

# *Enhancing writing pedagogy with learner corpus data*

ELENA COTOS

*Iowa State University, USA  
(email: ecotos@iastate.edu)*

---

## **Abstract**

Learner corpora have become prominent in language teaching and learning, enhancing data-driven learning (DDL) pedagogy by promoting ‘learning driven data’ in the classroom. This study explores the potential of a local learner corpus by investigating the effects of two types of DDL activities, one relying on a native-speaker corpus (NSC) and the second combining native-speaker and learner corpora. Both types of activities aimed at improving second language writers’ knowledge of linking adverbials and were based on a preliminary analysis of adverbial use in the local learner corpus produced by 31 study participants. Quantitative and qualitative data, obtained from writing samples, pre/post-tests, and questionnaires, were converged through concurrent triangulation. The results showed an increase in frequency, diversity and accuracy in all participants’ use of adverbials, but more significant improvement was made by the students who were exposed to the corpus containing their own writing. The findings of this study are thus interpreted as suggestive that combining learner and native-speaker data is a feasible and effective practice, which can be readily integrated in DDL-based instruction with positive impact.

Keywords: learner corpus, data-driven learning, learning driven data, L2 academic writing pedagogy, linking adverbials.

---

## **1. Introduction**

Since the mid-1980s, language corpora have shown tremendous potential in computer-assisted language learning, research, and teaching. Convergence between corpora and pedagogy has motivated fundamental changes in the ways we approach second language (L2) materials development, curriculum design, and teaching methodology. Although classroom applications of corpora, usually operationalized through concordancing programs (e.g., Bloch 2009), have not yet become mainstream practice, they have been very attractive to language teachers largely due to such advantages as salience of linguistic phenomena and extensive exposure to authentic language use in various registers and genres. These affordances have inspired some teachers to adopt data-driven learning (DDL; Johns, 1991) to create inductive, discovery-oriented learning opportunities whereby students analyze corpora undertaking a researcher role and engaging in active and autonomous learning (Chambers, 2010; Boulton, 2009, 2010; Braun, 2007). However, empirical evaluations of hands-on uses of corpora by L2 learners have remained relatively marginal (Rodgers, Chambers & Le Baron-Earle, 2011). Another concern is that empirical enquiry has been almost exclusively based on data produced by

native speakers of English (Gilquin, Granger & Paquot, 2007). While native-speaker corpora are indeed helpful in acquiring an L2 (Johansson, 2009), they cannot and should not be the only criterion for syllabus design because they “give no indication of what is difficult for learners” (Granger, Kraif, Ponton, Antoniadis and Zampa, 2007: 253). As Nesselhauf (2004: 125) reasonably points out: “For language teaching... it is not only essential to know what native speakers typically say, but also what the typical difficulties of the learners of a certain language, or rather of certain groups of learners of this language, are.” Learner corpora, i.e., electronic collections of authentic texts produced by L2 learners, can help to reveal those difficulties and to understand the differences between learner production and the features that characterize native-like language use.

To date, over a hundred learner corpora have been developed<sup>1</sup>, and interest in using them has increased steadily, especially in the area of academic writing. Numerous learner corpus findings have emerged from contrastive interlanguage analyses (Granger, 1996) identifying lexical, grammatical, phraseological, stylistic, and pragmatic features of learner language. Research suggests that English language learners clearly exhibit problems of frequency, register, positioning, semantics, and phraseology (Gilquin *et al.*, 2007). Although most of the findings are still largely at the level of implications, which may have delayed pedagogical use (Granger, 2009), learner corpus insights are slowly but surely making their way toward the classroom. Many have acknowledged the promise of what Seidlhofer (2002) termed the *learning driven data* (LDD) approach, which uses learner corpora for language teaching purposes. Local or in-house learner corpora, in particular, have been highly recommended to address the specific linguistic issues of a given learner population (Mukherjee & Rohrbach, 2006; Seidlhofer, 2002). Flowerdew (2001: 364) urges practitioners to implement “insights gleaned from learner corpora... to complement those from expert corpora for syllabus and materials design”, and Granger (2004) emphasizes the need for more publications advocating the use of learner corpora to inform pedagogical practice.

This paper attempts to bring together learner corpus research and pedagogy in an investigation of learner use of linking adverbials (LA), exploring DDL pedagogy that combines learner and native-speaker data. To achieve this purpose, the study unfolded in two stages. Initially, a local learner corpus of student writing produced as part of the coursework was created to examine LA patterns and conduct indirect comparison with learner corpus research and with native-speaker corpus-based reference materials. This informed the development of two types of DDL activities: the first included a native-speaker corpus of research articles, and the second added ‘learning driven data’ from the local learner corpus. These activities were then implemented in a pedagogical experiment in the context of advanced L2 academic writing instruction with two groups of graduate students, referred to as the NSC and LDD groups respectively. Observed LA performance data were obtained from the learner corpus compiled in the course of a semester and from pre- and post-tests. Learner perceptions of LAs and on the effects of the DDL activities were collected through questionnaires. Triangulation of results suggests that the use of local learner corpora combined with native-speaker corpora can strengthen DDL instruction by reinforcing students’ understanding of LAs, and potentially leading to improved knowledge as well as to an increase in the frequency, diversity, and accuracy of their use.

---

<sup>1</sup> An extensive LC account can be found on the website of the Centre for English Corpus Linguistics: <http://www.uclouvain.be/en-cecl-lcworld.html>

## 2. Learner corpus research and second language writing pedagogy

To provide a framework for the current study, this section briefly introduces LAs as part of the writing construct, then proceeds with an overview of learner corpus evidence on LA use, and concludes with a discussion of DDL classroom applications of native-speaker and learner corpora.

### 2.1. *Linking adverbials and the writing construct*

Organizational competence is key to producing good quality writing. An essential aspect of the writing construct inherent to organizational competence is textual knowledge, i.e., the ability to produce and understand features of extended discourse that facilitate text cohesion (Bachman & Palmer, 1996). In academic writing, cohesion is often realized with the help of linking adverbials. LAs are defined as devices that “serve to make semantic connections between spans of discourse of varying length” (Biber, Johansson, Leech, Conrad & Finegan, 1999: 558) and can be divided into several semantic categories: enumeration and addition, summation, apposition, result/inference, contrast/concession, and transition<sup>2</sup>. The complex nature of these adverbials is markedly difficult for L2 writers given their semantic as well as their grammatical and positional variation, as has been attested by learner corpus research.

### 2.2. *Linking adverbials in learner corpus studies*

Studies investigating learner corpora have generally concentrated on the role of adverbials in establishing inter-sentence relations and intra-text connectedness. Commonly found patterns reveal that learners of English tend to overuse additive, appositive, and transition LAs; underuse result and contrast LAs; misuse adverbial functions; employ fewer LA types; and use them inappropriately (Altenberg & Tapper, 1998; Blagojeva, 2002; Chen, 2006; Flowerdew, 1998; Granger & Tyson, 1996; He, 2002; Hyland, 2004; Hyland & Tse, 2004; Lei, 2012; Shaw, 2009; Tankó, 2004). Green, Christopher and Lam (2000: 99) also uncovered differences in the positioning of LAs in the sentence, arguing that inappropriate placement “has a deleterious effect on... both local and global text coherence”.

The focus in these studies was primarily on overall frequencies. Recent investigations have undertaken more elaborate analyses and report more insightful findings. For instance, Carrió-Pastor (2013) examined LA linguistic variation across different sections of research articles (which is also the register targeted by instruction in the context of this study), finding that it may depend on the writers’ linguistic, cultural and social background. Similarly, Charles (2011a, 2011b) deduced that the linguistic choice of a given adverbial may be influenced by factors such as genre, discipline and context, and communicative function.

Learner corpus studies are pedagogically-driven in that they “claim or at least imply that they attempt to make a contribution to language teaching” (Nesselhauf, 2004: 134). However, they have had little impact on pedagogy because researchers have largely scrutinized patterns of over/underuse, while teachers have mainly focused on eradicating errors (Milton, 1998; Shei & Pain, 2000; Hegelheimer, 2006). To bridge this gap, this study

<sup>2</sup> In the literature, LAs have also been referred to as internal connectors, logical connectives, adverbial connectors, cohesive markers, conjuncts, and conjunctive ties.

utilizes corpus-based materials that are supported by learner corpus findings about the use of linking devices.

### 2.3. *Data-driven learning practices*

DDL typically consists of having students analyze corpus data in the form of computer-generated concordances, frequency lists, clusters, distributions, etc., in order to illustrate patterns of L2 use. The DDL approach is renowned among practitioners for numerous advantages, especially for having a high degree of authenticity and salience. While acknowledging its benefits, Osborne (2004) argues that native-speaker corpora may not always be a suitable model for learners due to the fact that they can exhibit instances of language use that run counter to the L2 rules taught in class. Therefore, Osborne (2004: 253) proposes the learner corpus approach, which is believed to draw learners' attention to problematic areas in their own collective production. Seidlhofer (2002) calls it "learning-driven data" (LDD), adapting Johns' (1991) terminology for the use of learner corpora. According to Nesselhauf (2004: 139), LDD "can be attempted straight away by anybody who has access to a learner corpus or is willing to create one." Indeed, practitioners with first-hand access to their student writing can compile learner corpora without much difficulty and use them as a classroom resource. In a very practical vein, Millar and Lehtinen (2008: 62) showcase how to create a local learner corpus "in a relatively 'quick and dirty' way", offering a description of tools and approaches to learner corpus analysis. Rankin and Schiffner (2011) also provide a good example of how to exploit a local learner corpus by first identifying quantitative and qualitative differences with NS corpora, and then developing teaching materials in order to raise learners' awareness of the linguistic problems identified. Similar examples can be found in Pérez-Paredes (2003, 2004) and in Pérez-Paredes and Cantos Gómez (2004).

Even though some work has been conducted using LDD, the model is not as well supported empirically as implementations of DDL using native-speaker corpora. Nevertheless, the results of the few existing studies are clearly informative for classroom practice. Belz and Vyatkina (2005) and Ragan (2001) provide evidence showing that LDD can positively impact the frequency and accuracy of targeted linguistic items and contribute to an increase in learners' awareness of their meaning, function, and generic distribution. Seidlhofer (2002) suggests that exploring learner data can be a strong motivational factor for learners. Nesselhauf (2004) and Pérez-Paredes (2004) find that negative evidence noticed during LDD tasks is beneficial for language acquisition. Bernardini (2004) and Meunier (2002) further conjecture that combining native-speaker and learner corpora could make problematic language use even more salient to language learners. There seems then to be a need for continued empirical testing of learner corpus-driven materials and their influence on learning outcomes in order to enhance the potential of the DDL approach with effective LDD. It is this particular need that this study seeks to address.

### 3. The study

The present work investigates the effects of a pedagogical experiment that operationalized LAs as a specific writing difficulty based on the needs of individual learners, as established by a contrastive analysis of their pre-experiment writing compiled into a local learner corpus. In the experiment, two types of DDL activities were implemented with two groups

of students: one group completed activities using a native-speaker corpus only (NSC group); the other used both the native-speaker and the local learner corpus (LDD group). With an emphasis on applying the learner corpus as a research and teaching tool, the aim was to explore its learning potential and capacity to heighten the value of DDL.

### 3.1. Context and participants

The study was conducted in an advanced academic writing course for graduate students at a North American university. The course focused on the conventions of the research article genre. It was developed as described in Cortes (2007)<sup>3</sup>, using a specialized corpus of research articles and exercising top-down and bottom-up techniques (Charles, 2007) to complete corpus-driven language project tasks (Cheng, 2012). Each student in the course used a corpus of 35 to 45 articles (roughly 30,000 words) in his/her own discipline, which was part of a larger corpus that comprises 40 academic disciplines and amounts to 1,623 manuscripts with a total of 1,322,089 words. The larger corpus had been developed prior to this study to accommodate the disciplinary heterogeneity which is characteristic of this course. The articles in each discipline were published in reputable journals and evaluated as appropriate models of genre writing by faculty in respective fields.

Thirty-one students (19 male, 12 female) participated in the study. They were international graduate students in one of the following disciplines: Computer Engineering, Genetics, Physics and Astronomy, Statistics, Business Administration, Journalism, Sociology, and Curriculum and Instruction. These students, aged between 23 and 31, had different language backgrounds and had had an average of 6.3 years of L2 English instruction before enrolling in a graduate program at this university. Based on the TOEFL iBT, their overall English proficiency can be considered intermediate to advanced (scores 83 to 107), and their writing skills can be described as fair to good (scores 19 to 24). The institutional English Placement Test also indicated a similar level of writing ability, which is why they were placed in this particular course.

### 3.2. Approach

The study adopted a mixed-methods form-function analysis, focusing on the use of specific linguistic forms to explicate the functions they map on to (Ellis & Barkhuizen, 2005). The linguistic forms here are individual LAs, and the functions they perform are of a semantic, semantico-grammatical, and discourse nature. Two groups of students consented to participate in the study. Due to the constraints of the instructional context, which are often inevitable in instruction-embedded research, this constituted a convenience sample, but it is believed to be representative of international graduate students as a whole. The study design is quasi-experimental since a control group was not available and since both participating groups conducted a particular type of DDL activity. The LDD group (fifteen students) is considered experimental and the NSC group (sixteen students) suitable for comparison purposes. Quantitative and qualitative data were collected and analyzed through concurrent triangulation (Creswell, 2003). The quantitative data consisted of LA frequency counts and pre-/post-test results, which were used as a basis for evaluating the link between learners'

---

<sup>3</sup> Cortes (2007) provides descriptions of the corpus-based tasks, assignments, and concordancer.

Table 1 *Local learner corpus design criteria*

<i>Attributes</i>	
Class	Graduate students
Level of English	Intermediate to advanced
L1	Various
Learning context	ESL
Medium	Written texts
Genre	Research report

Table 2 *Local learner corpus composition (words and number of papers)*

<i>Groups</i>	<i>Pre-experiment</i>	<i>Immediate post-experiment</i>	<i>Delayed post-experiment</i>
NSC	51,472 (63)	13,939 (16)	36,752 (16)
LDD	46,978 (60)	12,392 (15)	41,424 (15)
Total	98,450 (123)	26,330 (31)	78,176 (31)

use of LAs and the type of DDL activity they completed. The qualitative data included participants' responses to questionnaires as well as their written productions, which were compiled into the local learner corpus.

### 3.3. *Local learner corpus*

The local learner corpus is an electronic collection of writing produced by the participants as course assignments, which encapsulates the design criteria that reflect the specifics of the instructional context (Table 1). It was the source of data for preliminary analysis of participants' use of LAs as well as for the content of the LDD activities.

The local learner corpus was collected in three stages – before, immediately following, and four weeks after the experimental implementation of the DDL activities (see Table 2). The texts in this corpus were written as minor and major course assignments: argumentative essays written for diagnostic purposes at the beginning of the semester, a set of reports on students' corpus observations of genre conventions in their discipline<sup>4</sup>, and final term papers in the form of research articles. It should be noted that, although the writing assignments through which the texts were obtained differed in topic and scope, they are comparable due to the similarity in their overall function – that of building and supporting an academic argument by reporting, analyzing, and discussing particular evidence, which requires effective use of LAs.

### 3.4. *Instruments and materials*

**3.4.1 Questionnaires.** Two open-ended questionnaires were administered to each group of participants before and after the pedagogical experiment. The first questionnaire was developed drawing from the principles of questionnaire design (Dörnyei 2003; Yoon &

<sup>4</sup> Cortes (2007) explains the rationale for assigning such corpus-linguistics research tasks.

Hirvela 2004) and was piloted with similar non-participants. It invited the students to reflect on their academic writing strengths and weaknesses in general, and on their knowledge and use of LAs in particular. The questions were designed to obtain an insight into the participants' understanding of the notion of cohesion and its realization in academic writing, their knowledge of LAs, and LA use in their own writing. The second, post-experiment questionnaire elicited similar perception data plus participants' evaluation of the effects of the type of activity they were exposed to.

*3.4.2 Pre- and post-tests.* Before and immediately after the implementation of the DDL activities in each group, the participants took tests which were presented to them as regular instruction activities. The first task was to identify LAs in a given excerpt, and the second was to complete a multiple choice LA classification task. The tasks were designed using Biber *et al.*'s (1999) semantic classification of adverbials. The test items were randomly selected from the specialized corpus of research articles used in class and were representative of the participants' various disciplines. In each item, 60-word stretches of text preceding and following the missing adverbial were provided in order to avoid ambiguity of meaning. For the same reason, all the distracters in the response options were chosen from semantic categories other than the category to which the correct answer belonged. To ensure that the distracters were not a possible correct variant, the tests were piloted with native-speakers.

*3.4.3 Corpus data-driven materials.* The materials were designed according to learner empowerment and DDL principles, the theoretical underpinning going back to schema theory (Bartlett 1932). From a language learning perspective, Barlow (1996: 30) maintains that schema-based restructuring occurs with repeated exposure to instances of language use and that data-driven instruction can reduce this lengthy process "by concentrating and manipulating instances of a language phenomenon, mak[ing] the patterns stand out clearly."

For both types of DDL activities, the materials were similar in that they included focused emphasis on the semantic roles, forms, and syntactic distribution of LAs. They provided a definition and a few examples of native-speaker use of adverbials in a given category, informed the students of the frequency with which they used those LAs prior to the pedagogical experiment as recorded in the local learner corpus, and then introduced three tasks (see example in Appendix A). The first task required them to examine a number of teacher-selected examples of LAs in terms of similarities and differences. The second task was motivated by the underlying DDL "assumption... that effective language learning is a form of linguistic research" (Johns, 1991: 30). The students were asked to query the corpus with a concordancing tool for individual LAs that belonged to a given semantic category and to extract examples that were helpful for understanding how to use them. The NSC group completed this activity with a specialized corpus in their discipline, and the LDD group explored both native-speaker corpus data and their own productions archived in the local corpus. The last was a follow-up reflection task where the students had to think about their own use of LAs and make recommendations based on what they noticed in the corpus<sup>5</sup>.

---

<sup>5</sup> The Moodle course management system was used to deliver the tasks, similar to the studies reported by Pérez-Paredes, Sánchez-Tornel, Alcaraz Calero and Aguado Jiménez (2011) and Pérez-Paredes, Sánchez-Tornel and Alcaraz Calero (2012), where L2 learners also conducted DDL explorations.



Table 3 *Experimental procedure*

<i>Stages</i>	<i>Week</i>	<i>Steps</i>	<i>Data</i>
Pre-experiment	1-8	Pre-experiment production	Writing performance
	8	Indirect comparison of local learner corpus data with learner corpus empirical findings and native-speaker corpus-based reference materials	LA use in pre-experiment local learner corpus
	9	Pre-experiment questionnaire	Metalinguistic/pragmatic awareness of LAs
	9	Pre-test on LAs	LA competence
Experimental implementation of DDL activities	10	Introduction to LA semantic categories	
	11	Activities on result/inference LAs	
	11	Activities on apposition and contrast/concession LAs	
	12	Activities on enumeration/addition and summation LAs	
Post-experiment	12	Post-test on LAs	LA competence
	12	Post-experiment questionnaire	Metalinguistic/pragmatic awareness of LAs
	13	Immediate post-experiment production	LA use in post-experiment local learner corpus
	17	Delayed post-experiment production	LA use in delayed post-experiment local learner corpus

### 3.5 *Experimental procedure*

A summative overview of the experimental procedure, which was the same for both groups, is presented in Table 3.

*3.5.1 Pre-experiment.* As previously mentioned, the pedagogical experiment was preceded by the collection of the first component of the local learner corpus and the identification of LAs in order to better understand student problems in this area. As shown in Table 4, the frequency counts display differences in the use of semantic categories. Of a total of 461 occurrences<sup>6</sup>, 43.4% belonged to the apposition category, 23.6% to result and inference, 20.4% to contrast and concession, 11.9% to enumeration and addition, and 0.6% to summation LAs. Also, in each semantic category there were individual adverbials that accounted for the high frequency of occurrence within the group (e.g., *for example* which appeared 171 times out of 200; *however* – 52 times out of 94; *first* – 23 times out of 55). This distribution can, to some extent, be considered congruent with previous learner

<sup>6</sup> Multi-word adverbials were considered one-token units.



Table 4 *LA frequency in the pre-experiment local learner corpus*

<i>Apposition</i>		<i>Result/ Inference</i>		<i>Contrast/ Concession</i>		<i>Enumeration/ addition</i>		<i>Summation</i>	
for example	171	so	47	however	52	first	23	in conclusion	2
for instance	8	therefore	42	though	20	in addition	14	to summarize	1
in other	5	thus	10	on the other	11	second	8		0
words		hence	6	hand		next	5	in sum	0
that is	5	consequently	4	nevertheless	2	further	4	to conclude	0
i.e.	5			yet	2	similarly	1	all in all	0
specifically	5			alternatively	1	lastly	0	overall	
namely	1			by	1	to begin	0		
which is to	0			comparison		with			
say				in contrast	0				
				conversely	0				
				in spite of	0				
Total (461)	200		109		94		55		3

corpus research that revealed an overuse of apposition and addition adverbials and an underuse of result, contrast/concession, and summative devices in formal academic essays (Granger & Tyson 1996; Altenberg & Tapper 1998; He 2002), which share the communicative functionality of the assignments in this study. While the use of LAs in the local corpus was not interpreted as 'overuse' and 'underuse' due to the absence of a comparable native-speaker corpus, it pointed to some tentative patterns that were worth the attention of the teacher, helping establish a baseline for addressing LAs in the classroom.

The indirect comparison extended to LA patterns in native-speaker academic prose as documented by Biber *et al.* (1999)<sup>7</sup>, and showed that the participants' use of LAs was distinct from native-speaker writing, where result/inference adverbials usually account for the largest proportion, apposition and contrast/concession the second largest, and enumeration/addition and summation adverbials the third. Additionally, in the local learner corpus, single adverbials (50.5%) were almost as frequent as prepositional phrases (49%) while native-speaker use is more varied (Biber *et al.*, 1999: 884). According to the same source, the beginning of a sentence is the most common position for LAs in native-speaker academic prose. The middle positions account for the second highest proportion of occurrences, and LAs in final position are rare but still present. In the local corpus, there were no instances of sentence-final placement, the adverbials being distributed mostly at the beginning and in the middle of sentences. This preliminary analysis confirmed the need to address this particular issue and informed the content of the instructional materials. This need was also substantiated by participants' responses to the pre-experiment questionnaire and by their results on the pre-test, which will be presented in Section 4.

*3.5.2 Experimental implementation of DDL activities.* The activities were implemented as language focus tasks in the tenth week of the semester in 80-minute class periods. First, the students were introduced to the semantic categories of LAs through explanations and examples. The instructor also referred to the findings in Biber *et al.* (1999), discussing how native speakers tend to use LAs in academic discourse and how the LAs identified in the pre-experiment local learner corpus appeared to suggest different patterns. The teaching points for the following three classes focused on specific categories of LAs (see Table 3). The NSC and LDD groups each worked with materials developed for the respective type of activity.

*3.5.3 Post-experiment.* Following the DDL activities, the participants took the post-test and responded to the second questionnaire. A week later they submitted a written assignment, and those texts formed the post-experiment component of the local learner corpus. At the end of the semester, the students submitted their final term papers, which constituted the delayed production component of the corpus.

---

<sup>7</sup> Biber *et al.* (1999) was chosen for this comparison as it is considered a generalizable corpus-based description of native-speaker language use, and also because their corpus of academic prose includes book extracts and research articles in a wide range of academic disciplines, which is similar to the texts in the local learner corpus.

### 3.6. Analysis

To gain an understanding of the potential of LDD as a branch of the well-established DDL approach, two main data analysis directions were pursued:

- observed and perceived performance – examining the changes in all participants' written production and knowledge of LAs before and after the experiment;
- effects of DDL activity types – comparing observed and perceived performance between the LDD and NSC groups according to the type of activity they completed.

Examination of observed performance of all participants began at the pre-experiment stage, when the local learner corpus was analyzed in terms of LA frequencies as well as semantic, syntactic, and positional realizations. The same analyses were carried out for the immediate and delayed post-experiment production to detect quantitative changes in written performance. To establish whether LAs were used appropriately, three sets of 31 pre-, immediate post-, and delayed post-experiment student texts (16 from the NSC group and 15 from the LDD group)<sup>8</sup> were evaluated qualitatively by the researcher and a second rater (Cohen's  $k = 0.92$ ).

Changes in knowledge of LAs and the effects of the DDL activity types were measured through a number of  $t$ -tests that juxtaposed binary scores for individual test items. Initially, comparability between the NSC and LDD groups (which was assumed given the students' placement in the course based on the results of the institutional test) was established through a two-tailed independent  $t$ -test using the scores from the pre-test, which showed no significant difference ( $t(22) = 0.98$ ,  $p = 0.338$ ). The effect size was relatively moderate ( $d = 0.41$ ). Since it could be hypothesized that students' knowledge of LAs after completing distinct types of activities may differ between the two groups, a one-tailed independent  $t$ -test was run to compare the NSC and LDD mean scores on the post-test. Hypothesizing that the corpus-based activities would contribute to improvement in students' knowledge in both the NSC and LDD groups, one-tailed paired  $t$ -tests were run for each group on pre- and post-tests. Cohen's  $d$  values were also calculated in order to quantify the magnitude of the difference between and within groups. Finally, the results were triangulated with participants' perceptions elicited by the two questionnaires.

## 4. Results and discussion

### 4.1 Observed and perceived performance

First, the impact of the DDL activities was investigated by evaluating all students' observed performance in the form of pre- and post-experiment writing and test scores as well as their perceptions of LA knowledge. Table 5 summarizes the production of both groups<sup>9</sup> in terms of the tokens normalized per 1,000 words, showing that the participants employed adverbials notably more often after having conducted the DDL tasks, although changes in

<sup>8</sup> Since the pre-experiment component of the local learner corpus comprised more papers than the other two components, 31 papers for the pre-experiment set were randomly selected for manual analysis (16 out of 63 papers from the NSC group and 15 out of 60 papers from the LDD group).

<sup>9</sup> The comparison was twofold, i.e., it was done for both groups individually, but the numbers for the two groups are combined in Table 5 and Table 6 for conciseness.

Table 5 LAs in the local learner corpus (normalized per 1,000 words)

	<i>Pre- experiment</i>	<i>Immediate Post- experiment</i>	<i>Delayed Post- experiment</i>
Apposition	2.03	1.94	1.83
Result/Inference	1.11	2.47	3.01
Contrast/Concession	0.95	2.62	2.25
Enumeration/Addition	0.56	1.67	1.84
Summation	0.03	0.11	0.09
<i>Total</i>	4.68	8.81	9.02

Table 6 Positional and syntactic realizations of LAs in the local learner corpus

	<i>Pre- experiment</i>	<i>Immediate Post- experiment</i>	<i>Delayed Post- experiment</i>
Position			
sentence-initial	87.3%	65.1%	63.7%
mid-sentence	12.7%	34.9%	33.4%
sentence-final	0%	0%	2.9%
Syntax			
adverb phrases	4.3%	5.6%	6.1%
finite/non-finite clauses	2.9%	3.2%	4.4%
prepositional phrases	47.4%	24.8%	38.2%
single adverbials	54.6%	66.4%	51.3%

LA frequency were not statistically significant (immediate post-experiment  $p = .07$ ; delayed post-experiment  $p = .09$ ). The use of apposition LAs decreased slightly while the use of other semantic categories increased. Analysis of qualitative changes demonstrated that the students began to employ more varied LAs and that over-reliance on some adverbials prior to the experiment decreased. This suggests that learners better understood how synonymous adverbials can perform the same or very similar functions. However, it seems that the students still hesitated to use LAs that they had not been very familiar with before, which may mean that they had the opportunity to consolidate the LAs they were somewhat familiar with but, perhaps, did not have sufficient opportunities for practicing unfamiliar ones.

Table 6 shows the changes that were detected with respect to position in the sentence and grammatical form. Post-experiment, mid-sentence placement of LAs became more frequent and sentence-initial position slightly less frequent. As for grammatical diversity, adverb phrases and finite/non-finite clauses were still rare in the post-experiment production, and the number of prepositional phrases decreased, unlike that of single adverbials. Comparisons with the delayed post-intervention component, overall, yielded similar results, which approximate Biber *et al.*'s (1999) description of native-speaker LA use.

While increased frequency may be partly accounted for by the fact that the students were taking note of particular forms rather often as they were completing the corpus-based tasks and recording them in their written responses, the results presented above are still encouraging because frequent citing of adverbial forms can be viewed as repeated

interaction with the target input, which is an essential condition for language acquisition. More importantly, the data showed that the adverbials were used appropriately. Manual analysis by two raters of pre-experiment, immediate post- and delayed post-experiment student texts revealed that before the DDL activities, learners used LAs appropriately 41% of the time, but this number increased to 81% shortly after the experiment, and to 76% in the delayed production. Consider these examples from the pre-, immediate post-, and delayed post-writing, respectively, by the same student:

- They also reveal some insights about the papers in my area. **That is:** try hard to clarify the procedures that are adopted and make sure their credibility at the same time.
- When the research is based on more than one hypothesis, in the discussion section these hypotheses are discussed in a cyclical organization. **That is,** findings or results are stated first and then followed by the abstract meanings or concept based on such results.
- However, men seem to pay attention to gaining ability to make more money and get a better job. **That is,** their expectations of what they can get from college are very goal-oriented. (LDD\_st4)

Higher-frequency LAs in the pre-experiment production were often used inappropriately, but their use improved and their frequency decreased in post-experiment writing. As in this example – “Though the girls and boys show similar interest in studying new things, however, generally speaking, females consider much more about the future job than males.” (pre-experiment, LDD\_st8) – some adverbials tended to often appear inappropriately in sentences that also contained other adverbials, especially contrast/concession, but this tendency was not observed in delayed post-writing. LAs that are more typical of informal speech became less frequent as well. Instead, the students used more formal synonyms (e.g., *therefore* instead of *so*; *for instance*, *namely*, or *specifically* instead of *for example*; *second*, *further* or *lastly* instead of *next*), as they may have realized that academic writing requires linking through more formal vocabulary.

The pre- and post-experiment questionnaires add an introspective insight into these findings. Participants’ perceptions of their knowledge about textual cohesion before and after the DDL activities appeared to be different. Before, many thought of cohesion as being realized through the use of short sentences, correct tenses, repetition of main points or terminology, logical organization, highlighting cause-effect / similarity-difference relationships between ideas, explaining concepts, justifying statements, or planning a draft. Some students admitted “I have no idea” or “I’m not sure” how to make writing cohesive. Nevertheless, 48% did mention linking devices, referring to them as “linking words” or “linking phrases,” “transition words” or “conjunctions.” When asked if they knew what LAs were, most participants said yes, though most showed lack of confidence with statements like “I think I know” and “I am not sure what I know is correct.” Their attempts to define and exemplify LAs confirm their incomplete understanding, as can be seen in these extracts:

- Linking adverbs are those adverbs that connect words in a sentence. (DDL\_st2)
- Linking adverbs are the words that connect the previous part and the following part together and help express opposite or similar attitudes of the author. (LDD\_st15)

Table 7 Descriptive statistics for pre- and post-tests

		Mean	N	SD	SE
NSC group	pre-test	6.5	16	3.20	.925
	post-test	11.5	16	2.81	.812
LDD group	pre-test	5.2	15	3.05	.880
	post-test	10.0	15	1.28	.369

The students mentioned a variety of LAs in their questionnaire responses: some that were found most frequently in their pre-experiment writing as well as some that were very rare. Such variability may be attributed to an inconsistency in the learners' form-function mappings. It may be that the mappings were better developed for some LAs than for others, and therefore those LAs were more actively used. In addition, the students named many coordinators and subordinators (*and, but, although*), which means that they were associating LAs with conjunctions but did not have a clear understanding of the similarities and differences in these connectors' roles. It is not surprising then that 94% of the students, when asked to identify the LAs in a given excerpt in the pre-test, highlighted words and expressions other than the expected items.

In the post-experiment questionnaire, many participants (73%) explicitly stated that they became much more "aware" and confidently claimed that they knew considerably more about LAs. Their explanations of what they learned included "new words," "types and meanings," "functions," "positions," "different forms for the same function," "punctuation," "how to appropriately use them," "how to link ideas correctly," "how to make writing more fluent," "how to make writing more varied," and, in general, "details" they did not know before. Finally, all the students affirmed that they realized the importance of LAs in academic prose and that they intended to improve the quality of their writing with more varied choices.

In addition to enhanced awareness, the results point to considerable improvement in students' knowledge of LAs after having observed their occurrences in the corpus. Table 7 presents the descriptive statistics based on students' scores on pre- and post-tests where they completed LA identification and classification tasks, showing higher means after the pedagogical experiment. One-tailed paired *t*-tests substantiated the difference in means with statistical significance for both NSC ( $t(11) = 5.34, p < .001$ ) and LDD ( $t(10) = 10.34, p < .001$ ) groups. The effect size was large in both cases ( $d = 1.66$  and  $d = 2.05$ , respectively), though generalization is difficult due to the small number of participants.

#### 4.2 Effects of LDD and NSC activities

To investigate the potential of combining native-speaker and learner output compared to exposing students only to native-speaker data, the frequency and variety of LAs used by the NSC and LDD groups in the immediate post-experiment production were compared. Although the syntactic and positional patterns of LAs in the writing of these groups were similar, the percentages in Table 8 signal a considerable difference in frequency – the LDD group exhibited an obvious tendency to employ various linking devices more often than the NSC group. However, while this is likely a positive result, it should be interpreted with

Table 8 Frequency of LA semantic categories in the immediate post-experiment component of the local learner corpus

Apposition	Result/ Inference		Contrast/ Concession		Enumeration/ Addition		Summation		
	NSC	LDD	NSC	LDD	NSC	LDD	NSC	LDD	
23.5%	76.5%	38.5%	61.5%	29.0%	71.0%	43.2%	56.8%	0%	100%

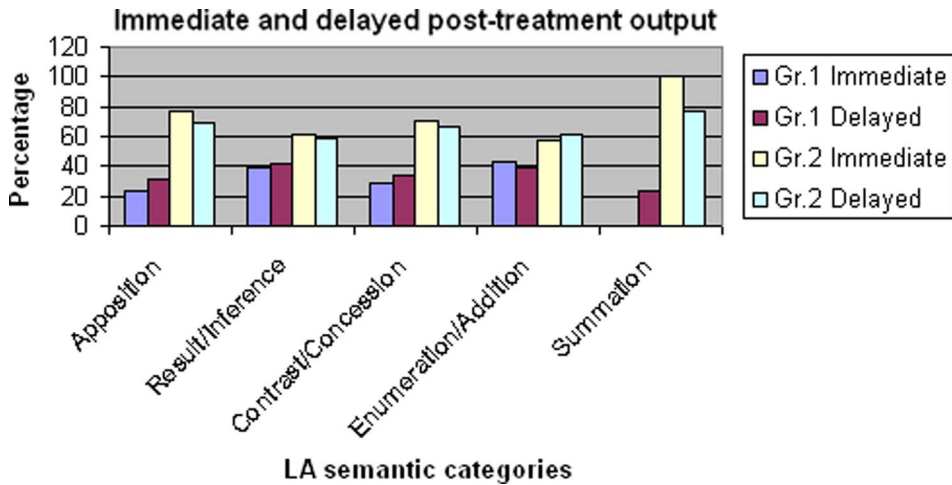


Fig. 1. LA use by the NSC and LDD groups in the immediate and delayed post-intervention local learner corpus

caution since the LDD students may have overused LAs. At the same time, this inference cannot be confidently made here due to the lack of comparable native-speaker data.

These findings were corroborated when comparing immediate and delayed post-experiment writing. Figure 1 shows how the use of LAs by the LDD group spanned more confidently across semantic categories compared to the NSC group. When LA use was manually analyzed for appropriateness by the two raters, it became clear that the LDD group performed better in the delayed post-experiment production, with 89.2% of appropriately used instances as opposed to 62.1% in the NSC group. These results are consistent with the findings of Belz and Vyatkina (2005), who implemented a developmental pedagogical intervention targeting modal particles, for which they collected a longitudinal corpus of learner German with a built-in control corpus of L1 German and conducted a comparison of successive language production. Additionally, a one-tailed independent *t*-test used to compare the mean scores of the LDD and NSC groups, which were obtained on the post-test, yielded statistical significance ( $t = 1.932$ ,  $df = 21$ ,  $p = 0.03$ ), suggesting better learning outcomes for the LDD group. The practical significance for the experiment impact was moderate to high as indicated by Cohen's effect size value ( $d = 0.68$ ). Consequently, it can be inferred that, in the context of this experiment, the LDD approach may be more conducive to improved knowledge and use of LAs.



These results were then reinforced with insights from the students' perspective. As reflected in the post-experiment questionnaire, both groups perceived the corpus-based work as helpful and beneficial for the learning of LAs. Reasoning about why it was helpful, they highlighted the advantages and disadvantages of the type of activity they completed. The advantages can be summarized as: plentiful exposure; attention to form, meaning and function; attention to the quality of writing; variety of contexts; and variety of syntactic structures – all viewed as beneficial effects in the DDL literature (see Bernardini, 2002; Flowerdew, 2012).

Although both groups expressed similar advantages, the LDD group found the local learner corpus excerpts particularly “helpful because it comes from our writing” (LDD\_st2). The majority of students in this group (92%) made a constant parallel with their own writing; for example, “through this approach, I realized my shortcomings; I found many misuses in the learner corpus, some of which I myself have” (LDD\_st9). It appears that observing the behavior of LAs in the native-speaker corpus helped them recognize items that were misused in their local corpus: “...we could learn from the good examples and find the flaws in our own writing” (LDD\_st12).

Furthermore, the students in the LDD group seem to have been more cognitively involved in the process of learning, for most of them (79%) mentioned drawing individual conclusions about the use of LAs after having thought about how they would personally use them in the examples provided. Interestingly, this appeared to be a common strategy when they compared the excerpts from native-speaker discourse with those from their own corpus.

The NSC students, on the other hand, pointed out that they found themselves attempting to memorize certain LAs in the examples produced by native speakers, which was not always easy because they did not have opportunities for immediate application and practice. It is interesting that eleven out of thirteen NSC students mentioned memorizing compared to only four LDD students. For the NSC students, memorizing was perhaps the most readily available and commonly used cognitive strategy. They may have “tried to remember them just like other things, like new words or verbs in past that are not *-ed* past” (NSC\_st4). Unlike their peer group, the LDD students may have found themselves memorizing less often because for them, comparing native-speaker and learner production was more of an engaging problem-solving task that triggered another cognitive mechanism – focus on negative evidence – which facilitated learning.

Another advantage surfacing from the post-experiment questionnaire responses of the LDD group is increased learning drive. For these students, the activities were interesting and exciting because they engaged with examples that came from them, and this personal appeal aroused interest in what they were learning in class. For example:

- I am surprised that we only use quite few words out of so many linking words. It sure motivates me to learn to use more of the other linking adverbials. (LDD\_st3)
- When I was reading the examples, I saw a sentence that was my sentence. I said wow! I didn't think that was wrong! I need to understand this better. (LD\_st16)

Evidence such as this tends to support Seidlhofer's (2002) conclusion about learner corpora being a strong motivational factor. Enabling the LDD students to analyze their local corpus, which was something new for them, created a motivating learning environment where the corpus-based tasks were perceived as a personally relevant learning experience.

Some disadvantages of each approach also emerged as themes in students' answers. Unlike the NSC group, the LDD students referred less often to such challenging factors as the hardship of drawing conclusions and the insufficiency of practice opportunities. The LDD group, however, rightfully brought up the time issue given that they had to search and examine concordance lines from two corpora. They also often pointed to the need for additional assistance from the instructor, especially for feedback on tasks that required them to make corrections in the sentences extracted from the local corpus. Providing individualized feedback on every classroom activity is very time consuming, so it is advisable to supply learners with answer keys for at least a sample of corrected misuse to help them confirm their linguistic hypotheses. Bernardini (2000) warns that along with the excitement of discovery and problem-solving, students can feel uncertain and frustrated as they notice problematic patterns in their target language production; this also appeared to be the case for seven of the fifteen students in the LDD group. To prevent that, the teacher can carefully select both negative and positive evidence from the learner corpus, first providing students with teacher-generated worksheets and modeling the analysis of select examples, and then giving them the opportunity to explore the corpus themselves and discuss their interpretations with each other and the teacher.

Like the comparable group, the LDD students mentioned that sometimes they could not comprehend the meaning of the input provided in the concordance lines because it was extracted from various academic disciplines. It is not surprising that a student majoring in Biology might not understand an excerpt from Electrical Engineering. Here again, the teachers' role is paramount for they can anticipate this difficulty by carefully selecting more representative examples, or even by editing good but difficult ones.

Both NSC and LDD activities engaged the students in language research, provided them with access to valid input, and helped them draw conclusions from observed LA occurrences. However, although the participants positively evaluated these types of activities, LDD appears to be more appealing and more conducive to learning. The evidence obtained in this study further suggests that a major strength of learner corpora, also emphasized by Seidlhofer (2002), is that they can trigger focus on negative evidence by giving the students access to their own problem areas and, consequently, accelerate acquisition by enhancing discovery of the differences between their interlanguage and the target language.

## 5. Conclusion

This study investigated the implementation of activities based on both native-speaker and learner corpora, attesting to the power of the DDL approach to increase learners' awareness of observed linguistic phenomena, facilitate knowledge development, and improved use of target linguistic forms. It also confirms the value of the extended, 'learner-driven data' DDL approach, suggesting that the potential of combining learner and native-speaker data is as strong as, if not stronger than, exposing learners to authentic native-speaker discourse only. The main implication here is that supplementing corpus materials with learner output promises to be effective practice that can be readily integrated in DDL instruction with positive impact. Therefore, L2 writing practitioners are encouraged to develop in-house learner corpora in order to identify and better tackle issues that are immediately relevant to their students.

For ‘learner-driven data’ to advance effective DDL, more research needs to focus on pedagogical applications. Studies in this vein are likely to face certain constraints due to the nature of instructional environments, as in this study where it was not possible to recruit a larger sample size, have an additional control group for a true experimental design, randomly select participants, or have a native-speaker corpus appropriate for direct comparison. Nevertheless, the findings obtained here contribute to the important but surprisingly under-researched area of learner corpus-based pedagogy. Hopefully, despite the realities that may affect methodological choices, this work will incite future CALL endeavors that will take collections of learner language as a useful source of data for much more than “error” analysis, which has been a prominent tendency in learner corpus research. It will be very informative for teaching practice if learner corpus investigations focus on L2 writers of different levels of proficiency, different linguistic features, different implementation conditions, different writing tasks, different learner corpus-based CALL programs, etc. Also, because conclusions from single experiments are difficult to generalize, it is important to establish an agenda of longitudinal research that would help practitioners gain insights into longstanding effects because, as Boulton (2011: 39) remarks, the biggest advantages of learning driven by corpus data lie in longer-term benefits, both linguistic and cognitive/constructivist.

### Acknowledgements

The author is grateful to the anonymous reviewers and the editors of this special issue for their constructive feedback and insightful comments on this manuscript. Gratitude is also due to Dr. Viviana Cortes for pedagogical guidance and for allowing the use of WordSearch.

### References

- Altenberg, B. and Tapper, B. (1998) The use of adverbial connectors in advanced Swedish learners’ written English. In: Granger, S. (ed.), *Learner English on computer*. London: Longman, 80–93.
- Bachman, L. and Palmer, A. (1996) *Language testing in practice*. Oxford: Oxford University Press.
- Barlow, M. (1996) Corpora for theory and practice. *International Journal of Applied Linguistics*, 1(1): 1–37.
- Bartlett, F. (1932) *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.
- Belz, J. and Vyatkina, N. (2005) Learner corpus analysis and the development of L2 pragmatic competence in networked intercultural language study: The case of German modal particles. *The Canadian Modern Language Review/La Revue Canadienne des Langues Vivantes*, 62(1): 17–48.
- Bernardini, S. (2000) Serendipity expanded: Exploring new directions for discovery learning. *TaLC 2000*. Graz, 19–22 July.
- Bernardini, S. (2002) Exploring new directions for discovery learning. In: Kettemann, B. and Marko, G. (eds.), *Teaching and learning by doing corpus analysis*. Amsterdam: Rodopi, 165–181.
- Bernardini, S. (2004) Corpora in the classroom: An overview and some reflections on future development. In: Sinclair, J. (ed.), *How to use corpora in language teaching*. Amsterdam/Philadelphia: John Benjamins, 15–36.
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999) *Longman grammar of spoken and written English*. London: Pearson.
- Blagoeva, R. (2002) Demonstrative reference as a cohesive device in advanced learner writing: A corpus-based study. *Language and Computers*, 49(1): 297–307.
- Bloch, J. (2009) The design of an online concordancing program for teaching about reporting verbs. *Language Learning & Technology*, 13(1): 59–78.

- Boulton, A. (2009) Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, **2**(1): 37–54.
- Boulton, A. (2010) Data-driven learning: Taking the computer out of the equation. *Language Learning*, **60**(3): 534–572.
- Boulton, A. (2011) Language awareness and medium-term benefits of corpus consultation. *New trends in CALL: Working together*. Madrid: Macmillan ELT.
- Braun, S. (2007) Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL*, **19**(3): 307–328.
- Carrió-Pastor, M. L. (2013) A contrastive study of the variation of sentence connectors in academic English. *Journal of English for Academic Purposes*, **12**: 192–202.
- Chambers, A. (2010) What is data-driven learning? In: O’Keeffe, A. and McCarthy, M. (eds.), *The Routledge handbook of corpus linguistics*. London: Routledge, 345–358.
- Charles, M. (2007) Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes*, **6**(4): 289–302.
- Charles, M. (2011a) Adverbials of result: Phraseology and functions in the problem solution pattern. *Journal of English for Academic Purposes*, **10**(1): 47–60.
- Charles, M. (2011b) Corpus evidence for teaching adverbial connectors of contrast: ‘However’, ‘yet’, ‘rather’, ‘instead’ and ‘in contrast’. In: Kübler, N. (ed.), *Corpora, language, teaching, and resources: From theory to practice*. Bern: Peter Lang, 113–132.
- Chen, C. W. (2006) The use of conjunctive adverbials in the academic papers of advanced Taiwanese EFL learners. *International Journal of Corpus Linguistics*, **11**: 113–130.
- Cheng, W. (2012) *Exploring corpus linguistics: Language in action*. London/New York: Routledge.
- Cortes, V. (2007) Exploring genre and corpora in the English for academic writing class. *The ORTESOL Journal*, **25**: 8–14.
- Creswell, J. W. (2003) *Research design: Qualitative, quantitative and mixed methods approaches*. Thousand Oaks, CA: Sage.
- Dörnyei, Z. (2003) *Questionnaires in second language research: Construction, administration, and processing*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Ellis, R. and Barkhuizen, G. (2005) *Analyzing learner language*. Oxford: Oxford University Press.
- Flowerdew, L. (1998) Integrating expert and interlanguage computer corpora findings on causality: Discoveries for teachers and students. *English for Specific Purposes*, **17**(4): 329–345.
- Flowerdew, L. (2001) The exploitation of small learner corpora in EAP materials design. In: Ghadessy, M. and Roseberry, R. (eds.) *Small corpus studies and ELT*. Amsterdam: John Benjamins, 363–379.
- Flowerdew, L. (2012) Exploiting a corpus of business letters from a phraseological, functional perspective. *ReCALL*, **24**(2): 152–168.
- Gilquin, G., Granger, S. and Paquot, M. (2007) Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, **6**(4): 285–374.
- Granger, S. (1996) From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In: Aijmer, K., Altenberg, B. and Johansson, M. (eds.) *Languages in contrast. Text-based cross-linguistic studies*. Lund: Lund University Press, 37–51.
- Granger, S. (2004) Computer learner corpus research: Current status and future prospects. In: Connor, U. and Upton, T. (eds.) *Applied corpus linguistics: A multidimensional perspective*. Amsterdam/Atlanta: Rodopi, 123–145.
- Granger, S. (2009) The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In: Aijmer, K. (ed.) *Corpora and Language Teaching*. Amsterdam/Philadelphia: John Benjamins, 13–32.
- Granger, S., Kraif, O., Ponton, C., Antoniadis, G. and Zampa, V. (2007) Integrating learner corpora and natural language processing: A crucial step towards reconciling technological sophistication and pedagogical effectiveness. *ReCALL*, **19**(3): 252–268.

- Granger, S. and Tyson, S. (1996) Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, **15**: 19–29.
- Green, C., Christopher, C. and Lam, J. (2000) The incidence and effect on coherence of marked themes in interlanguage texts: A corpus-based enquiry. *English for Specific Purposes*, **19**(2): 99–113.
- He, A. (2002) On the discourse marker ‘so’. In: Peters, P., Collins, P. and Smith, A. (eds.), *New frontiers of corpus research*. Amsterdam: Rodopi, 41–52.
- Hegelheimer, V. (2006) Helping ELS writers through a multimodal, computer-based, online grammar resource. *CALICO Journal*, **24**(1): 1–28.
- Hyland, K. (2004) Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*, **13**: 133–151.
- Hyland, K. and Tse, P. (2004) Metadiscourse in academic writing: A reappraisal. *Applied Linguistics*, **25**(2): 156–177.
- Johansson, S. (2009) Some thoughts on corpora and second-language acquisition. In: Aijmer, K. (ed.), *Corpora and language teaching*. Amsterdam: John Benjamins, 33–44.
- Johns, T. (1991) Should you be persuaded: Two examples of data-driven learning materials. *English Language Research Journal*, **4**: 1–16.
- Lei, L. (2012) Linking adverbials in academic writing on applied linguistics by Chinese doctoral students. *Journal of English for Academic Purposes*, **11**: 267–275.
- Meunier, F. (2002) The pedagogical value of native and learner corpora. In: Granger, S., Hung, J. and Petch-Tyson, S. (eds.), *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam/Philadelphia: John Benjamins, 55–76.
- Millar, N. and Lehtinen, B. (2008) DIY local learner corpora: Bridging gaps between theory and practice. *JALT CALL Journal*, **4**(2): 61–72.
- Milton, J. (1998) Exploiting L1 and interlanguage corpora in the design of an electronic language learning and production environment. In: Granger, S. (ed.), *Learner English on computer*. London/New York: Addison Wesley Longman, 186–198.
- Mukherjee, J. and Rohrbach, J.-M. (2006) Rethinking applied corpus linguistics from a language-pedagogical perspective: New departures in learner corpus research. In: Kettemann, B. and Marko, G. (eds.) *Planing, gluing and painting corpora: Inside the applied corpus linguist’s workshop*. Frankfurt: Peter Lang, 205–232.
- Nesselhauf, N. (2004) Learner corpora and their potential for language teaching. In: Sinclair, J. (ed.), *How to use corpora in language teaching*. Amsterdam: John Benjamins, 125–152.
- Osborne, J. (2004) Top-down and bottom-up approaches to corpora in language teaching. In: Connor, U. and Upton, T. (eds.), *Applied corpus linguistics: A multi-dimensional perspective*. Amsterdam: Rodopi, 251–265.
- Pérez-Paredes, P. (2003) Integrating networked learner oral corpora into foreign language instruction. In: Granger, S. and Petch-Tyson, S. (eds.), *Extending the scope of corpus-based research: New applications, new challenges*. Amsterdam: Rodopi, 249–261.
- Pérez-Paredes, P. (2004) Learner oral corpora and network-based language teaching: Scope and foundations. In: Sinclair, J. (ed.), *How to use corpora in language teaching*. Amsterdam: John Benjamins, 249–268.
- Pérez-Paredes, P. and Cantos Gómez, P. (2004) Some lessons students learn: Self-discovery and corpora. In: Aston, G., Bernardini, S. and Stewart, D. (eds.), *Corpora and Language Learners*. Amsterdam: John Benjamins, 247–257.
- Pérez-Paredes, P., Sánchez-Tornel, M. and Alcaraz Calero, J. M. (2012) Learners’ search patterns during corpus-based focus-on-form activities. *International Journal of Corpus Linguistics*, **17**(4): 483–516.
- Pérez-Paredes, P., Sánchez-Tornel, M., Alcaraz Calero, J. and Aguado Jiménez, P. (2011) Tracking learners’ actual uses of corpora: Guided vs non-guided corpus consultation. *Computer Assisted Language Learning*, **24**(3): 233–253.

- Ragan, P. (2001) Classroom use of a systemic functional small learner corpus. In: Ghadessy, M., Henry, A. and Roseberry, R. (eds.), *Small corpus studies and ELT: Theory and Practice*. Amsterdam: John Benjamins, 207–236.
- Rankin, T. and Schiftner, B. (2011) Marginal prepositions in learner English: Applying local corpus data. *International Journal of Corpus Linguistics*, **16**(3): 412–434.
- Rodgers, O., Chambers, A. and Le Baron-Earle, F. (2011) Corpora in the LSP classroom: A learner-centred corpus of French for biotechnologists. *International Journal of Corpus Linguistics*, **16**(3): 391–411.
- Seidlhofer, B. (2002) Pedagogy and local learner corpora: Working with learning-driven data. In: Granger, S., Hung, J. and Petch-Tyson, S. (eds.), *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam/Philadelphia: Benjamins, 213–234.
- Shaw, P. (2009) Linking adverbials in student and professional writing in literary studies: what makes writing mature. In: Charles, M., Pecorari, D. and Hunston, S. (eds.), *Academic writing: At the interface of corpus and discourse*. London: Continuum, 215–235.
- Shei, C. and Pain, H. (2000) An ESL writer's collocational aid. *Computer Assisted Language*, **13**(2): 167–182.
- Tankó, G. (2004) The use of adverbial connectors in Hungarian university students' argumentative essays. In: Sinclair, J. (ed.), *How to use corpora in language teaching*. Amsterdam: John Benjamins, 157–181.
- Yoon, H. and Hirvela, A. (2004) ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, **13**: 257–283.

### Appendix A. Example of LDD intervention activities

#### Result and Inference linking adverbials: **therefore, thus, so, consequently, hence**

These adverbials show that the second unit of the discourse states the result or consequence of the preceding discourse; in other words, they mark the conclusions that the writer has led the reader to make or the conclusion that can be drawn from a previous supporting idea/fact.

Example: Modern improved varieties of rice are unable to attain their full potential yield in the absence of good husbandry and efficient water control. **Thus** the prosperity of the rice industry is largely in the hands of the irrigation engineer.

Example: In 'immigrant' cities, the indigenous political elite remained dominant, while the age-, ethnic- and sex-selective process of immigration tended to overwhelm residential variation in family status. **As a result**, social rank and migration/ethnicity /family status emerge as the major dimensions of residential structure.

#### *Working Learner Corpus*

The analysis of our learner corpus yielded the following counts for words that are used as Result and Inference linking adverbials:

---



---

#### **Result and Inference = 90**

---



---

Therefore – 42

Thus – 10

So – 47

Consequently – 4

Hence – 6

---



---



In the following activities, you will focus on the use of Result and Inference linking adverbials in the English-speaker corpus of published research articles in your discipline and in our learner corpus.

### Task 1:

Read carefully the examples extracted from the English-speaker corpus and from our learner corpus. Compare the use of linking adverbials in the two columns. What similarities and what differences can you see?

<i>English-speaker corpus</i>	<i>Working Learner Corpus</i>
<p>Oligofructose was purer than inulin (75% pure); <b>therefore</b>, doses of inulin were increased by a factor of 1.25 in order to compare similar concentrations of the active ingredients of both fructans.</p>	<p>After authors pointed out a gap in the previous research in Move 3, they are able to say their purpose of the present paper. <b>Therefore</b>, Move3 could be a sort of transition part of introduction.</p>
<p>This indicates a potential for increased amounts of amine production in the presence of high decarboxylase levels and, <b>therefore</b>, a possible contribution to carcinogenesis.</p>	<p>And sometimes there is no Move 3 with no DPR. <b>Therefore</b>, it really depends on the paper and the author's purpose.</p>
<p>Based on the experimental design, it <b>thus</b> cannot be assessed whether energy intake effects on the partitioning of retained energy are best represented based on constraints on body composition (minimum lipid/protein) or on composition of growth (minimumLD/PD).</p>	<p>It is obvious at this point that the author is giving justification to support his findings. <b>Therefore</b>, it could be categorized in this group.</p>
<p>Skeletal muscle is the major site of glucose disposal in the animal, and leptin enhances that utilization. <b>Thus</b>, leptin augments the partitioning of nutrients toward muscle and away from adipose tissue accretion.</p>	<p>Because of the peculiar character of social science, the authors withhold showing their own attitudes, at least superficially. <b>Therefore</b>, they are apt to using tentative verbs such as "suggest" or "state."</p>
<p>It is widely believed that differentiation allows a firm to fully exploit its competitive advantage, and <b>hence</b> to achieve a high level of return.</p>	<p>For example: this author used the technique of Immunocytochemistry here. So he described his specially designed Immunocytochemistry experiment here. And <b>thus</b>, he chose past tense here.</p>
<p>As noted, the two existing subjects, IFP and F&amp;BOps, were selected for PBL treatment because one of the researchers taught these subjects; <b>hence</b>, they were selected for the sake of convenience.</p>	<p>While women always look after their children and take care of their home, they do not have to go out for work except for some cases. <b>Thus</b> women may pay more attention to just learning things that interest them.</p>
<p>A complete description and analysis of all of the comments cannot be presented in this article; <b>hence</b>, only a sample of the comments in each category is presented here.</p>	<p>There are so many beginning of move 4 in that type <b>thus</b> I would like to say most of the authors prefer this type.</p>
<p>That is to say, firms may not need to differentiate themselves from others in order to achieve a high level of performance. <b>Hence</b>, the issue of differentiation or conformity remains to be explored.</p>	<p>There are hundreds thousands of factors to affect human behaviors. <b>So</b>, the amount of relevant previous studies should be massive.</p>
<p>Osborne (1977) found a 1:1 sex ratio in New England, shifting towards a preponderance of females is common in molluscan species generally, increasing with age (Fretter and Graham, 1962). <b>Consequently</b>, in those species that continue to grow throughout their lives the larger size-class contain progressively more females (Coe, 1944).</p>	<p>I could not find "stating a hypothesis" type in several papers. <b>So</b>, it might be said very carefully that this type is not very frequently used in my discipline.</p>
	<p>Some of the instruments may be not precise enough for the experiment, or the background noise is too obvious that it affect the outcome of the experiment. <b>So</b> the authors have to adjust the instruments to eliminate or reduce the noise.</p> <p>Authors try to show that, the topic they researched is widely being researched and <b>hence</b> is important.</p>



Task 2:

A. Use the WordSearch concordancer to search for Result and Inference adverbials in the English-speaker corpus. Provide a few examples of each adverbial that help you better understand how to use it.

B. Use the WordSearch concordancer to search for Result and Inference adverbials in our learner corpus. Select examples where you think the linking adverbials were used inappropriately and try to correct the misuse (include both the original example and your corrected version).

Task 3:

Now that you have compared the Result and Inference linking adverbials in the English-speaker corpus and the learner corpus, write about the patterns you noticed. What recommendations for the use of these linking adverbials can you give? How do you think you personally have been using these words in your writing?

At the end of this class period, upload your written responses to our course Moodle.