

# The Impact of the Data-Driven Learning Approach on ESL Writers' Citation Patterns

Ebtisam Saleh Aluthman<sup>1</sup>

<sup>1</sup> Department of Applied Linguistics, College of Languages, Princess Nourah bint Abul Rahman University, Riyadh, Saudi Arabia

Correspondence: Ebtisam Saleh Aluthman, Department of Applied Linguistics, College of Languages, Princess Nourah bint Abul Rahman University, Riyadh, Saudi Arabia. E-mail: esaluthman@pnu.edu.sa

Received: July 6, 2017 Accepted: August 7, 2017 Online Published: September 25, 2017

doi:10.5539/ijel.v7n6p109 URL: <http://doi.org/10.5539/ijel.v7n6p109>

## Abstract

This study reports the impact of the data-driven learning (DDL) approach on ESL Saudi writers' general citation patterns that contribute to their general authorial voice. Specifically, the study examines the effects of the DDL activities on ESL writers' use of integral and non-integral citation patterns based on Swales' (1981, 1986, and 1990) modal of citation analysis and the extended scheme of classification set by Thompson & Tribble (2001). Guided use of both the Michigan Corpus of Upper-Level Student Papers (MICUSP) and WordandPhrase.info has been designed, implemented, and assessed with a representative sample of 32 ESL upper-intermediate and advanced writers in the Department of Translation in College of Languages at Princess Nourah bint Abul Rahman University (PNU). The effectiveness of the DDL activities in improving the writers' use of the citation patterns in composition of assignments is measured via a repeated measure paired *t* test. The study evaluates writers' authorial voice in terms of their use of integral and non-integral citation patterns. The quantitative analysis reveals that participants' integral patterns ( $n = 398$ ) of citation significantly outnumbered non-integral patterns ( $n = 126$ ). The verb-controlling pattern occurred the most ( $n = 320$ ), constituting 61% of total citation patterns. Results of the paired sample *t* test reveals a significant statistical difference between participants' performances before and after the integration of the DDL activities, with the mean value being increased from 2.285 to 3.778. These results inform pedagogical implications of the DDL approach in ESL writing. The conceptual framework implementing the DDL approach in the present study provides guidance for applying corpus-informed tools when designing writing activities for upper-intermediate to advanced ESL learners.

**Keywords:** data-driven learning, authorial voice, citation patterns, academic writing

## 1. Preliminaries

The authorial voice in academic writing has received growing attention among writers and researchers over the past two decades (Swales, 1981, 1986, 1990; Thompson, 2000, 2001; Hyland, 1999, 1999, 2000, 2002, 2002, 2005). The premise underlying this interest is the fact that the voice of authorship should be carefully integrated within the academic text to establish the text's credibility and authenticity in a disciplinary field. The authorial voice is a dynamic projection of the writer's own voice within the academic text (Bitchener, 2010), contributing to the disciplinary field of epistemology (Basturkmen, 2012), communicating with a particular readership (Hyland 1999a), foregrounding the contribution of the text (Swales, 1990), and seeking acceptance and inclusion within a particular academic discourse community (Warchal, 2010). The authorial voice referees the author-reader dialogue (Hyland, 2005), where the writer's authoritativeness is revealed through assessing, claiming, evaluating, and claiming gaps in related literature (Navratilova, 2013).

The construction of the authorial voice is achieved using a variety of linguistic and rhetorical resources, among which citation patterns receive great emphasis (Thompson & Tribble, 2001). Citation patterns are discursive resources implemented in the construction of the authorial voice for the purpose of acknowledging and evaluating related knowledge as well as identifying gaps in existing research (Petric, 2007). Within an academic research context, Mansourizadeh & Ahmad (2011) maintain that citation is a resource used by writers to locate their research within the related context, claim relevance of their inquiry, confirm their competency in the field, postulate the significance and legitimacy of their investigation, and depict the relevance of their contribution. From a linguistic point of analysis, Swales (1981, 1986, & 1990) has pioneered the citation research. He classified the

citation patterns in terms of their syntactic function within the inter-sentence level into integral and non-integral. Integral citation patterns are those playing a grammatical function within the sentence in which they occur. Non-integral citation patterns, on the other hand, are those that play no grammatical function within the level of sentence.

Based on Swales' classification of citation patterns, Thompson & Tribble (2001) analyze disciplinary variations of citation in the corpus of doctoral dissertations in two disciplines in the Department of Agriculture: Agricultural Botany and Agricultural Economics. The authors developed a more detailed scheme of classification by identifying citation patterns in terms of integration or non-integration within sentences as well as by the function achieved by these patterns. Table 1 is a description of this scheme (see Thompson & Tribble, 2001 for examples of these citation examples from their corpus).

Table 1. Citation scheme

	Categorization	Description
Integral Citation	Verb- controlling	The citation function is performed via a controlling active or passive verb.
	Naming	(a) The citation constitutes a noun phrase. (b) The citation constitutes a part of a noun phrase.
	Non-Citation	The citation refers to the name of the writer without the year. This usually happens when the reference author's name is mentioned and subsequently followed by a citation to the same reference author.
Non-integral Citation	Source	The citation function is performed by attributing an assertion, fact, or results to a reference name.
	Identification	The citation is performed by identifying and specifying an agent in the sentence it references in an information-prominent citation pattern rather than an author-prominent citation pattern (Weissberg & Buker, 1990, cited in Thompson & Tribble, 2001).
	Reference	The citation is indicated by the insertion of the directive verb. For example (see).
	Origin	The citation pattern specifies the originator of a theory, product, or concept.

Source: Thompson & Tribble (2001).

Implementing tools of corpus analysis facilitated the investigation of citation patterns from a variety of perspectives during the past two decades. As an essential resource in the construction of disciplinary knowledge, sub-disciplinary variations of citation patterns have been repeatedly emphasized (Hayland, 1999). The underlying assumption is that writers in different disciplines employ different rhetorical practices to construct particular disciplinary knowledge. Hyland (1999) and Thompson (2000) investigated sub-disciplinary variations in citations in research articles and doctoral theses conducted in different disciplines with tools of frequency and concordance analysis. Their analysis demonstrates the novice writers' tendency to use limited citation patterns; thus, they should be guided to use a variety of citation patterns based on genre awareness. Based on Thompson & Tribble's (2001) model of citations, Manan (2015) examined integral citation patterns in postgraduate students' master's theses at the National University of Malaysia. Among the integral citation patterns, verb-controlling patterns have been found to have the highest rate of occurrence in Manan's data ( $n = 198$ ), followed by naming patterns ( $n = 48$ ). Manan (2015) emphasizes ESL writers' barriers of language proficiency being a hindrance to both academic writing and employing citation patterns appropriately. In line with Thompson & Tribble (2001), Manan (2015) insists on the integration of training courses within academic writing courses, focusing on the proper implementation of citation patterns, with the writers being exposed to authentic texts to raise cognitive awareness of different citation patterns.

Variations across local and international writers have been also stressed in a variety of studies. Karimi & Asadnia (2014), for example, found that local Iranian writers employed more integral citation patterns than their international counterparts. Rabab'ah & Al-Marshadi (2013) investigated citation patterns among native and non-native postgraduate students based on Sawles' (1990) framework of citation patterns. They analyzed five masters' theses written by Arab writers and five masters' theses written by native English writers. The results revealed the Arab writers' low proficiency in citing appropriately. McCallum (2016) investigated the different functions of the first-person pronouns in shaping the authorial voice in a small, specialized corpus of 45 female Saudi writers' papers. Among the most prominent functions were commenting, giving opinions, claiming, reporting an experience, and maintaining a desire.

The aforementioned studies of citation patterns during the last two decades have proven a number of issues and implications (e.g., non-native English writers' lack of adequate citation, and the necessity of familiarizing writers

with authentic texts to enhance their persuasive authorial voice in their academic writing). Citation patterns of undergraduate students and intermediate to advanced writers, however, have been marginalized in this area of research. At the same time, the DDL approach has demonstrated a significant impact in a variety of educational contexts (see, e.g., Flowerdew, 2012, 2015; Boulton, 2009, 2010; Gaskell & Cobb, 2004; Yoon, 2014). The aim of the present study is to investigate the impact of direct application of corpus use on ESL upper-intermediate to advanced undergraduate writers' authorial voice in an educational writing context. The current study reports on a pedagogically-oriented investigation of citation patterns among 32 ESL Saudi upper-intermediate and advanced writers in terms of the impact of the DDL approach on their usage and selection of citation patterns. The investigation is guided by the hypothesis that the DDL approach has a positive impact on participants' overall improvement of authorial voice. The improvement is also explored in terms of the selection of integral versus non-integral citation patterns.

## 2. Review of Literature

### 2.1 Data-Driven Learning Approach

A few decades after the establishment of corpora as an identifiable field in linguistic study in 1950s, the use of electronic readable format for different kinds of corpora have been integrated into most, if not all, linguistic disciplines. In parallel to this rapid development of the use of corpora were the growing claims of the authentic materials movement in the language teaching field in the 1980s, which advocated the use of real-world authentic materials in language classrooms. The "naturally occurring texts" of the language corpus were rapidly recognized as great material for this movement. Theories of second language acquisition, such as Vygotskian sociocultural theory (Vygotsky, 1986) and the "noticing" hypothesis (Schmidt, 1991), have underpinnings, both explicitly and implicitly, in the use of authentic corpus material in language classrooms.

The pedagogical applications of the use of corpus have been categorized into direct and indirect applications (Römer, 2010, as cited in Flowerdew, 2012). Indirect applications include those applications that have indirect effects on teaching and learning materials introduced by researchers and material designers. Direct applications, on the other hand, refer to those applications that involve both teacher-corpus interaction and learner-corpus interaction. In the last few years, this direct interaction with the corpus tools has been coined Data-Driven Learning (DDL) by Johns & Kim (1991, iii), referring to "the use in the classroom of computer generated concordances to get students to explore the regularities of patterning in the target language, and the development of activities and exercises based on concordance output." This call to implement the DDL approach in education has been the foundation of a new era of learning in terms of pedagogical methods, curriculum design, and assessment over the last two decades (see Champers, 2010 for a brief account of DDL history). A large body of scholarly research has been conducted to increase the awareness of teachers, graduate students of applied linguistics, and trainers about the practical applications of corpus in language classrooms. An account of DDL advantages that are well documented in the DDL literature and an overview of the empirical studies that implemented the DDL approach in skill-based activities (i.e., writing tasks) follow.

The most prominent advantage stated by DDL scholars is the fact that DDL involves both a discovery-based deductive approach and inductive approach. Johns' (1991) description of DDL approach expands to include the implementation of authentic, naturally occurring data in exploratory activities where students are observing, exploring, inferring patterns, and generalizing using an inductive approach. Johansson (2009), however, maintains that the implementation of the DDL approach implies an emphasis on deducing insights; thus, it is a guided combination of both inductive and deductive approaches where the authentic, naturally occurring forms of information are tailored for students' particular needs. Within the context of this guided DDL approach, students are viewed as researchers (Bernardini, 2004) who are conducting multiple tasks of observation, analysis, and reasoning. In such a context of learning, Krieger (2003) argues, the instructor has the role of a research mentor rather than a presenter of knowledge.

The second advantage of the DDL approach is the facilitation of learner's involvement in a lexical grammatical approach to learning. Compared to traditional approaches that separate grammar from lexicology, the DDL approach provides exploration of lexical items and patterns within grammatical structures via concordance tools (Flowerdew, 2015). Third, the DDL approach has been shown by empirical studies to be influential in fostering learners' autonomy and enhancing their awareness and observation skills (Luo, 2016). This influence is largely due to the exploratory-based activities that are offered by the DDL activities. These activities serve as cognitive learning tools to enhance learners' observation skills (O'Sullivan, 2007; Sun, 2003).

### 2.2 Data-Driven Learning Approach and Writing Tasks

Writing is dominating the research on the effect of the DDL approach on skill-based activities (see Luo, 2016). The

effect is tackled from a variety of perspectives including errors correction, enhancement of fluency and complexity, and particular tasks of writing. The majority of the empirical studies that have been conducted on issues of DDL and writing tasks agree that the DDL approach has positive effects on the overall improvement of writing. Following is a summary of the most prominent studies in the field, with references to the key issues in this area of investigation.

Error correction is the most prominent issue tackled when investigating the effect of the DDL approach on writing. Gaskell & Cobb (2004) used a pretest and posttest design study to investigate the effect of concordance tools on eliminating errors at the level of sentences with low intermediate learners. In an exploratory study, Luo (2015) reported on a pretest and posttest experimental study where he investigated the effect of utilizing a DDL approach on forty eight second language learners (L2) Chinese learners' writing in terms of three variables: accuracy, complexity, and fluency. The experimental group, who consulted the British National Corpus (BNC) and Baidu tools in their writing, performed significantly better than the control group in terms of writing accuracy and fluency, but not as to complexity. Luo & Lia (2015) conducted a small exploratory study that investigated the effect of Beijing Foreign Studies University (BFSU CQP) web as a reference tool for eliminating writing errors of ESL intermediate Chinese learners. Results comparing the experimental and control groups revealed a positive effect of using an online corpus to minimize lexical and grammatical errors compared to the use of online dictionaries. Tono et al. (2014, as cited in Luo, 2015), however, indicated the significance of the DDL approach for identifying and correcting omission and addition errors, whereas misinformation errors have been difficult to identify and correct. Yoon & Jo (2014) conducted a case study to investigate the effects of indirect and direct implementation of DDL approaches on error correction. They reported in their results that learners' self-correction was significantly higher when applying indirect corpus-based activities than when using direct activities with low-proficiency learners. This finding agrees with that of Gaskell & Cobb (2004), who highlighted that an indirect DDL approach is best considered as a transitional stage to a direct DDL approach.

The positive effects of the DDL approach on writing have been also observed at the level of writing strategies. Gilmore (2009) investigated the usefulness of two online corpus resources—the BNC and the Collins Birmingham University International Language Database (COBUILD)—for improving L2 Japanese learners' writing strategies. He observed a significant effect on learners' drafting strategies after a 90-minute practical tutorial on implementing these online resources into writing tasks. Kennedy & Miceli (2010) found a positive effect of the DDL approach on writing, planning strategies, and idea generation with Italian L2 intermediate learners when integrating corpus applications in writing tasks. Boulton (2009) claimed the usefulness of the DDL approach in learning and mastering linking adverbs. Emphasizing the importance of choosing the appropriate type of corpus for a particular task, Chang (2014) investigated the roles of both general and specific corpora in improving L2 writing. He conducted a case study to evaluate L2 Korean engineering learners' consultation of both kinds of corpus; results showed a significant effect of using specific corpora as consultation tools in learners' academic writing.

The aforementioned studies on the implementation of the DDL approach in writing tasks have proven the benefits of direct and indirect corpus-based activities. They provide learners with linguistic cognitive consultation tools that improve L2 learners' overall writing accuracy. This improvement results from four different factors documented in previous studies: error-correction elimination, refining learners' use of grammatical structures, developing learners' strategies for drafting, planning, and generating ideas, and improving noticing and cognitive skills. The investigation of the DDL use on L2 writers' rhetorical and stylistic development, however, has received little attention in prior investigations.

The aim of the present study is to investigate the impact of the implementation of the DDL approach on the improvement of learners' authorial voices in terms of citation patterns. The investigation is tangibly identified in terms of designing appropriate corpus-based activities as well as testing the effect of the DDL implementation on learners' use of citation patterns. Beyond the mere identification and testing of the DDL approach's effects on learners' overall improvement in their use of citation patterns, another aim of the study is to examine the learners' selection of integral and non-integral citation patterns before and after the implementation of the DDL approach. From a methodological point of view, many of the experimental studies conducted in this area of investigation followed the pretest and posttest design. Due to the aims of the present study, and to decrease the effect of other influences on the results, the present study utilized a repeated measure design where 32 participants were given the same amount of DDL instruction. The participants' performance in terms of their mastery of citation patterns, as well as their selection of integral and non-integral citation patterns were analyzed before and after the implementation of the DDL approach in their writing instructions.

### 3. Research Questions

In order to achieve the aims of this study, the following research questions have been proposed.

RQ1: What are the general patterns of citation employed by upper-intermediate to advanced Saudi ESL writers?

RQ2: Is there a positive significant difference in the Saudi ESL writers' use of citation patterns before and after implementing the DDL approach in their writing tasks?

RQ3: Is there a significant difference in the Saudi ESL learners' selections of integral versus non-integral citation before and after the implementing the DDL approach in their writing tasks?

### 4. Research Design

The effective implementation of the DDL approach in previous experimental studies was guaranteed by ensuring a number of key issues including the choice of the appropriate corpus best suited for the task, training participants as well as participants' familiarity with corpus tools. The research design of the current study has been determined in accordance with these key issues. Based on the aims of the present study and the desire to decrease the effect of other influences on the results, the present study utilized a repeated measure design where 32 participants were given the same amount of the DDL instruction during the a training session , and were asked to write literature reviews before and after the DDL training. The participants' performance (i.e., mastery of citation patterns and selection of integral and non-integral citation patterns) were measured before and after the DDL training. The participants' pre- and post-DDL literature reviews were analyzed in terms of both citation proficiency and citation patterns. The paired sample *t* test measure was applied to participants' submissions before and after the DDL use. Based on Thompson & Tribble's (2001) framework of citation patterns, the participants' citation patterns were identified, analyzed, and tabulated to ascertain general patterns of citation. Quantitative results were obtained and pedagogical implications were drawn.

The participants were all enrolled in the Department of Translation in the College of Languages at PNU, Saudi Arabia. All participants were Level Three female upper-intermediate to advanced learners who had passed two courses in writing: Writing I and Writing II. Writing I introduced learners to essay writing, and Writing II prepared students to use different types of academic writing including argumentative essay, cause-and-effect essay, and research essay. The participants were all familiar with how to cite previous studies in their reports, how to quote sources, and how to report on related literature. In addition, they were expected to be enrolled in the Writing III course, which requires the presentation of a full research paper, including a literature review section. This requirement helped motivate the participants to be actively involved in all DDL activities conducted in the present study.

### 5. Instrumentation

#### 5.1 Data-driven Learning Methods

The selection of the corpus resources has a great effect on the usefulness of the DDL approach. Some researchers rely on a general corpus that offers massive numbers of authentic examples, such as British National Corpus BNC (e.g., Yoon, 2008; Gilmore, 2009; Luo, 2016). Other researchers integrate a specialized corpus into DDL activities (e.g., Kennedy & Mceli, 2010). The purpose of the current research has necessitated the use of two different corpus resources.

The first source is the Michigan Corpus of Upper-level Student Papers (MICUSP), available at <http://micusp.elicorpora.info/>, which provides participants with authentic examples of how integral and non-integral citation forms are crafted by highly proficient upper-intermediate counterparts in a variety of disciplines (see Römer, 2010, for a detailed description of MICUSP). Figure 1 shows an illustration of the MICUSP home interface.

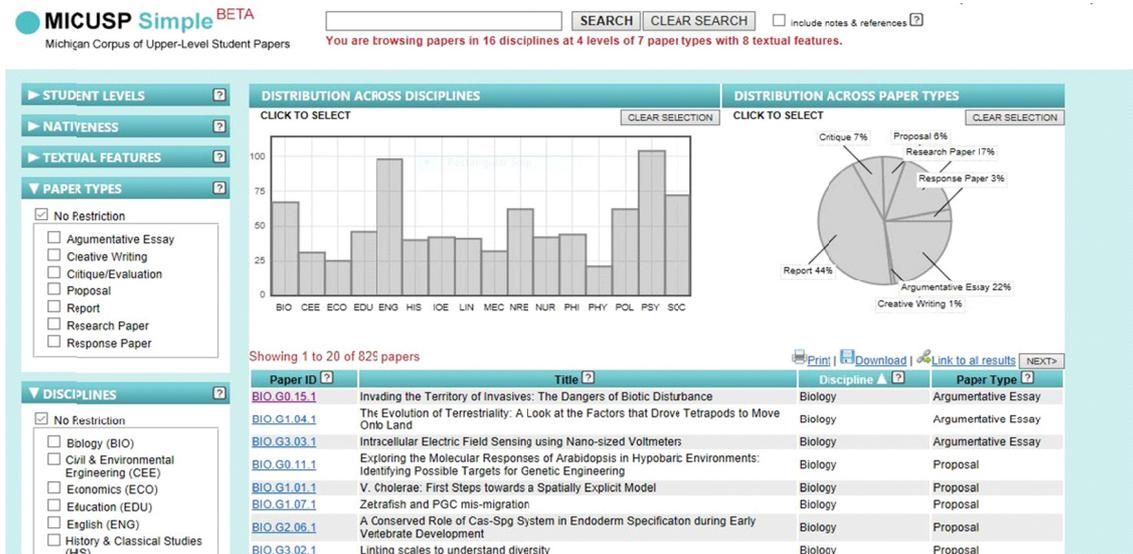


Figure 1. Screenshot of MICUSP simple BETA version interface

MICUSP is compiled of 829 papers (i.e., 2.6 million words) in a user-friendly, easy-to-use interactive design. All MICUSP files are saved in TEI-compliant XML format, and display the main divisions of the papers as well as metadata such as the writer’s gender, discipline, first language, academic level, and paper type. Römer (2010) expounded on how MICUSP can be integrated pedagogically within academic writing activities to assist writers in making good decisions about their writing. The basic pedagogical purposes of MICUSP that are functionally used in the current study are: (a) surfing papers, and (b) investigating the use of words and phrases. Figure 2 shows a simple search on the phrase “according to,” and shows that it occurs 835 times in 365 papers representing 16 disciplines at 4 levels, including 7 paper types with 8 textual features. The distributions of “according to” across paper types and disciplines are shown using charts. The search also shows the phrase in an authentic context with all the metadata illustrated.

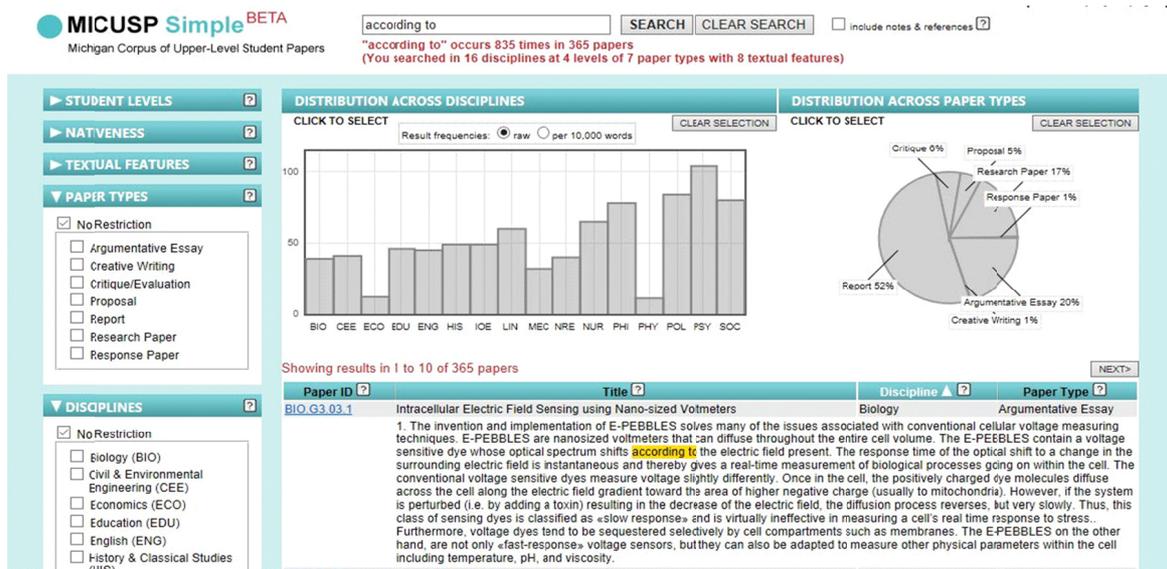


Figure 2. Screenshot of MICUSP interface: The phrase “according to” in context

The second corpus resource is the Word and Phrase database at <http://www.wordandphrase.info/academic/frequencyList.asp>, which is part of the Corpus of Contemporary American English (COCA). It consists of 120 million words from academic English, in different disciplines and

genres, within the time span 1990-2010. Searching this large corpus provides information about word frequency, n-grams, and collocation. Participants are advised to survey the “Academic” section, not “All Genres,” which includes both academic and non-academic corpus. Figure 3 illustrates a search of the Word and Phrase database using the word “argues.” Information about the distribution of the word under investigation across disciplines (ratio per million), the word in context, as well as suggested synonyms are all summarized.

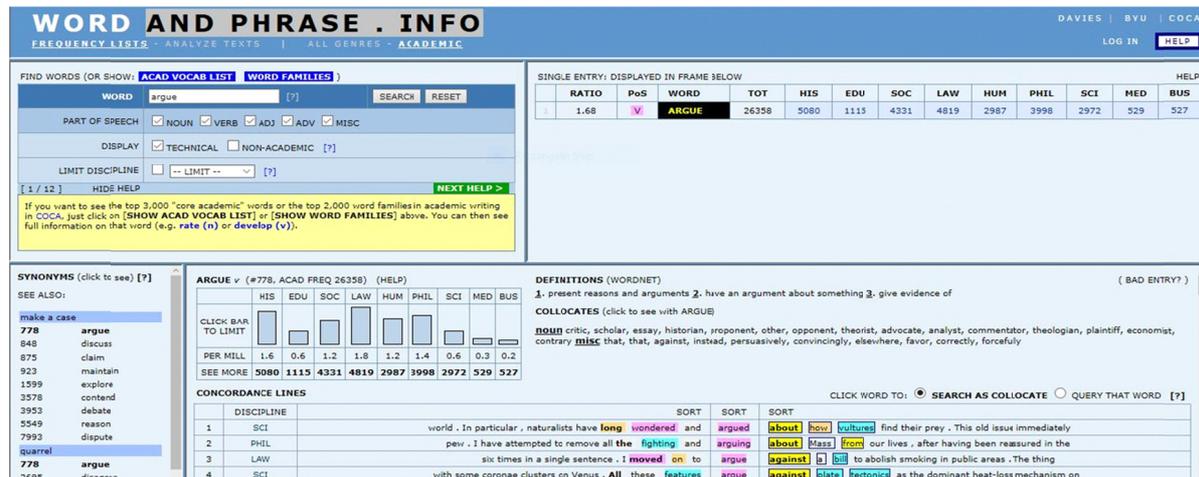


Figure 3. Screenshot of the Wordandphrase.info main interface using the word “argue” as a search term

## 5.2 Procedures and Administration

The main procedures of data collection are summarized as follows:

- Participants were asked to write a literature review section (about 1000 words long) prior to the implementation of the DDL methods. They were introduced to techniques and tips on how to write a literature review section in a research article. The topics assigned to this task included ones the participants were familiar with throughout their studies, such as “Language and Media,” “Animal Communication,” “Saudi Varieties,” “Orthography,” and “Language and Brain.”
- The participants were given indirect and direct DDL activities throughout their writing activities over three weeks. Initially, a 60-minute training session was conducted during which the participants were familiarized with consultation skills and the tools of corpus-searching activities prior to the participants’ independent consultations of both corpus resources. The significance of a training session in empirical DDL studies has been repeatedly emphasized (Gilmore, 2009; Smart, 2014; O’Sullivan & Chambers, 2006). During the training session, the participants learn how to limit their search, conduct frequency analysis, and explore concordance features.
- The participants were all asked to compose another literature review on another topic from the suggested list, and the same evaluators using the same rubric described below evaluated their second submission.
- A rubric was used by two different evaluators to assess the participants’ literature reviews in terms of their use of citation patterns. The rubric was designed after extensive reading and consultation with instructors of academic writing and researchers. The main criteria of the evaluation were:
  - Appropriateness of the citation patterns.
  - Integration within the context of the literature review section.
  - Citing information properly.
  - Coherence.
  - Cohesion
  - Grammar.

## 6. Results

In order to answer the three research questions posed in the present study, both evaluations of the participants’ submissions were analyzed in terms of (a) the participants’ overall use of citation patterns, and (2) the participants’

selection of integral and non-integral citation patterns. The results related to the three research questions are summarized in this section.

### 6.1 The Participants' General Citation Patterns

RQ1 in this study is concerned with tracing the general patterns of citation employed by upper-intermediate to advanced Saudi ESL writers. Thompson & Tribble's (2001) framework is used in the process of identifying the citation patterns employed by the participants in all the literature reviews submitted (both prior to and after the implementation of DDL activities). Citation patterns are categorized following Thompson & Tribble's (2001) framework. The results are tabulated in Table 2 and then illustrated in Figure 4.

Table 2. The participants' general citation patterns

	Category	N	%
Integral citation	Verb-controlling	320	61%
	Naming	78	15%
	Non-citation	0	0%
	Total	398	76%
Non-integral citation	Source	96	18%
	Identification	25	5%
	Reference	5	.01%
	Origin	0	0%
	Total	126	24%
Total		524	

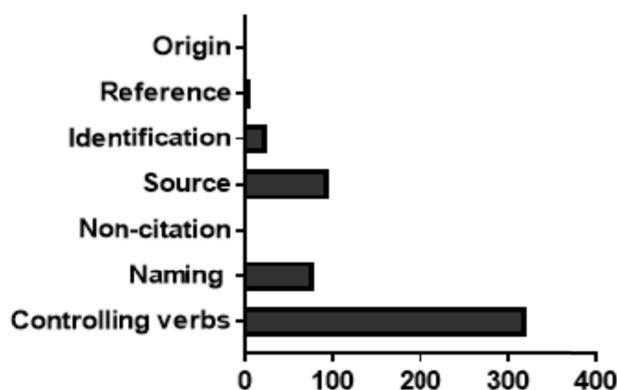


Figure 4. The participants' general citation patterns

Table 3. The frequency of citation patterns in the first and second submissions

	Category	No. in 1st Submissions	No. in 2nd Submissions
Integral citation	Verb-controlling	164	156
	Naming	51	27
	Non-citation	0	0
	Total	215	183
Non-integral citation	Source	24	72
	Identification	3	22
	Reference	0	5
	Origin	0	0
	Total	27	99
Total submissions by group		242	282
Total combined submissions		524	

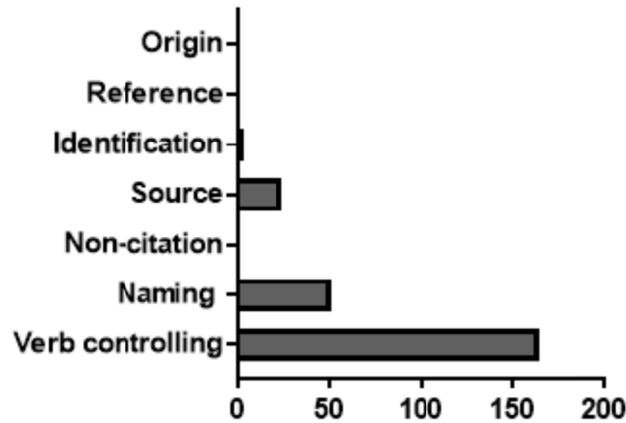


Figure 5. The participants’ citation patterns in the first set of submissions

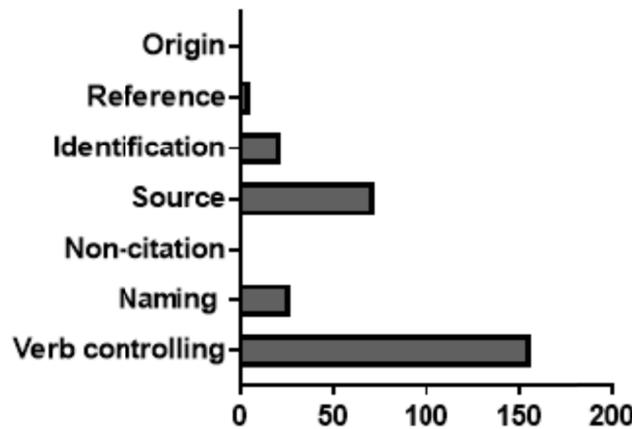


Figure 6. The participants’ citation patterns in the second set of submissions

6.2 The Paired Samples *t* Test: Results

RQ2 is concerned with finding out whether there is a significant difference in the Saudi ESL writers’ use of citation patterns before and after implementing the DDL approach in their writing tasks. Participants’ first and second submissions have been evaluated and scored. A paired samples *t* test has been conducted on the participants’ scores for both submissions. Table 4 and Table 5 show the descriptive statistics for the participants’ first and second submissions.

Descriptive analysis of the participants’ scores on the first submission reports a mean of 13.4 (SD = 2.38). The median is 14.0 (Range = 20.0-9.0). Table 5 demonstrates a significant difference in the second submission, with a mean of 17.0 (SD = 2.21). The median is 18.0 (Range = 20.0-12.0). This difference indicates an evident improvement in the participants’ general use of citation patterns between the first and second submissions.

Table 4. Results of the participants’ scores on first submissions

Measurement	Value
<i>N</i>	32
<i>M</i>	13.4
<i>SD</i>	2.38
<b>Median</b>	14.0 (High = 20.0; Low = 9)
95% CI for Mean: [12.52, 14.23]	
Average Absolute Deviation from Median = 1.75	

Note. CI = confidence interval.

Table 5. Results of the participants' scores on second submissions

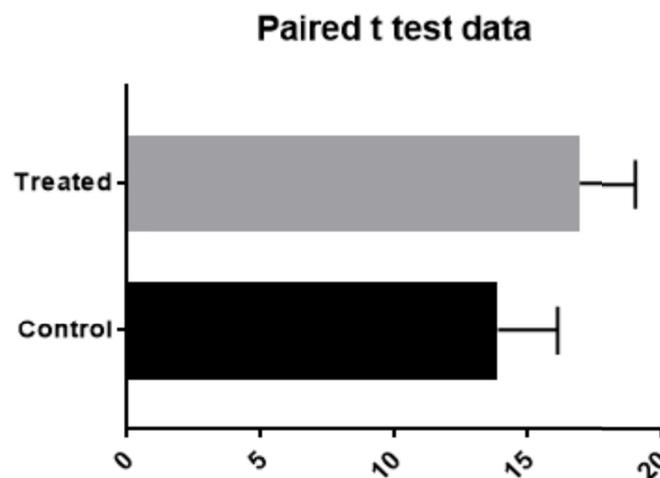
<b>No. of items</b>	32
<b><i>M</i></b>	17.0
<b><i>SD</i></b>	2.21
<b>Median</b>	18.0
Hi = 20.0 Low = 12.0	
95% CI for Mean: [16.17, 17.76]	
Average Absolute Deviation from Median = 1.78	

Note. CI = confidence interval.

Table 6 shows the results of the paired sample *t* test of the participants' use of citation patterns in both submissions. As illustrated, the results of the *t* test indicates a significant statistical difference between the two submissions (i.e., before and after the integration of DDL activities) with  $p < 0.0001$  and degrees of freedom = 31. Analysis shows that the mean of differences is 3.031 ( $SD = 2.071$ ) and the mean value increases from 2.285 to 3.778. However, the improvement in the participants' use of citation patterns was not only due to an increase in the quantification of patterns (as illustrated in Table 3), but also due to the selection of the citation patterns. An analysis of the participants' selection of the citation patterns is summarized in the following section.

Table 6. Results of the paired sample *t* test

Measurement	Value
<i>p</i> value	<0.0001
Significantly different ( $p < 0.05$ )?	Yes
One- or two-tailed <i>p</i> value?	Two-tailed
<i>t</i> , <i>df</i>	$t = 8.279$ , $df = 31$
No. of pairs	32
<i>M</i> of differences	3.031
<i>SD</i> of differences	2.071
<i>SEM</i> of differences	0.3661
95% CI	2.285 to 3.778
$R^2$ (partial eta squared)	0.6886
<i>r</i> (correlation coefficient)	0.5368
<i>p</i> value (one-tailed)	0.0008

Figure 7. Results of the paired sample *t* test

### 6.3 The Participants' Selection of Integral versus Non-integral Citation Patterns

In order to trace the improvement in the participants' performance, it is necessary to identify any significant differences in their selections of integral versus non-integral citation before and after implementing the DDL approach.

Table 3 shows a noticeable difference in the total quantification of citation patterns in the participants' first and second submissions (242 to 282 patterns, respectively). The frequency of the integral citation patterns decreased from 215 to 183, with a moderate decrease in both the verb-controlling pattern (from 164 to 156) and a significant decrease in occurrences of the naming patterns (from 51 to 27). Related to the non-integral citation patterns, an identifiable increase has been traced in the total number of occurrences (from 27 to 99), indicating an adequate mastery of non-integral citation patterns and an awareness of these patterns in constituting the authorial voice in academic writing. This result is clear from the increases in the numbers of the source patterns (24 to 72) as well the identification pattern (3 to 22).

The results summarized above are in line with results obtained by Rabab'ah & Al-Marshadi (2013) when they found that citation patterns used by non-native writers ( $n = 857$ ) exceeded patterns employed by native writers ( $n = 525$ ). This finding was justified by the fact that non-native writers were lacking some of the strategic competence and linguistic repertoire that are necessary for constituting a persuasive authorial voice (i.e., rephrasing their own ideas for the sake of credibility and projecting an evaluative critical voice). The DDL activities, however, raise the non-native participants' cognitive and metacognitive awareness of the variety of linguistic, rhetorical, and discursive tools that constitute a persuasive authentic authorial voice by exposing participants to authentic texts and reliable corpus consultations.

## 7. Conclusion

This research is a response to both the lack of studies investigating the authorial citation patterns employed by ESL Saudi undergraduate writers, and a wider lack of research on the impact of a corpus-based DDL approach in enhancing these citation patterns. The purpose of the present study is threefold. First, it provides an empirical investigation of the impact of utilizing two well-known corpora—the MICUSP and WordandPhrase.info—in a writing task using a guided-exploratory approach. Details concerning the implementation of these corpus tools, designing activities, and developing assessment tools are based on a careful reading of the existing literature. Second, this study investigates the impact of using the DDL approach on the improvement of ESL writers' general citation patterns, thereby enhancing their authorial voices. The investigation is conducted via a repeated measure of a paired sample  $t$  test. Third, the study investigates the impact of the utilization of a corpus-based DDL approach on the ESL writers' integral and non-integral citation patterns by careful quantification of all categories of citation patterns in both submissions. The investigation yielded significant results demonstrating the well-established assumption about the effectiveness of guided instruction using corpus-based activities in improving ESL writers' general proficiency.

Regarding general citation patterns, the participants showed significant dependability on the verb-controlling pattern in constructing their authorial identities ( $n = 320$ ). The dominance of this pattern is also reflected by related exploratory studies in the field. Another prominent finding of the present study was the participants' preference for integral citation patterns ( $n = 398$ ) over non-integral citation patterns ( $n = 126$ ). This preference, however, is not in line with the native writers' general preferences as documented in the literature (Mansourizadeh & Ahmad, 2011; Rabab'ah & Al-Marshadi, 2013). According to Thompson & Tribble (2001), writers tend to deemphasize the visibility of researchers in constructing their authorial identity in academic writing. This explains why native writers, who are regarded as more proficient writers than their non-native counterparts, tend to prefer employing non-integral patterns rather than integral citation patterns. This also explains the dominance of the verb-controlling patterns in non-native writers' authorial voices.

The DDL approach is found to enhance the participants' authorial voices. Results of the paired sample  $t$  test shows that the mean value increases from 2.285 to 3.778. This significant statistical result is explained by a number of factors. First, the participants' reduction of the occurrences of the integral patterns ( $n = 215$  to  $n = 183$ ) and significant increase of the occurrence of the non-integral citation patterns ( $n = 27$  to  $n = 99$ ). In particular, a significant increase has been observed in the occurrence of the source patterns (from 24 to 72) as well the identification pattern (from 3 to 22). This indicates that the DDL activities, including exposure to authentic academic texts and corpus-based consultations, have a positive effect on the participants' awareness of how to construct their invisibility while citing in their academic writing. The DDL activities improve the participants' abilities in reshaping their authorial identities by relying on their rhetorical and discursive repertoires. The increase in both the source and identification patterns indicates the participants' increased abilities in constructing their invisible evaluative authorial voices in a native like way one.

## 8. Implications of the Study

The results of the present study suggest a number of significant pedagogical observations and implications relevant to the area of DDL and academic writing in an EFL context. More specifically, by reporting on the participants'

employment of citation patterns, this study highlights an area of investigation that has been never addressed in research even though the topic is closely related to academic writing.

The participants in the present study, though being upper-intermediate to advanced ESL learners, lack adequate proficiency in constructing their authorial identity. This is indicated by the descriptive results of their scores on the first submission, with a mean value of 13.4 and median value of 14.0. The participants' writing in the present study, however, manifests very proficient English and a sophisticated vocabulary. Thus, the inadequacy identified in their authorial voices in their first submission is due to lack of awareness of the different pragmatic functions of the citation patterns. Along the same line, other researchers related the ESL writers' inadequacy in citation to lack of both the necessary knowledge and exposure to the researching tools in authentic academic texts (Loi, 2012; Azlan, 2013). Authentic texts provide an opportunity for familiarizing non-native writers with the pragmatic and discursive functions of citation patterns (Thompson & Tribble, 2001). The DDL activities showed an observable effect in bridging this gap in the participants' communicative strategic competence. This manifests the significance of integrating the DDL approach into the academic writing curriculum.

The positive effects of the DDL approach documented in the present study is in line with the well documented research that calls for the integration of DDL in ESL context (e.g., Sun, 2003; Bernardini, 2004; O'Sullivan, 2007; Johansson, 2009; Flowerdew, 2012, 2015; Luo, 2016). The DDL activities implemented in the present study have enhanced the participants' exploratory deductive and inductive insights through activities of observing, exploring, and inferring patterns in authentic, naturally occurring language texts. The activities also improved the participants' autonomy and noticing strategies through different searching and consultation activities. Investigations should be conducted to assess and compare the effects of the available corpus-based material in ESL contexts for different academic writing purposes. Future research should be directed at cross-discipline studies investigating disciplinary variations in different aspects of academic writing, such as authorial identity and metadiscourse strategies. The research in this area should benefit from the availability of different academic corpora in different disciplines that are compiled on the basis of varying criteria. Results of such investigations will help in designing corpus-informed academic courses that will best serve particular purposes in different disciplines.

## References

- Azlan, N. (2013). Citation typologies and rhetorical functions of citations in master dissertations. *Journal of Creative Practices in Language Learning and Teaching (CPLT)*, 1(2), 61-74.
- Basturkmen, H. (2012). A genre-based investigation of discussion sections of research articles in dentistry and disciplinary variation. *Journal of English for Academic Purposes*, 11(2), 134-144. <https://doi.org/10.1016/j.jeap.2011.10.004>
- Bennett, G. (2013). *Using corpora in the language learning classrooms: Corpus linguistics for teachers*. Ann Arbor, MI: University of Michigan Press.
- Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 15-36). Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.12.05ber>
- Bitchener, J. (2010). *Writing an applied linguistics thesis or dissertation: A guide to presenting empirical research*. Basingstoke: Palgrave Macmillan.
- Boulton, A. (2009). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, 21(1), 37-54. <https://doi.org/10.1017/S0958344009000068>
- Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Lang Learn*, 60(3), 534-572. <https://doi.org/10.1111/j.1467-9922.2010.00566.x>
- Chambers, A. (2010). What is data-driven learning? In A. O'Keefe & McCathy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 345-358). London: Routledge.
- Chang, J. (2014). The use of general and specialized corpora as reference sources for academic English writing: A case study. *ReCALL*, 26(2), 243-259. <https://doi.org/10.1017/S0958344014000056>
- Flowerdew, L. (2012). *Corpora and language education*. New York: Palgrave Macmillan.
- Flowerdew, L. (2015). Data-driven learning and language learning theories. *Multi Afford Lang Corpora Data-Driven Learn*, 69, 15-36. <https://doi.org/10.1075/scl.69.02flo>
- Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, 32(3), 301-319.

<https://doi.org/10.1016/j.system.2004.04.001>

- Gilmore, A. (2009). Using online corpora to develop students' writing skills. *ELT Journal*, 63(4), 363-372. <https://doi.org/10.1093/elt/ccn056>
- Helali, M., & Kuhi, D. (2014). The use of citations in academic writing: Analysis of introduction sections of Iranian and native English master's theses. *Journal of Social Issues & Humanities*, 2(3), 216-220.
- Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics*, 20(3), 341-367. <https://doi.org/10.1093/applin/20.3.341>
- Hyland, K. (1999). Talking to students: Meta-discourse in introductory textbooks. *English For Specific Purposes*, 18(1), 2-26. [https://doi.org/10.1016/S0889-4906\(97\)00025-2](https://doi.org/10.1016/S0889-4906(97)00025-2)
- Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. Harlow, UK: Longman.
- Hyland, K. (2002). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics*, 34, 1091-1112. [https://doi.org/10.1016/S0378-2166\(02\)00035-8](https://doi.org/10.1016/S0378-2166(02)00035-8)
- Hyland, K. (2002). Options of identity in academic writing. *ELT Journal*, 56, 351-358. <https://doi.org/10.1093/elt/56.4.351>.
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-192. <https://doi.org/10.1177/1461445605050365>
- Johan, T., & King, P. (1991). Classroom concordancing. *English Language Research Journal*, 4, 27-45.
- Karimi, M., & Asadnia, F. (2014). Meta-discourse Strategies and Citation Patterns as Resources of Authorial Identity in Research Articles across International/Local Journals and Applied Linguistics Sub-disciplines. *The Journal of Teaching Language Skills*, 6(1), 49-72.
- Kennedy, C., & Miceli, T. (2010). Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning and Technology*, 14(1), 28-44.
- Krieger, D. (2003). Corpus linguistics: What it is and how it can be applied to teaching. *The Internet TESL Journal*, 9(3). Retrieved from <http://iteslj.org/Articles/Krieger-Corpus.html>
- Loi, C. K. (2012). *A corpus based analysis: The use of citation*. Paper presented at the International Conference of Innovative Research in a Changing and Challenging World. Puket, Thailand.
- Luo, Q. (2016). The effects of data-driven learning activities on EFL learners' writing development. *Springer Open*, 5(1255), 1-13. <https://doi.org/10.1186/s40064-016-2935-5>
- Luo, Q., & Liao, Y. (2015). Using corpora for error correction in EFL learners' writing. *Journal of Language Teaching and Research*, 6, 1333-1342. <https://doi.org/10.17507/jltr.0606.22>
- Manan, N. (2015). The use of integral citations in master's degree thesis. *International Journal of Education and Research*, 3(7), 233-246.
- Mansourizadeh, K., & Ahmad, U. K. (2011). Citation practices among nonnative expert and novice scientific writers. *Journal of English for Academic Purposes*, 10(3), 152-161. <https://doi.org/10.1016/j.jeap.2011.03.004>
- McCallum, L. (2016). Exploring authorial presence through the use of first person pronouns: Evidence from a Saudi university. *Arab World English Journal*, 7(2), 118-135. <https://doi.org/10.24093/awej/vol7no2.8>
- Navratilova, D. (2013). Authorial presence in academic discourse: Functions of author reference pronouns. *Linguistica Pragensia*, 1, 9-30.
- O'Sullivan, Í. (2007). Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL*, 19(3), 269-286. <https://doi.org/10.1017/S095834400700033X>
- Petric, B. (2007). Rhetorical functions of citations in high- and low-rated master's theses. *Journal of English for Academic Purposes*, 6(3), 238-253. <https://doi.org/10.1016/j.jeap.2007.09.002>
- Rabab'ah, G., & Al-Marshadi, A. (2013). Integrative vs non-integrative citations among native and nonnative English writers. *International Education Studies*, 6(7), 78-87. <https://doi.org/10.5539/ies.v6n7p78>
- Römer, U. (2010). Using general and specialized corpora in English language teaching: Past, present and future. In M. Campoy-Cubillo et al. (Eds.), *Corpus-based approaches to English language teaching* (pp. 18-38). London, UK: Continuum.

- Römer, U. (2012). Corpora and teaching academic writing: Exploring the pedagogical potential of MICUSP. In J. Thomas & A. Boulton (Eds.), *Input, process and product: Developments in teaching and language corpora* (pp. 70-82). Brno, Czech Republic: Masaryk University Press.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129-15. <https://doi.org/10.1093/applin/11.2.129>
- Sun, Y. (2003). Learning processes strategies and web-based concordances: A case study. *British Journal of Educational Technology*, 34(5), 601-613. <https://doi.org/10.1046/j.0007-1013.2003.00353.x>
- Swales, J. M. (1981). *Aspects of article introductions*. Birmingham, UK: Aston University Languages Study Unit.
- Swales, J. M. (1986). Citation analysis and discourse analysis. *Applied Linguistics*, 7(1), 39-56. <https://doi.org/10.1093/applin/7.1.39>
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge, UK: Cambridge University Press.
- Swales, J., & Feak, C. (1994). *Academic writing for graduate students*. Ann Arbor, MI: University of Michigan Press.
- Thompson, P. (2000). Citation practices in PhD theses. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 91-101). Frankfurt: Peter Lang.
- Thompson, P., & Tribble, C. (2001). Looking at citations: Using corpora in English for academic purposes. *Language Learning and Technology*, 5(3), 91-105.
- Vygotsky, L. (1986). *Thought and language*. Cambridge, MA: The MIT Press.
- Warchal, K. (2010). Moulding interpersonal relations through conditional clauses: Consensus-building strategies in written academic discourse. *Journal of English for Academic Purposes*, 9(2), 140-150. <https://doi.org/10.1016/j.jeap.2010.02.002>.
- Yoon, H., & Jo, J. W. (2014). Direct and indirect access to corpora: An exploratory case study comparing students' error correction and learning strategy use in L2 writing. *Language Learning and Technology*, 18(1), 96-117.

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).