nature of instruction supported by DDL. It is important to note this in this study, as it has long been recognized in the field of language acquisition and language learning (both in L1 and L2) that affective factors, such as motivation, confidence, and self-esteem, play a crucial role in successful language learning and acquisition. These factors have come to be known as 'affective filters' in Krashen's (1982) hypothesis.

In light of the suggestions of some DDL studies underscoring the importance of scaffolding in developing learner's autonomy, this study has adopted the scaffolding prompts used by Chang and Sun (2009), illustrated below:

Table 1: Scaffolding Prompts (Chang & Sun, 2009, p. 288)

Steps	Strategies
Keyword selection	Enter an appropriate keyword or string of keywords. Try a shorter string of keywords. Try different keywords.
Concordance analysis	Read words surrounding the keyword in examples. Pay attention to pattern frequency Skip unclear examples, or link to full texts.
Rule formulation	Look back at the question's keyword and its surrounding words. Compare the keyword in the question with examples. Select the best usage pattern on the basis of surrounding words in questions.
Outcome evaluation	Read examples to confirm whether or not the formulated language pattern exists. Read examples to confirm whether or not a formulated language pattern is used frequently. Review the learned usage patterns and skills.

These prompts have been particularly useful for conducting the experiment and for making the course more learner centered. The prompts' role was to provide a framework for the researcher to guide participants in the DDL activities and exercises. In effect, this scaffolding technique has been beneficial for both the researcher and the participants in the successful implementation of DDL in the experiment. Overall, DDL "helps [learners] to become better language learners outside the classroom" (Johns, 1991, p. 31). All these advantages promote the learner-centered classroom – a paradigm that the current educational system of the Philippines is grounded in.

2.10 Modal Verbs

It is a generally accepted wisdom in the field of grammar (i.e. generative and functional) that modality is one the most complex features of the English language, thus it is no surprising that this grammar feature of English has received quite an attention from grammarians (Coates, 1983; Halliday, 1970; Huddleston, 1984; Palmer, 1979) from all around the world. Studies have examined modal verbs from various grammatical and semantic standpoints, for example, as grammatical markers of stance (Biber et al., 1999; Halliday, 1994; Orta, 2010) and as hedging devices (Hyland, 1996; Hyland & Milton, 1997). Some studies have also explored cross-cultural variation in the use of modal verbs (Bao, 2010; Carrio-Pastor, 2014; Coates & Leech, 1980; Mukundan & Khojasteh, 2011), aiming to outline the similarities and differences in the use of modal verbs across English varieties.

Erhman (1966) defines modal verbs as "... closed class verbs which may occupy the first position of a verb phrase, which may not be immediately preceded by another verb, which may invert with the subject in interrogation, and which is negated by 'not'" (p. 9). Quirk et al. (1985) and Biber et al. (1999) also consider modal verbs as the first constituent in a verb phrase, which is immediately followed by another verb, known as the main verb. Although most grammarians agree on its definition, many differ in their opinions regarding its classification.

This disparity can be traced to the overlapping semantics of these modal verbs. Nartey and Yankson (2014) argue that “classifying modal auxiliaries is by no means unproblematic” as some modals can mean differently depending on how they are used in a sentence. For example, the modal ‘can’ may function differently, as illustrated below by sentence (a) and sentence (b):

- (a) Can you hold your breath underwater for two minutes?
 (b) Can I now enter the room?

In sentence (a), ‘can’ clearly conveys ability, while ‘can’ in sentence (b) connotes permission. This ambiguity of classification makes this particular grammatical feature of English challenging to learn and teach in both ESL and EFL contexts. Hence, Bran-ford (1967, as cited in Nartey & Yankson, 2014) strongly suggests that “[i]t is probably better to avoid labelling any modal too specifically (e.g. ‘may’ = ‘possibility’) but to study each according to the context as one finds it’ (p. 22). This ambiguity and complexity, along with the challenges they pose to learners, particularly ESL and EFL learners, make modal verbs an appropriate subject of study in the researcher’s attempt to test the effectiveness of DDL as a viable approach to grammar teaching.

In the same vein, his proposal explicitly calls for a contextual study of modal verbs; hence, a traditional approach (i.e., armchair linguistics) to describing them should be considered implausible. The researcher believes that the same argument should apply to the teaching and learning of modals. The anomalous and polysemous nature of modal verbs in English arguably requires a more responsive and effective approach to teaching them that captures their complex semantic behavior, which, in turn, can lead to better understanding and learning.

The researcher believes that the DDL approach meets this criterion, as DDL uses a reasonably large amount of data (from a large corpus/corpora) that is arguably representative of a given genre, register, or even an English variety. These data are considered authentic and are derived from the actual language use of their speakers, whether native or non-native. Therefore, the language data used in a DDL classroom were taken from a given context of language use and can pave the way for a ‘contextual’ study of certain grammatical features of English, are modal verbs. Furthermore, this study is limited only to Quirk et al.’s (1985) central modals (i.e., can, could, will, would, might, must, shall, should, may), including their negated and contracted forms.

3. Methodology

3.1 Research Design

Since this study is two-pronged, the study operated under a mixed-methods research design. For the quantitative part, the researcher employed an experimental design using a pretest-post-test method to compare the performance of two groups (i.e., the control group and the experimental group) and determine the effectiveness of the DDL approach in teaching modal verbs. For the qualitative part, two focus groups were convened: one for the learner participants and the other for the teacher observers. Probe and follow-up questions were asked of the

participants. Their answers to the questions were transcribed and thematically analyzed.

3.2 The Corpora

The corpora that were used in this study were the corpora from the International Corpus of English. Three corpora were purposively selected for use in the study: the ICE-Great Britain and the ICE-Philippine component. These corpora were used by the researcher because they are readily available in a concordancing software that is free to use. In terms of composition, the corpora are composed of over a million words coming from 500 texts, with 2000 words per text, 300 for spoken English, with approximately 600,000 words, and 200 for written English, with approximately 400,000 words. This inclusion of ICE-Great Britain was deemed essential as it represents an inner-circle variety of English and thus provides a robust baseline for comparison with the localized Philippine component. This contrastive perspective allowed the study to capture both cross-varietal differences and shared features in modal verb usage, thereby offering a more comprehensive and contextually grounded analysis.

3.3 The Participants

The participants are 60 grade eight (8) learners from a public national high school in the Philippines. They were randomly divided into two groups: 30 learners were assigned to the experimental group, and the remaining 30 were assigned to the control group. The former received modal verb instruction under normal ELT conditions, while the latter was subjected to DDL instruction. For the focus group discussion, two groups were convened: One for the learner participants and the other for teacher observers.

The learner group was composed of ten (10) randomly selected participants from among the 30 learners assigned to the experimental group, while the teacher group was composed of five (5) English teachers invited to observe the DDL group. One of them is the head of the English department, a master's degree holder, who has been teaching in the school for over 20 years. Two of them are master teachers: one holds an MA degree in English Language Education, and the other has an MA degree in Education with a major in English.

Both have been employed by the school for more than ten (10) years now. The remaining two (2) teacher observers were considered novice teachers as both had been teaching in the school for just five (5) years. They are both holders of a bachelor's degree in Secondary Education, majoring in English. The researcher purposively chose these teacher observers for their diverse educational backgrounds, training, and teaching experience as he deemed them valuable in extracting critical insights during the focus group discussion. Before conducting the study, ethics approval was obtained from the Research Ethics Office of the university. Following approval, informed assent was secured from learner participants and informed consent from teacher participants before their involvement in the research.

3.4 Course Design

The experiment ran for about five (5) weeks, excluding pretest and posttest. The DDL grammar course was divided as follows:

3.4.1 Week 1: *Introduction to Corpus Linguistics*

This part of the course aimed to introduce the learners to the basic concepts of corpus linguistics. In particular, it acquainted learners with a brief history, its contribution to other fields, and its application in language pedagogy. Compared to previous studies, which provided a concise introduction to CL, this study offered a week-long discussion about CL. This is in light of the suggestions forwarded by several DDL scholars mentioned above, who noted that ample time for introduction is needed for learners to grasp the concept of using corpus and concordancing software in the classroom. This helped them appreciate its importance and therefore gain a specific connection with the course. This may be arguably longer than usual course introductions but investing sufficient time would do more good than harm for both the learners and the teacher.

3.4.2 Week 2: *AntConc Tutorial*

Here, the learners were exposed to the nitty-gritty of using the AntConc software, as the study employed hands-on concordancing activities. Learners must achieve a certain level of proficiency, if not expertise, in using this software. For the tutorial, videos from Anthony Lawrence's YouTube channel, the owner and inventor of the software, were played for the learners. Learners were encouraged to practice them at home as they were requested by the teacher to download and install the software on their respective laptops and notebooks. This part is deemed necessary by the researcher so the learners can learn the basics of using AntConc, thereby avoiding confusion.

3.4.3 Week 3: *Concordancing Exercises*

This part of the course allowed the learners to explore the three corpora. The researcher provided many individual and small group exercises, ranging from frequency and collocation to grammar and word-based concordancing activities, using scaffolding prompts. The learner participants in class reported on these activities. Here, other intriguing grammar issues, such as the use of prepositions, served as the basis of the activities. Modal verbs were intentionally excluded by the researcher to prevent pre-emption of the experiment and to avoid learners' mastery of the topic.

3.4.4 Week 4-5: *Teaching of Modal Verbs*

This part of the course is where the actual teaching of modal verbs using the DDL approach transpired. Just as the activities and exercises provided in the third week of the course, the teaching also employed a combination of individual, pair, and group concordancing activities, utilizing both inductive and deductive methodologies. The researcher served as the facilitator of the lesson. The study adopted Chang and Sun (2009) scaffolding prompt. It is important to note here that for all the weeks that the course ran, only one (1) hour/day from Mondays to Fridays was allotted for each weekly course activity. For the traditional (control) group, the same procedure used in weeks four and five for the experimental

(DDL) group was employed, except that the traditional approach was applied to the control group.

3.5 Data Gathering Procedure and Instruments

The researcher designed a 30-item test about modals to serve as a pretest and a post-test. Three high school English teachers validated it in the same public national high school where the study was conducted. After validation, it was then pilot tested on another set of grade eight learners from a different section for possible revision. Similar to previous studies, this test served as both a pretest and posttest, administered to the control and experimental groups two weeks before the lesson proper (pretest) and two weeks after the last day of the lesson on modal verbs (post-test). To avoid familiarity with the tests, the pretest results were not discussed or disclosed to the participants, and the items in the post-test were rearranged to prevent resemblance to the pretest.

The exams were checked. Scores were then tabulated and subjected to a t-test to test the hypotheses that were drawn in the first part of this paper. For the focus group discussion part, learner-participants from the DDL group were immediately convened three days after their last lesson on modal verbs. Probe and follow-up questions were asked to elicit their insights regarding their DDL experience. The focus group discussion lasted for almost one hour. The teacher group was convened the next day, after the researcher had delivered the final lesson, to gather fresh insights from the teacher-participants. The interview lasted for almost an hour.

The researcher, before starting the interview, oriented both groups of participants to say that the conversation would be audio recorded purely for research purposes and that their names would not be disclosed to protect their privacy. They were also reminded that they had the option not to answer the researcher's questions. The participants' answers were transcribed and analyzed thematically to gain a clear picture of their insights, perceptions, and opinions about DDL. The researcher invited two inter-coders, both holding an MA degree in English Language Education, to evaluate the recurring themes that he had identified. This ensures the reliability of the themes discussed in the next chapter of this paper.

Any disagreements in the coding were discussed and debated thoroughly until the final version was written. Finally, the questions during the FGD were self-developed and guided by existing questionnaires in the literature. These questions were discussed thoroughly with the adviser for revision. To compare learners' test scores, a two-sample t-test (also known as an independent-samples t-test) was used. This was also used to determine the significant difference between the performance of the controlled group and the experimental group in their pretest and post-test.

4. Results and Findings

The first goal of this paper is to test whether DDL as an approach to teaching modal verbs is effective vis-à-vis the traditional teaching methodology employed by secondary English teachers. To establish this, the researcher employed an

experimental pretest-posttest design in the research. Hence, this section of the paper presents the findings of this experiment coupled with a comprehensive analysis of the data obtained and its implications. Considering the reliability of the test scores, the researcher deemed it necessary to establish first that the randomly selected participant groups were heterogeneous. This is to ensure that both groups are of the same level of competence and knowledge of the topic, even before they were exposed to the experiment. Ensuring this would prevent the scores from skewing the results in favor of one another, thereby undermining the validity and reliability of the test scores.

Beyond statistical significance, it is crucial to evaluate the practical significance of the findings, particularly the magnitude and educational value of the learners' gain after instruction. Although the pretest results revealed no meaningful difference between the two groups, the negligible mean difference of 0.20 and the extremely small *t*-value (0.1273) further underscore that both groups began the experiment at virtually the same level of proficiency. This baseline equivalence is important because it strengthens the internal validity of the study: any subsequent improvement can be confidently attributed to the type of instruction rather than preexisting differences in ability.

Table 2: Mean Scores and Computed *t*-value in the Pretest of the Two Groups

Sample Groups	Mean	Standard Deviation	Mean difference	Computed <i>t</i> value	Interpretation
Control Group	15.87	3.80	0.20	0.1273	Not statistically Significant
DDL Group	16.07	4.76			

Table 2 aims to support this by comparing the mean scores of the two groups and subjecting both to a *t*-test. As can be seen from the table, the computed *t*-value of 0.1273, which is well below the critical value of 2.048 at the 0.05 level of significance or degree of freedom, clearly does not indicate statistical significance, thereby signalling that both groups are at the same level of competence before instruction. This comparison between the control and DDL group's pretest scores is crucial to this study as it needs to statistically establish DDL's effectiveness (or otherwise) as an approach to language teaching. It is essential to note that previous studies (e.g., Tian, 2005) that employed experimental pretest-posttest methods did not explicitly state in their papers that such measures were taken to establish internal reliability between their control and experimental groups.

Table 3: Mean Scores and Computed *T*-value in the Pretest and Posttest of the Two Groups

Sample Groups	Test Versions	Mean	Mean difference	Computed <i>t</i> value	Interpretation
Control Group	Pretest	15.87	2.66	2.3808	Significant
	Posttest	18.53			
DDL Group	Pretest	16.07	6.40	4.8972	Extremely significant
	Posttest	22.47			

When examining the posttest results, a more nuanced understanding emerges as shown in Table 3. The control group demonstrated a statistically significant improvement, increasing their mean score from 15.87 to 18.53—a gain of 2.66 points. While this difference is meaningful within a traditional classroom context, its practical impact appears modest. Based on the magnitude of the gain relative to the group's standard deviation, this improvement corresponds to what is typically considered a small to approaching medium effect in educational research. This suggests that conventional instruction remains effective in facilitating incremental learning of modal verbs, consistent with approaches that rely on explanation, example-based practice, and rule application. However, the modest size of the gain indicates that this mode of instruction may only support superficial or procedural knowledge rather than deeper grammatical competence.

The performance of the DDL group presents a markedly different picture. Their mean score increased from 16.07 in the pretest to 22.47 in the posttest, resulting in a 6.40-point gain. This difference is not only statistically extremely significant, but also represents a large practical effect, suggesting a meaningful and substantial enhancement in learners' understanding and mastery of modal verbs. Estimated effect-size measures indicate that the magnitude of learning in this group is more than twice that of the control group, reflecting a level of instructional impact rarely observed in small-scale classroom interventions. Such a substantial gain implies that DDL enabled learners to engage in deeper cognitive processing—allowing them to infer patterns, analyze authentic examples, and internalize usage more effectively than through teacher-delivered explanations alone.

This contrast between the two instructional approaches illustrates that while both methods lead to improvement, the quality of learning produced by DDL is significantly greater. The DDL group's large gains suggest enhanced retention, a stronger command of usage, and a more intuitive grasp of modal verbs in authentic contexts. These findings align with earlier research (e.g., Lee & Liou, 2003; Sun & Wang, 2003), yet the present study extends these insights by demonstrating that high school learners—often considered less suited to corpus-based learning—can benefit substantially from DDL. The results, therefore, challenge the assumption held by some researchers (e.g., Tian, 2005; Cresswell, 2007) that only advanced or intermediate learners can succeed in DDL environments. Instead, the data from this study affirm Boulton's (2009) assertion that while DDL may be ideal for advanced learners, it can still “bring something to the learning process even at lower levels of ability (p. 14).”

Overall, the results offer compelling evidence that DDL is not only a statistically effective approach for teaching modal verbs but also one that produces substantial, meaningful, and educationally significant learning gains. The practical implications of these findings suggest that incorporating DDL into secondary ESL classrooms may provide learners with richer opportunities for discovery, deeper grammatical awareness, and stronger long-term understanding compared to traditional methods.

The second goal of the paper is to present and analyze the opinions and insights of learner-participants and teacher-observers about DDL. Below is the thematic analysis of their responses during the convened focus group discussion.

4.1 Learner Participants

4.1.1 Theme 1: Difficulty and usefulness

When asked about their initial thoughts about DDL, all participants agreed that it was very demanding. One learner said: *“The class used a lot of activities that are time-consuming but are very exciting to do and very informative indeed”*. Another one answered: *“Mahirap po pero nakaka-enjoy gawin yung mga activities at natututo ka talaga [it is difficult but we enjoyed doing the activities because we really learn]”* This simply shows that despite the challenging nature of DDL as an approach to language teaching, learners still pay attention to the learning that they will get from the activities that they do. It is also surprising that, despite the task-centered nature of the activities, the learner participants deemed them worthwhile. The participants’ approval signals DDL’s applicability and plausibility as a methodology for use by secondary English language teachers in their classrooms.

This also signifies that even low-level learners can gain an appreciation of DDL despite the complexity and difficulty that it brings, thereby furthering the argument and the debate that indeed DDL can be applied not just to advanced learners or university learners but can prove very useful and beneficial to secondary learners as well (Boulton, 2010). This observation is also consistent with findings from international DDL studies involving young or novice learners, which similarly report initial difficulty followed by positive uptake (e.g., Lee & Liou, 2003), strengthening the comparative relevance of the present results. Moreover, from a critical realist perspective, this shows how learner agency operates despite structural constraints such as limited exposure to technology.

4.1.2 Theme 2: Increased attention to the target language

One interesting observation worth noting is that four of the participants deduced that it helped them pay more attention to learning modal verbs. One participant opined that *“nafofocus niya po ako dun sa modal verbs kasi the activities are organized [activities were organized; hence they guide me to the modal verbs]”*. What this learner meant was probably that the activities are highly structured, thereby allowing them to focus more on modal verbs. Another participant explained that: *“systematic at organized po yung lessons kaya nakakapagconcentrate po kami sa modal verbs [lessons are systematic and organized, allowing us to concentrate on modal verbs]”*.

These responses of the learners bring to the table another pedagogical advantage of DDL that has been consistently noted in previous studies (Boulton, 2012, 2015; Flowerdew, 1998; Yoon, 2008), that is, DDL helps learners to focus (noticing) their attention to the target language, in this case, on modal verbs. As such, it is essential to underscore in this paper the crucial role of noticing (Schmidt, 1990) in the successful acquisition or learning of linguistic features in an English classroom. This result parallels international findings where noticing played a mediating role between DDL exposure and grammar gains, suggesting that younger Filipino learners exhibit similar cognitive engagement patterns identified in global studies. In terms of practical implications, this implies that teachers may explicitly

design corpus tasks that heighten noticing – e.g., guided concordance questions, highlighted patterns, or scaffolded worksheets – which can be implemented even in low-tech or paper-based DDL setups.

This result, however, is not surprising, as there are only a handful of DDL studies that have investigated learner-participants' insights into their DDL experience through a survey questionnaire. This implies that a qualitative approach, such as focus group discussions, can be a useful tool in shedding light on learners' perceptions of DDL. This could, in turn, bring valuable insights for the possible rectification of the DDL approach, making it more learner-friendly and more useful in secondary ESL classrooms. Furthermore, this result signifies learners' positive feedback on DDL, thus showing signs of viability as a methodology for language teaching in the country.

4.1.3 Theme 3: Knowledge of corpora and concordancing tools

Ang mahirap po taalaga dun sa lesson eh yung pag gamit ng application (AntConc and ICE) [it was the use of the applications that made the lesson a lot harder]". This is one of the observations made by a participant. This observation exposes a disadvantage of DDL: it requires mastery of concordancing software and the use of corpora. This has also been noted in previous research studies (Allan, 2006; Boulton, 2008, 2009, 2010; Tian, 2005), which is why many researchers opted to use a paper-and-pen interface instead of a hands-on and program interface when applying DDL in their experiments. These researchers believed that it adds a burden on learners and can play a crucial role in learners' test performance.

However, in this study, this did not hinder the participants from outperforming their control group counterparts on the test. This shows that learners do recognize these difficulties and thereby give credence to the importance of beginning a DDL classroom with concurrent training. This finding informs concrete strategies for Philippine ELT: brief orientation modules, teacher demonstrations, peer-support groups, and simplified concordance tasks can scaffold learner interaction with corpora even in low-tech environments. This is the very reason concordancing training was included in this study's course.

It was observed in previous research (e.g., Yeh et al., 2007) that learners need to be trained and oriented first with concordancing tools and software before they are exposed to hands-on concordancing activity during the lesson proper. From a critical realist standpoint, this shows that structural constraints (e.g., digital skill limitations) coexist with emergent mechanisms (e.g., inquiry-based problem solving), both influencing but not fully determining learner outcomes.

4.1.4 Theme 4: Use of language and authentic learning

Learner participants were very keen and excited to examine the corpora and noted that *"nakaka-amaze po na yung inaanalyze po naming examples ay mga totoong salita [it's amazing to learn that we are actually analyzing authentic language use]*". Another learner said *"Totoong language po pala yung pinag-aaralan naming [we are studying authentic language use]*". This demonstrates that the use of corpora makes DDL even more engaging, as learners appear to be interested in examining authentic language data. This helps the learner become more highly motivated, making

learning an enjoyable activity. Research in psycholinguistics and second language acquisition has consistently demonstrated that affective factors, such as motivation, play a pivotal role in successful learning and acquisition of a target language. Furthermore, one participant argued that they become more independent learners, saying that *“binigyan po kami ng chance to examine it [modals in the corpora] on our own [we are given the chance to examine it on our own]”*. The learner's answer is very encouraging and puts DDL in a better position to be used as a methodology in language teaching, as it supports discovery and independent learning, both of which the Philippine Department of Education highly recommends.

The research studies have underscored the learner-centered nature of DDL, as reviewed earlier, and have been its most significant selling feature to date. DDL no doubt increases learner autonomy and puts the learner at the helm of their learning, thereby developing their metacognitive skills. This result supports various studies' (Boulton, 2012, 2015; Turnbull & Burston, 1998) observation that DDL promotes independent learning.

4.2 Learner Participants

4.2.1 Theme 1: Learner-centeredness of DDL

The teacher-observers were in chorus when they argued during the focus group discussion that DDL is supportive of independent and discovery learning. Teacher 1 said, *“This methodology is no doubt supportive of learner autonomy”*. Teacher 2 extrapolates that *“since the approach banks more on inductive method, the approach helps the learner to build their own understanding of modal verbs, hence training them to become more independent learner.”*

These arguments solidify DDL's advantage as an approach to language teaching as current language teaching framework under the new K-12 curriculum centers on learner-centered paradigm thereby signalling its viability as a new language teaching methodology in Philippine secondary ESL classroom. This result is not surprising, as the problem-solving nature of DDL, which taps into learners' independent learning, has been widely acknowledged in the literature. It posits that DDL helps learners become researchers themselves, thereby increasing their metacognitive skills.

4.2.2 Theme 2: Learner-centeredness of DDL

The teacher observers were keen to note that DDL can finally pave the way for teaching the variation of English within their English classrooms. Teacher 3, during the focus group discussion, noted that DDL *“can be the springboard to teach learners the different English varieties as ICE-Corpora are readily available and AntConc is free for download”*. Another teacher also pointed out that DDL *“might be the only methodology that could modernize the teaching of English varieties around the world”*. She further argued that *“I cannot recall a methodology in language teaching that could use a cross-variety approach to teaching English.”*

These unique observations of teachers bring to the table another issue, that has not been touched on yet by existing studies on DDL, that is, using DDL as an approach to teaching world Englishes in a secondary classroom. This is one of the

hotly contested issues in the current ELT and applied linguistics debate in the Philippines. Although the existence of these varieties is now widely acknowledged, it remains a challenge for educators to fully integrate the notion of English varieties into the Philippine secondary ESL classroom. DDL can be used as a tool to begin integrating these into ESL classes, not just for greater awareness and appreciation of the learner, but, more importantly, to foster cultural appreciation among and between these English-speaking nations. This is also an interesting finding, as there appears to be no single study that has examined the possible integration of several ELT models, specifically the construct of world Englishes, alongside the application of DDL in ESL classrooms. This could be an interesting area of inquiry for future research

4.2.3 Theme 3: Complex yet promising methodology and intensive teacher training

Teacher observers did not mince words when arguing that DDL is a fairly complex methodology compared to other approaches. One teacher explained, *“This methodology is complicated than others and would require intensive teacher training, but if mastered, it would be a very encouraging new teaching methodology in our classroom”*. Elsewhere, another teacher observed that *“while DDL requires complex skills from teachers, it would be a handy methodology to use in our class”*. These observations suggest that teachers recognize the complexity of DDL but also acknowledge it to their English class. This suggests the openness of the teacher observers to incorporating them into their daily teaching repertoire.

Moreover, these arguments from the teachers suggest that if ever considered beneficial by authorities, it would require the design of an intensive training program for teachers on CL and DDL, as well as the use of software tools (Lenko-Szymanska, 2017). It can be noted that there are three primary competencies that teachers of DDL must possess namely: technical skills which zeroed in on the proficiency of teachers in using software tools; corpus linguistics skills which focused on the teachers comprehensive knowledge about corpus linguistics; and lastly the teachers’ pedagogical skills which focused on designing and integrating suitable teaching practices that would suit best for corpus-based or corpus-informed language classroom (Callies et al., 2016; Mukherjee, 2006). This would ensure the proper delivery of DDL instruction and ultimately its success as a viable methodology in Philippine secondary ESL classrooms.

It is essential to note that this issue of teacher training has not been adequately addressed in the existing literature or in the current DDL debate. As Lenko-Szymanska (2017) observed: *“The development of a model of effective teacher training in data-driven learning is a challenge that has not been adequately tackled in the literature to date”* (p. 217). This finding underscores the importance of teacher training (both in-service and pre-service) as a crucial component of the success of DDL instruction. Since it is the teachers who will implement this methodology, exposing them to robust DDL training would not only ensure the accurate implementation of DDL instruction in the classroom but also maximize learner learning and participation.

Furthermore, the teachers’ positive feedback shows a promising new insight into the DDL scholarship, as teachers’ reluctance to use DDL in their instruction has

been consistently acknowledged in the literature (e.g., Breyer, 2011; Farr, 2008; Mukherjee, 2006; Römer, 2006; Tribble, 2015; Zareva, 2016). Despite the challenge posed by DDL's heavy technical reliance, teachers seem to lean more on its promising gains than on the difficulty of application. Moreover, in the humble view of this researcher, this signals a promising start for DDL in the Philippines. Arguably, this positive attitude among teachers can be interpreted as strong motivation and a willingness to learn more, expanding their pedagogical horizons to ensure they remain skilled, informed, and relevant in teaching English in their classrooms. As O'Keeffe and Farr (2003) put it: "native and non-native teachers need to learn to manipulate language corpora for their own pedagogic ends..." (p. 412).

4.2.4 Theme 4: Technology-dependent and unsuitable for some school setups

The public-school system's lack of or inadequacy in technology infrastructure is another problem identified by the teachers. One teacher was candid enough to air her frustration regarding the quality and availability of educational technologies not just in their school but also in other national high schools in the province. She argued that *"the limited number of computers that are available in our school cannot satisfy the needs of all learners if DDL is applied in our classrooms"*. This, however, is not surprising as a great majority of public national high schools in the country have a poor computer-to-learner ratio. This would indeed prove very problematic for DDL in Philippine ESL secondary schools.

The good news, however, is that DDL can rely on a paper interface instead of a hands-on or program interface. Studies by Boulton (2008, 2009, 2010), Tian (2005), and Allan (2006), among others, provided robust evidence and methodology for applying paper-based concordancing activities. Their findings could be very instructive in applying a paper-based DDL instruction. This problem opens a new avenue for research inquiry in the Philippines regarding the use of paper interfaces in DDL teaching and learning.

5. Conclusions and Recommendations

Since its introduction in the early 1990s (Johns, 1991), data-driven learning (DDL) has achieved notable international success in ELT research and pedagogy. Studies consistently report positive outcomes across diverse contexts, suggesting that its benefits reflect deeper generative mechanisms in language learning. Yet in the Philippines, these mechanisms remain largely unactivated. The absence of DDL in local research and classroom practice indicates historical, institutional, and infrastructural conditions that have constrained its emergence as a viable pedagogical alternative.

Guided by a critical realist perspective, this study examined both the empirical effects of DDL in a Philippine secondary classroom and the conditions shaping its plausibility alongside existing approaches such as Communicative Language Teaching (CLT), English for Specific Purposes (ESP), and Task-Based Language Teaching (TBLT) (Martin, 2014). Using a mixed-methods design, the study investigated observable events—student performance and stakeholder

perceptions—and the underlying structures that influence these outcomes, including institutional readiness, technological capacity, and teacher expertise.

Quantitative findings showed statistically significant improvements in the experimental group's mastery of modal verbs, indicating that DDL can produce measurable learning gains under conducive conditions. However, qualitative data revealed contingencies that mediate these gains: the need for extensive teacher training to use corpus tools and the persistent inadequacy of technological infrastructure in public secondary schools. These structural constraints limit the sustainability and scalability of DDL despite positive learner and teacher responses.

Overall, the study demonstrates that DDL's limited presence in Philippine ELT stems not from a lack of interest but from deeper institutional, technological, and policy-related mechanisms. Addressing these constraints requires systemic—not merely individual—action. Future research may explore DDL in senior high school settings, compare hands-on and paper-based modalities, and examine its integration within a World Englishes framework. While DDL holds clear potential, its actualization will depend on realigning structural conditions so that, as Mukherjee (2006) notes, corpora can be adopted by “ordinary teachers and learners in ordinary classrooms” (p. 14).

6. Acknowledgments

The researcher sincerely extends heartfelt gratitude to the teacher and learner respondents who willingly shared their time, insights, and experiences to help complete this study. Their active participation and cooperation have been invaluable in making this research possible. Deep appreciation is also given to the Southern Luzon State University (SLSU) Office of Research Services and the University Administration for their unwavering support of my research endeavors. Their encouragement, guidance, and the scholarship grant they have extended have significantly contributed to the pursuit and completion of this scholarly work.

7. References

- Allan, R. (2006). Data-driven learning and vocabulary: Investigating the use of concordances with advanced learners of English. *Occasional paper-Centre for Language and Communication Studies*, (66), 1-52.
- Baker, P., Gabrielatos, C., Khosravini, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & society*, 19(3), 273-306.
- Bao, Z. (2010). Must in Singapore English. *Lingua*, 120(7), 1727-1737.
- Bautista, M. (Ed.) (2011a). *Studies of Philippine English: Exploring the Philippine component of the International Corpus of English*. Mandaluyong, Philippines: Anvil Publishing.
- Bautista, M. (2011b). Some notes on 'no' in spoken and written Philippine English. In M. L. Bautista (Ed.). *Studies of Philippine English: Exploring the Philippine component of the International Corpus of English* (pp. 75-90). Philippines: Anvil Publishing.
- Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. *Language and Computers*, 26, 181-190.

- Biber, D., Johansson, S., Leech, J., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.
- Boontam, P., & Phoocharoensil, S. (2024). Effects of Data-Driven Learning on Writing Complexity, Accuracy, and Fluency (CAF) of Thai EFL Learner. *PASAA: Journal of Language Teaching and Learning in Thailand*, 68, 229-271.
- Borlongan, A. (2011). Some aspects of the morphosyntax of Philippine English. In M. L. Bautista (Ed.). *Studies of Philippine English: Exploring the Philippine component of the International Corpus of English* (pp. 187-200). Philippines: Anvil Publishing.
- Boulton, A. (2008). Looking for empirical evidence of data-driven learning at lower levels. In B. Lewandowska-Tomaszczyk (ed.), *Corpus Linguistics, Computer Tools, and Applications: State of the Art* (pp. 581-598). Frankfurt: Peter Lang.
- Boulton, A. (2009). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, 21(1), 37-54.
- Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60 (3), 534-572.
- Boulton, A. (2012). Language awareness and medium-term benefits of corpus consultation. *Proceedings of the Eurocall 2009 Conference: New Trends in CALL-working together* (pp. 39-46). Spain.
- Boulton, A. (2015). Applying data-driven learning to the web. In A. Leńko-Szymańska & A. Boulton (eds), *Multiple Affordances of Language Corpora for Data-driven Learning*. Amsterdam: John Benjamins, pp. 267-295. [https://DOI:10.1075/scl . 69.13bou](https://doi.org/10.1075/scl.69.13bou)
- Breyer, Y. (2011). *Corpora in language teaching and learning: Potential, evaluation, challenges*. Frankfurt, Germany: Peter Lang.
- Callies, M., Kreyer, R., Schaub, S., & Gűldenring, B. (2016). Towards corpus literacy in foreign language teacher education: Using corpora to examine the variability of reporting verbs in English. *Angewandte linguistik in schule und hochschule*, 391-415.
- Carrió-Pastor, M. L. (2014). Luisa Carrió-Pastor Cross-cultural Variation in the Use of Modal Verbs in Academic English. *Finnish Journal of Linguistics*, (27), 153-166.
- Chan, T. P., & Liou, H. C. (2005). Effects of web-based concordancing instruction on EFL learners' learning of verb-noun collocations. *Computer assisted language learning*, 18(3), 231-251.
- Chang, W. L., & Sun, Y. C. (2009). Scaffolding and web concordancers as support for language learning. *Computer Assisted Language Learning*, 22(4), 283-302.
- Coates, J., & Leech, G. (1980). The Meanings of the Modals in British and American English. *York Papers in Linguistics*, 8, 23-34.
- Coates, J. (1983). *The semantics of modal auxiliaries*. London/Canberra: Croom Helm.
- Cobb, T. (1999). Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning*, 12 (4), 345-360.
- Cotos, E., Link, S., & Huffman, S. (2017). Effects of technology on genre learning. i(3), 104-130. Retrieved from <http://llt.msu.edu/issues/october2017/cotoslinkhuffman.pdf>
- Cresswell, A. (2007). Getting to "know" connectors? Evaluating data-driven learning in a writing skills course. In E. Hidalgo et al. (Eds). *Corpora in the Foreign Language Classroom* (pp. 267-287). Amsterdam: Rodopi.
- Crosthwaite, P. (2020). Taking DDL online: Designing, implementing, and evaluating a SPOC on data-driven learning for tertiary L2 writing. *Australian Review of Applied Linguistics*, 43(2), 169-195..
- Crosthwaite, P., & Steeples, B. (2024). Data-driven learning with younger learner: Exploring corpus-assisted development of the passive voice for science writing with female secondary school learners. *Computer Assisted Language Learning*, 37(5-6), 1166-1197.

- Dita, S. (2011). The grammar and semantics of adverbial disjuncts in Philippine English. In M. L. Bautista (Ed.). *Studies of Philippine English: Exploring the Philippine component of the International Corpus of English* (pp. 33-50). Philippines: Anvil Publishing.
- Ehrman, M. (1966). *The meanings of the modals in present-day American English*. The Hague: Mouton & Co.
- Farr, F. (2008). Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users. *Language awareness*, 17(1), 25-43.
- Farr, F., & Leńko-Szymańska, A. (2024). Corpora in English language teacher education: Research, integration, and resources. *TESOL Quarterly*, 58(3), 1181-1192.
- Flowerdew, L. (1998). Integrating 'expert' and 'interlanguage' computer corpora findings on causality: Discoveries for teachers and learners. *English for Specific Purposes*, 17(4), 329-345.
- Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching. *The Routledge handbook of corpus linguistics*, 359-370.
- Halliday, M. A. K., (1970). *A course in spoken English: Introduction*. Oxford: Oxford University Press.
- Halliday, M. A. (1994). *Functional grammar*. London: Edward Arnold.
- Hammer, D. (1997). Discovery learning and discovery teaching. *Cognition and instruction*, 15(4), 485-529.
- Harwood, N. (2005). 'Nowhere has anyone attempted... In this article I aim to do just that': A corpus-based study of self-promotional I and we in academic writing across four disciplines. *Journal of pragmatics*, 37(8), 1207-1231.
- Hsu, H. L., Lai, S. L., & Chen, H. J. H. (2024). Examining in-service senior high school English teachers' perspectives on corpus use and the effects of corpus training. *Applied Corpus Linguistics*, 4(3), 100116.
- Huddleston, R. (1984). *Introduction to the Grammar of English*. Cambridge University Press.
- Hyland, K. (1996). Writing without conviction? Hedging in science research articles. *Applied Linguistics*, 17(4), 433-454.
- Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 learners' writing. *Journal of Second Language Writing*, 6(12), 183-205.
- Hyland, K. (1998). Boosting, hedging and the negotiation of academic knowledge. *TEXT*, 18(3), 349-382.
- Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning.. In T. Johns & P. King (Eds), *Classroom Concordancing English Language Research Journal* (pp. 1-16).
- Johns, T. (1997). Contexts: The background, development, and trialling of a concordance-based CALL program. In *Teaching and language corpora* (pp. 100-115). London: Longman.
- Johns, T., & King, P. (Eds.) (1991). *Classroom concordancing*. ELR Journal 4. University of Birmingham.
- Krashen, S. 1982. *Principles and Practice in Second Language Acquisition*. Oxford: OUP.
- Lee, C. Y., & Liou, H. C. (2003). A study of using web concordancing for English vocabulary learning in a Taiwanese high school context. *English Teaching & Learning*, 27(3), 35-56.
- Leńko-Szymańska, A. (2017). Training teachers in data-driven learning: Tackling the challenge. *Language Learning & Technology*, 21(3), 217-241. Retrieved from <http://llt.msu.edu/issues/october2017/lenko-szymanska.pdf>
- Leńko-Szymańska, A. (2022). Training teachers and learner to use corpora. In *The Routledge handbook of corpora and English language teaching and learning* (pp. 509-524). Routledge.
- Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh: Edinburgh University Press.

- Lusta, A., Demirel, Ö., & Mohammadzadeh, B. (2023). Language corpus and data-driven learning (DDL) in language classrooms: A systematic review. *Heliyon*, 9(12).
- Mair, C. (2002). Empowering non-native speakers: the hidden surplus value of corpora in continental English departments. In *Teaching and learning by doing corpus analysis* (pp. 119-130). Brill.
- Martin, I. P. (2014). English language teaching in the Philippines. *World Englishes*, 33(4), 472-485.
- Martínez, I. A. (2005). Native and non-native writers' use of first-person pronouns in the different sections of biology research articles in English. *Journal of second language writing*, 14(3), 174-190.
- McEnery, T., & Gabrielatos (2006). English corpus linguistics. In B. Aarts & A. McMahon (Eds). *The handbook of English linguistics* (pp. 34-71). USA: Blackwell Publishing.
- McEnery, T., & Wilson, A. (2001). *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- Milizia, D. (2013). Phrasal verbs and phrasal units: Political corpora within the walls of the classroom. In C. Desoutter, D. Helter, & M. Sala (Eds.) *Corpora in specialized communication* (pp. 135-164).
- Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art—and beyond. *Corpus technology and language pedagogy: New resources, new tools, new methods*, 5-24.
- Mukundan, J., & Khojasteh, L. (2011). Modal auxiliary verbs in prescribed Malaysian English textbooks. *English language teaching*, 4(1), 79-89.
- Nartey, M., & Yankson, F. E. (2014). A semantic investigation into the use of modal auxiliary verbs in the manifesto of a Ghanaian political party. *International Journal of Humanities and Social Science*, 4(3), 21-30.
- Orta, I. V. (2010). A contrastive analysis of the use of modal verbs in the expression of epistemic stance in Business Management research articles in English and Spanish. *Ibérica, Revista de la Asociación Europea de Lenguas para Fines Específicos*, (19), 77-95.
- O'Keeffe, A., & Farr, F. (2003) Using language corpora in initial teacher education pedagogic issues and practical applications. *TESOL Quarterly*, 37 (3). 389-418.
- O'Sullivan, I. (2007). Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL*, 19 (3), 269-286.
- Palmer, F. (1979). *Modality and the English modals*. London: Longman.
- Quirk, R., Svartvik, J., Leech, G., & Greenbaum, S. (1985). *A comprehensive grammar of the English language*. New York: Longman.
- Romer, U. (2006). Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future development. *ZAA* 54 (4), 121-134.
- Satake, Y. (2024). The effects of teacher, peer, and self-feedback on error correction with corpus use. *Applied Corpus Linguistics*, 4(3), 100114.
- Sinclair, J. M. (2008). New evidence, new priorities, new attitudes. In *How to use corpora in language teaching* (pp. 271-299). John Benjamins Publishing Company.
- Schmidt, R. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics*, 11, 129-158.
- Stubbs, M. (1996). *Text and corpus analysis: Computer-assisted studies of language and culture* (p. 158). Oxford: Blackwell.
- Sun, Y. C., & Wang, L. Y. (2003). Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning*, 16(1), 83-94.
- Tian, S. (2005). Data-driven learning: Do learning tasks and proficiency make a difference? In *Proceedings of the 9th Conference of the Pan-Pacific Association of Applied Linguistics*. Tokyo: Waseda University Media Mix Corp (pp. 360-71).

- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: Benjamins.
- Tribble, C. (2015). Teaching and language corpora: Perspectives from a personal journey. In *Multiple affordances of language corpora for data-driven learning* (pp. 37-62). John Benjamins Publishing Company.
- Turnbull, J., & Burtston, J. (1998). Towards independent concordance work for learners: Lessons from a case study. *ON-CALL*, 12(2), 10-21.
- Vannestål, M. E., & Lindquist, H. (2007). Learning English grammar with a corpus: Experimenting with concordancing in a university grammar course. *ReCALL*, 19(3), 329-350.
- Yeh, Y., Liou, H., & Li, Y. (2007). Online synonym materials and concordancing for EFL college writing. *Computer Assisted Language Learning*, 20 (2). 131-152.
- Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 Academic Writing
- Yoon, H. & Hirvela, A. (2004). ESL learner attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13 (4), 257-283.
- Zareva, A. (2016). Incorporating corpus literacy skills into TESOL teacher training. *ELT Journal*, 71(1), 69-79. doi:10.1093/elt/ccw045
- Zhong, X., & Wakat, G. (2023). Enhancing grammar proficiency of EFL learner through corpus-integrated lessons. *Ampersand*, 11, 100139. <https://doi.org/10.1016/j.amper.2023.100139>