DDL INTEGRATION IN DIFFERENT EAP SCENARIOS OF ACADEMIC WRITING

INTEGRACIÓN DE ABD EN DIFERENTES ESCENARIOS DE ESCRITURA ACADÉMICA EN IFA

Alejandro Curado Fuentes
Universidad de Extremadura

Abstract

In EAP (English for Academic Purposes) contexts, a holistic view is desirable for research and pedagogy. This mixed-methods approach usually includes quantitative data from language learning situations. For example, pre-tests, post-tests, and delayed post-tests may be exploited for concrete linguistic aspects so that the learners' performance evolvement with them is measured and contrasted. Other instruments of qualitative observation such as surveys, questionnaires, interviews, and classroom discussion also tend to refine the depiction of the learning profiles and outcomes. In this paper, the goal is to compare different EAP contexts over an 8-year period so that some corpus-related pedagogy issues in EAP may be explored. In particular, the study sets out to compare three different academic scenarios where DDL (Data-Driven Learning) techniques were developed with and for students during the academic writing sessions of the courses. The students were university faculty members (N=20, most with a B1 level), 15 other (graduate) students (mostly B2), and 15 other faculty and graduate students (B2). The DDL method was explored with all three groups, and salient linguistic-discursive features were compared (use of first-person pronouns / awareness of authorship, and importance of active / passive voice in the texts). Furthermore, by using some questionnaires and interviews with the students, additional feedback via their reflections on academic written English was collected. Overall, most students responded positively to the recognition of the learning opportunities offered with DDL for written academic language improvement across and within their fields, although some variations exist within and across groups in terms of learning opportunities and outcomes.

Key Words: EAP, writing, authorship, DDL, corpora

Resumen

En contextos de IFA (Inglés para fines académicos), es deseable aplicar una visión holística a la hora de investigar temas pedagógicos. Este enfoque de métodos múltiples generalmente incluye datos cuantitativos sobre los ítems específicos que se exploren en la situación de aprendizaje. El uso de pre- y post-tests / post-tests posteriores en torno a aspectos lingüísticos concretos puede proporcionar información detallada sobre la adquisición lingüística del alumnado. Otros instrumentos de observación cualitativa tales como encuestas, cuestionarios, entrevistas y debates pueden refinar el dibujo de los perfiles de aprendizaje y resultados del mismo. En este trabajo, el objetivo es comparar diferentes contextos de LFE a lo largo de un periodo de ocho años para que se exploren diferentes temas importantes en relación con la pedagogía de corpus en clases de Inglés para fines académicos (IFA). En concreto, este estudio tiene como base comparar tres escenarios distintos de escritura en IFA donde se desarrollaron técnicas de ABD (Aprendizaje basado en datos) con los alumnos. Los discentes eran miembros del

profesorado universitario (N=20, con nivel B1 de idioma en su mayoría), otros 15 alumnos provenientes de posgrado (con B2 la mayor parte), y otros 15 profesores y alumnos de posgrado (con B2). Se exploró el método de ABD con los tres grupos, y se contrastaron aspectos lingüístico-discursivos concretos (uso de primera persona, concienciación de autoría en el escrito e importancia de la voz activa / pasiva en los textos). Además, con el uso de cuestionarios y entrevistas con los estudiantes, se recogieron diferentes reflexiones en torno a la escritura académica en Inglés. En general, la mayoría de los discentes respondieron de forma positiva al reconocer las oportunidades de aprendizaje ofrecidas para la mejora del Inglés académico escrito en sus respectivas áreas, aunque se registraron variaciones dentro de y entre los grupos en cuanto a esas posibilidades y sus resultados de aprendizaje.

Palabras clave: IFA, escritura, autoría, ABD, corpora

1. INTRODUCTION

Writing in academic contexts of English as a foreign (EFL) or additional language (EAL) has been approached by a myriad of studies from different angles (e.g., corpus-based, social rhetoric, genre analysis, life-history, ethnographic, etc). Such L2 academic communities must often adapt to writing conventions outside their own in L1, especially in the Hard and Social Sciences (cf. Gea-Valor, Rey-Rocha and Moreno 2014), whereas in the Humanities, there is still a certain degree of resistance and skepticism within some disciplines and subject niches (e.g., History, literary studies, linguistics,cf. Burgess 2017). Nonetheless, mainly because of academic recognition and quality criteria pressure from assessment agencies (e.g., ANECA in Spain), the acceptance among higher education communities that they must either publish in English or perish becomes forefront, leading to an abundance of EAP (English for Academic Purposes) and ERPP (English for Research Publication Purposes) materials and courses where language and rhetorical strategies are mainly treated, but not so much "the practices surrounding the route to publication" (Hyland 2016: 190).

In current university settings of Spain, this phenomenon is evident, and one way by which it stands out is the increasing number of faculty members enrolled in EAP / ERPP courses and workshops at tertiary education over the past 8-10 years. All four communicative macro-skills are usually integrated in course / material design, and although our university colleagues tend to emphasize the importance of speaking, writing skills are underlined as intricately needed and fundamental. In general, EAP learners notice the importance of lexical / grammatical issues (e.g., verb tense, voice, article usage, hedging, and so forth) expected for clearer and more concise writing. Parallel to linguistic problems, however, it seems that an appropriate knowledge of text macro-structure and discipline-related rhetorical strategies (i.e., academic and disciplinary literacies) is not seen as important, even if it may work as an advantage prior to paper submission in a given journal (Paltridge 2015). An example is analyzed by Martín and León Pérez (2017) in Medicine, where the importance of rhetorical variation awareness is described for "immunity and allergy" research articles.

In our own courses, over the past decade, and probably going hand-in-hand with and related to higher language levels (from B1 to B2 / C1) over the years, both tenured faculty and fledgling researchers at university tend to recognize the significance of concrete linguistic features (grammatical, mainly) for writing style. In contrast, stylistic pointers such as structural / rhetorical devices along the

158

DDL INTEGRATION IN DIFFERENT EAP SCENARIOS OF ACADEMIC WRITING
Curado Fuentes, A.   *Encuentro 27*, 2018, ISSN 1989-0796, pp. 157-172

text / genre, are, generally, less significantly followed, unless such items are explicitly stated during the reviewing / editing process. This gap in their realization of structures and moves / sub-moves taps into language variation awareness needs across disciplines. For writing tasks in ESP / EAP courses, the integration of Data-driven Learning [DDL], e.g., use of on-line corpora and software tools for writing corrections, can help learners to examine target linguistic / rhetorical issues related to their own specialized contexts of research writing (cf. Tribble 2001; Boulton 2012; Anthony 2016). This type of specialized writing approach via authentic electronic corpora has emerged strong over the past decade as a significant instrument for the construction, analysis, and management of written data from L2 learners either by using existing academic material or having learners build their own corpora of target writing in their fields (cf. Charles 2015; Anthony 2016).

Our study takes place in an EAP type of scenario, where the DDL approach may range from more general academic concordances (e.g., COCA, Corpus of Contemporary American English) to the interaction with corpus data in a given field or genre. The language level (and age) also varies. Over the past decade, our courses have mainly aimed to focus on general academic English skills (EGAP, English for General Academic Purposes). However, during the writing sessions (devoted to general issues such as lexis, grammar, and rhetorical phraseology), a more specific approach is included with the application of DDL after the examination of students' written output via a specific type of task (ESAP, i.e., English for Specific Academic Purposes). These tasks demand their use of identity and authorship stance pointers. The university semesters when these micro-sessions took place were: Fall of 2010, Spring of 2013, and Fall 2017.

This paper thus aims to first describe the common as well as distinctive academic phraseological findings derived from the work with DDL during those three ESAP sessions. Secondly, it attempts to evaluate the students' gained knowledge from DDL by observing pre- and post-tests (via reversed translations of key phraseology) as well as post-task questionnaires and discussions. In the theoretical section below, the literature review examines why this approach is taken in the courses (i.e., DDL for writing and academic skills analysis); then, the Methods will present the instruments for data collection in our case studies. In the Results and discussion, classroom-based observations and data are provided in terms of language awareness and linguistic findings from the students' use of the corpora. In the Conclusions section, we will derive key issues from our context analysis as well as from the current higher-education EAP scenario.

2. THEORETICAL FRAMEWORK

Two main lines of corpus-driven work tend to be applied to EAL / EFL written performance: 1) Learner corpora, where learners' use of linguistic items can be studied for different purposes (e.g., Error analysis), and 2) Native corpora for reference and use (i.e., DDL). Corpus linguistics and language learning seem to have merged naturally over the years due to the authenticity conveyed by corpora either as a reference tool for error / inter-language analysis (e.g., Granger 1998) or as key resources for linguistic exploitation (e.g., Johns 1990). In the second approach, the term DDL, automatically coming to mind, was coined by Johns (1990) and expanded and reformulated by several others (e.g., Sinclair 1991; O'Keeffe, McCarthy and Carter 2007; Boulton 2011).

Corpus-based approaches prove to be significant for the construction, analysis, and management of L2 materials and learners' data (cf. Boulton 2012). This method has been considered and often successfully applied to academic skills over the years, with different interests and focuses as classroom situations have varied (Boulton and Cobb 2017: 350-352). In the case of writing, this approach seems suitable because more positive results tend to be found with this academic skill when students use DDL (Boulton and Cobb 2017: 379), especially with learners' uses of corpora for the correction and / or evaluation of linguistic items (Boulton and Cobb 2017: 380).

Most LSP students realize the potential benefits derived from corpora for specialized writing (Boulton 2012; Anthony 2018), and in the case of advanced EAP learners, the tools seem appropriate for the students' own reflections on their discipline-oriented text types, styles, and forms (Charles 2015). In specialized writing and translation, learners also deem both on-line monolingual and parallel corpora as effective tools for writing for specific purposes (cf. Moon and Oh 2018; Ruivivar and Lapierre 2018). Integrating DDL approaches in the EAP classroom for the benefit of our colleagues and graduate students seems like a convenient strategy to follow. As Mizomuto and Chujo (2016: 62) put it, "guided DDL induction (…) has the potential to be employed more proactively as a teaching methodology with the understanding that (a) it may be better than conventional teaching approaches and (b) it does not favor learners with specific learning styles (i.e., inductive or deductive)."

DDL can thus focus effectively on the connections between words, grammar and discourse (cf. Smart 2014; Mizomuto and Chujo 2015). Among the discursive features regarded as important for academic writing across disciplines, meta-linguistic use, identity and authorship stance stand as objects of analysis (cf. Hyland 2002a). They can be defined within "interpersonal meaning" (cf. Hood 2010; Cheung and Low 2017), a phenomenon by which the writer's voice (i.e., attitude and position) can be conveyed in relation to the reading audience, writing conventions, and / or expectations in that specific discourse community. Hyland (1999) defines this as academic stance, associated to a great extent with the writer's interaction with his/her readers. In this respect, ESAP intricately relates to the specific requirements of, among other aspects, identity usage in the written texts.

In Spain, the importance of academic / professional writing and stance, meta-discourse, and identity is clear by just taking *Ibérica*, the journal of the European Association of Languages for Specific Purposes (AELFE, AsociaciónEuropea de Lenguas para Fines Específicos), as an example to observe that more than a dozen papers have been published over the past decade. The actual perceptions, attitudes, and ideas across different academic communities in Spain as regards writing differences, needs, and problems has been examined (e.g., Gil-Salom and Soler-Monreal 2014; Gea-valor, Rey-Rocha and Moreno 2014). Remedial suggestions include not only the attention to key verb tenses, hedges, and article usage in the texts / subtexts, ever important in our EAP writing courses, but also working with explicit discourse and genre features in L2 to which learners may not entirely relate or be familiar with (e.g., using first person pronouns across research sections, cf. Sheldon 2009).

An on-going pursuit in the DDL integration within academic writing courses is the notion of disciplinary / sub-disciplinary focuses. There are not only critical content-based differences among

160

DDL INTEGRATION IN DIFFERENT EAP SCENARIOS OF ACADEMIC WRITING
Curado Fuentes, A.   *Encuentro 27*, 2018, ISSN 1989-0796, pp. 157-172

students (e.g., non-Philology versus Linguistics-knowledgeable students in the classroom) but also writing competence level variations (e.g., academic stance is not usually embedded yet in younger researchers, even if their language levels are high). Even if the literature tells us that DDL can adapt to different learning styles (cf. Mizomuto and Chujo 2016; Boulton 2009), the issue here is whether EAP students can effectively integrate their roles as "detectives of language use" (cf. Johns 1990) in order to become effective ESAP practitioners, communicating for specific academic purposes (e.g., chart explanations, cf. Sancho-Guinda 2012) or within a specific genre (e.g., argumentative essays, cf. Schneer 2013).

Another important issue is the notion of L1 transfer in academic literacies. L2 students may concentrate on certain lexical bundles and un-notice or underestimate others, in agreement with Pérez-Llantada's (2014: 88) observation that "few non-native learners ever fully accumulate the native repertoire of formulaic sequences". Yet, a good deal of L1 and L2 convergence exists, and when insufficient, the repertoire of lexical bundles and resources can result in an unsuccessful alignment or ''frustrated proximity'' (Sancho-Guinda 2012: 166). This reinforcement need for lexical / grammatical competence tends to recur as one of the main issues related to L2 academic writing.

3. METHODS

The three cases described in this study are examples of engagement with DDL for written academic language revision / improvement. The type of analysis presented is concrete and qualitative, as it was the only common approach in all three groups. All the EAP students dealt with academic evaluation in their writing when using personal and meta-textual references (i.e., expressing authorship and stance). The three groups of students were: 1) 20 students (mostly faculty at B1 levels) taking four hours of academic writing (in an intensive 20-hour EAP course) during the Fall semester of 2010. 2) 15 graduate students (from Philology and Education degrees) being exposed to 30 hours of corpus work in a 60-hour course for the Master of Education (most students at the B2 level) during the Spring semester of 2013, and 3) 15 (mostly graduate) students but also faculty at B2 levels taking 2 hours of DDL within a 12-hour EAP module (Fall of 2017).

The corpus tools used varied from course to course. Since more corpus interfaces, big data, and faster internet access have made on-line corpora easier and probably friendlier to use, our courses tend to have also integrated them more effectively in the classroom. Smaller corpora are also equally accessible, and better software tools enable the building of ad-hoc resources for specialized text type classifications and analyses (cf. Heuboeck, Holmes and Nesi 2010).

The corpora used in the courses were: 1) For 2010, the Hong Kong Polytechnic Corpus of Engineering (HKPCE) (more faculty came from Science) and a home-made corpus of academic blogs compiled by the teacher in 2009 and processed via WordSmith Tools (2004). 2) For 2013, many different corpus resources were used, but for the evaluation of authorial identity in the texts, three on-line resources were important: MICUSP (The Michigan Corpus of Upper-level Student Papers), the academic section of COCA (Corpus of Contemporary American English), and BAWE (British

Academic Written English corpus). 3) In 2017, BAWE and COCA were selected to showcase some academic language usage of verbs and nouns from the students' compositions.

The writing tasks assigned to the students were the following:

1. Stating personal opinion (arguing for or against a given topic): Years 2010 and 2013.
2. Describing a favorite hobby / interest: Years 2010 and 2017.
3. Describing a favorite teaching method: 2013.

In the following section, Results and Discussion, the precise instructions and details of the tasks will be described, with the observation of specific linguistic devices that may have gone either unnoticed or overused in the L2 writings. For the evaluation of their potential improvement with the recognition and production of such items, after the DDL sessions, in 2010 and 2013, a pre-test was handed out to the students, followed by a post-test on the same content on the following day (2010), and further apart in 2013 (more than two weeks). In 2010, the teacher conducted the DDL revision between tests by showing the data on-line on the screen, while in 2013, the students themselves made note of the possible linguistic improvements by using the resources available on-line. In 2017, there was no evaluation, as corpus use was restricted to illustrating some linguistic items that those volunteering to read their texts unveiled (the teacher took note of some verbs and nouns and compared them with corpus-driven examples).

3.1 Data collection

How the data was collected varied across the groups. In 2010, the students submitted their written texts in paper form and then, the possible changes to be made were underlined by the teacher and returned to them on the following day. The pre- and post-tests in 2010 included a list of frequent academic expressions to be translated. In 2013, the same thing was done as in 2010, but a questionnaire on students' impressions and reactions to corpora was also included at the end of the sessions. In 2017, as mentioned above, there was no data collection. In all three cases, discussions and some interviews with the students were made via classroom talks. In addition, institutional on-line course questionnaires were obligatory for students to complete in 2010 and 2017.

The data collection derived from this study was compared under the hypothesis that "an increase in the accuracy with which partially acquired features are used" is actually made after DDL work (Ellis 2010: 344). The writing assignments were timed and constrained within classroom situations for the purpose of language revision via DDL. There was no post-writing task as a follow-up exercise, nor other mechanisms for progress surveying. Therefore, the increase in accuracy to be observed, if any, would be relegated to specific academic phraseology evaluation by means of reversed translation (see Appendix, Table 1). Thus, academic writing proficiency would not be an objective in this evaluation (as mentioned, writing post-tasks were not administered to any of the groups). Yet, examining their improvement in the translations might lead to, at least partially, observations that DDL may effectively foster intake for written academic language usage.

162

DDL INTEGRATION IN DIFFERENT EAP SCENARIOS OF ACADEMIC WRITING
Curado Fuentes, A.   *Encuentro 27*, 2018, ISSN 1989-0796, pp. 157-172

Furthermore, academic expressions can be influenced by the L1-related language (cf. Sancho-Guinda 2012). Reversed translation may become an important instrument for linguistic revision / reinforcement. One way of redressing such deficiencies may be found in the explicit marking of L1 and L2 correspondences not only via existing academic parallel corpora, but also home-made collections of academic writing in L1. Figure 1 presents the most frequent alignments from an ad-hoc corpus of English and Spanish journals in the fields of Computer Science and Education (cf. Curado-Fuentes 2010). These expressions were used as linguistic reference for the translation exercise in the pre- and post-tests. The labels "English corpus ok" and "Spanish corpus ok" refer to the existence of such items as highly frequent in the two corpora, while the "English corpus no" means that those items were not frequent in the English journals, and yet overused in an L2 corpus.

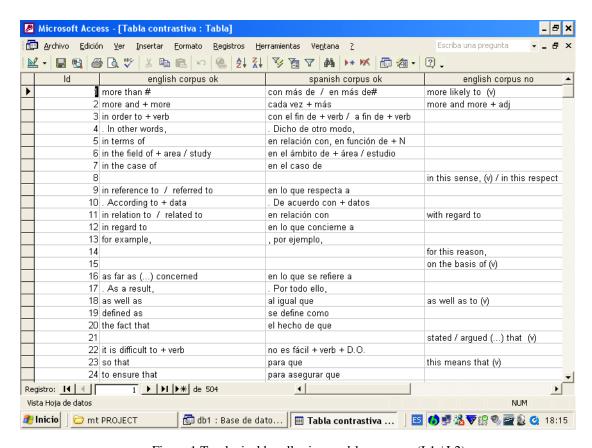| Id	english corpus ok	spanish corpus ok	english corpus no
1	more than #	con más de / en más de#	more likely to (v)
2	more and + more	cada vez + más	more and more + adj
3	in order to + verb	con el fin de + verb / a fin de + verb	
4	. In other words,	. Dicho de otro modo,	
5	in terms of	en relación con, en función de + N	
6	in the field of + area / study	en el ámbito de + área / estudio	
7	in the case of	en el caso de	
8			in this sense, (v) / in this respect
9	in reference to / referred to	en lo que respecta a	
10	. According to + data	. De acuerdo con + datos	
11	in relation to / related to	en relación con	with regard to
12	in regard to	en lo que concierne a	
13	for example,	, por ejemplo,	
14			for this reason,
15			on the basis of (v)
16	as far as (...) concerned	en lo que se refiere a	
17	. As a result,	. Por todo ello,	
18	as well as	al igual que	as well as to (v)
19	defined as	se define como	
20	the fact that	el hecho de que	
21			stated / argued (...) that (v)
22	it is difficult to + verb	no es fácil + verb + D.O.	
23	so that	para que	this means that (v)
24	to ensure that	para asegurar que	

Figure 1.Top lexical bundles in an ad-hoc corpus (L1 / L2).

Finally, the type of writing explored in our courses was not research-based, as above-mentioned, but chiefly dealt with academic writing related to description and argumentation. No discourse structure model regarding any of the text types treated was considered (e.g., Hyland, 2004; Schneer, 2013), although it may be observed that various texts (especially argumentative essays in 2013) followed the thesis statement / argument / conclusion form. In all three cases, for the evaluation of their writing, three linguistic elements were considered: 1) Using personal pronouns (mainly first person); 2) Using impersonal statements (mainly *it is* + adjectival / participle forms); 3) Using modality in the passive versus active voice. All these elements could point at how identity formulation and reflection might be made in the texts (cf. Hyland, 1994; Hyland, 2002a).

4. RESULTS AND DISCUSSION

4.1 The 2010 group

After the essays were returned to the students in class, we focused on the use of the first person pronoun, passive versus active voice, and impersonal statements. In contrast with other groups (see Table 1), these students' texts presented simpler constructions for argumentation (e.g., *I think that*, *I believe that*, *I am in favor of…*). Active statements outnumbered passive constructions by far (an average of one passive sentence per 300 words). Impersonal statements with the use of *it is* + adjective / participle were a bit more noticeable in argumentation, but not to describe interests or hobbies (e.g., *it is possible that*, *it is good that*, *it is considered that…*).

Year	Students' overused constructions	Corpus-informed constructions noticed / exploited by the students
2010	*I think that* / *I believe that* *It is possible that* *Is possible* + subject	*I* + many verbs (state / others) (no *that*) *We* + involving verbs (e.g., take / see) *It* / *This* +could / may / can + *be* + past participle
2013	*I think* / *I prefer* / *I would choose* *This can happen* / *This can be* + adj. *If you take* / *consider…*	*I* + many verbs (other verb tenses / engagement with the reader) *We* / *I* + modal + *argue* (active) *It* + modal + *be* (passive statements) *It is* + possible / important + *to…*
2017	*I find* + concept + adjective *I predict that…* *It is* + adjective + *that*	*it* + *can* / *may* + *be* + past participle + *that*

Table 1. Contrastive examples from the linguistic aspects examined in the courses

The information contrasted with the on-line HKPCE revealed that no use was made of the first person singular pronoun in HKPCE, as the texts in this corpus were research papers and written projects. In contrast, the students realized how the use of *we* could be made in active statements throughout the corpus. Less clear was their realization of how some verbs, usually activated with personal pronouns in their minds (e.g., *observe*, *predict*, *consider*), could be re-written in the passive voice with modals (e.g., *it could be predicted that…*) to confer a more detached academic stance (perhaps, this awareness difficulty was due to their lower language levels, the fact that all the students were non-Humanities / non-Linguistics experts, or both factors).

In turn, from the corpus of blogs, a wealth of choices was induced with the use of *I* with other verbs. A common structure, unused by the students, was the omission of *that* in *I believe / think* clauses. In addition, the possible use of *we* to involve the reader was shown and understood as a form of interpersonal meaning with the community of readers (a strategy that is also common in academic Spanish). Finally, they could double-check in their compositions that impersonal passive statements

might be preferred "for better English", as some learners claimed in the classroom discussions. Some structures reviewed with these corpora included some items from the pre- and post-tests (e.g., *this could be defined as…*).

For the pre- and post-tests, the lexical content (see Table 2 in the Appendix) was derived from a corpus, but none used in the courses. As mentioned above, this corpus was a home-made collection (239,000 tokens) of academic papers from the areas of Computer Science and Education (cf. Curado-Fuentes 2010). The aim of this corpus was the provision of highly frequent academic expressions. The 20 items were the most frequent academic lexical bundles found in that corpus, coinciding in part with other academic writing corpora (e.g., BAWE and Academic Word List by Coxhead2000). The translation to be made was reversed translation (Spanish to English), but four items were in English (deemed as more difficult for this B1 level) for direct translation in the tests.

The calculation of the scores was made with a non-parametric measurement (Sign test calculator 2018), since the number of students was below 30 and a kurtosis test in Excel revealed negative scores (-0.7 and -0.4) deviating from 1.0. The z-value from the pre- and post-test score comparison yielded 4.47214, the *p*-value $<.00001$, and the result was significant at $p<0.05$. Therefore, an obvious lexical recognition and gain could be deduced from the score comparisons, at least in the short term, since the post-test followed the DDL treatment and pre-tests within a very short time (one day later). In post-course questionnaires, some students referred to this particular DDL activity as positive, albeit more requests pointed to other needs in the courses (e.g., more hours for writing, speaking).

4.2 The 2013 group

The 15 students in the graduate course of the Master of Education wrote longer compositions as an average (a mean of 423 words) and submitted their writings on the virtual university platform. This facility (together with more time afforded for the activities) allowed for their own management of the corpus resources and tools. They also had access to concordance software (WordSmith) with which to examine their own word usage and frequencies. Overall, they could check how, in their descriptions of favorite teaching methods, their own use of first person pronouns was frequent with opinion and state verbs (e.g., *I think that*, *I prefer*, *I will choose…*). However, they made almost no use of the plural form *we*, and only some use of the second person *you* as an impersonal statement for activity description (e.g., *if you take…*). The most common verb tense used with pronouns was the present simple, followed by the future.

Impersonal statements were not frequent, but appeared in an average rate of 1.25 per composition. The passive voice was used in a 2.2-to-15 sentence ratio, but rarely used with modality (only three statements in total in arguments with *can* and *may*). Impersonal clauses were chiefly integrated in the explanation of concepts and observations of teaching practices. Most modal verbs appeared in the active voice in the compositions (e.g., *this can happen*, *this can / may be…*).

When we used on-line corpora for language revision, MICUSP provided a wide range of verbs and tenses as options with first person pronouns (e.g., *I* + continuous forms in the future and past, and *we* + modals engaging the reader). The students made notes of these differences and submitted them as a final exercise for the activities. The results were quite satisfactory as a whole. One student even wrote her final project for the course on linguistic differences noticed and exploited with the on-line tools. Here is an excerpt of the conclusions in her project:

> *I consider that the corpus instruments that we have used are very positive for our language learning competence. The other tools that improved this learning approach can include the use of translation and vocabulary applications, but the best for positive reinforcement of language use is the corpus tool, undoubtedly.*

The use of BAWE also led to important discoveries of academic verb use. The case of frequent verbs used with modality and in passive statements in argumentative texts was realized. For example, with the verb *argue*, 88 percent of its use was in the form *it* + modal + *be argued*). Another issue was the different choices of verb tenses available. COCA unveiled this for our graduate students: the future tense with personal pronouns in essays, where up to 50 percent of the items occurred in the future continuous tense (e.g., *I will be arguing / I will be referring to…*). Regarding impersonal statements, our students were able to observe some of their linguistic choices reflected in COCA, but important differences emerged in terms of collocations and structure usage (e.g., *it is possible* was followed by *to* more frequently than by *that*, and the use of *that* was often omitted in the corpus).

The scores from pre- and post-tests were noticeably higher than in the 2010 group, but the number of students (N=15) made the application of parametric statistics dubious (with kurtosis values for their scores of -0.2 and -0.5). The sign test was performed, and the z-value yielded 3.87298, the *p*-value <.00011, and the result was significant at $p<0.05$.

The production of lexical bundles in their writing was not measured in a longitudinal study that might have added more reference data. Yet, all the learners' improved test results correlated with their recognition of corpus work as a good approach to academic language gain (in agreement with Curado-Fuentes 2014). In Table 3 (Appendix), the high means resulting from scores on a likert scale (1-5), especially the first five items, may point to such positive reactions to corpora for L2 learning among students. In contrast, Table 3 (Appendix) identifies lower means for negative statements on corpus use. From classroom discussions, some positive and negative comments were:

> *MICUSP and BAWE are so cool, they give you the actual use in academic writing, and you can see in which areas of study.*
> *I prefer to use the concordances if we choose fewer texts, because they clarify things better.*
> *The best thing is that I learned to use them for my own courses of English writing. I can now check my vocabulary in context.*
> *COCA fails too much. I lose the information and I have to start again.*
> *I don't like these electronic tools. They are confusing. Grammar is not explained well.*
> *I just don't like computers, and less for language learning.*

4.3 The 2017 group

Some examples in this group were related to usage of frequent academic verbs in the compositions. A salient contrast with the corpus information was the use of such verbs in passive statements with modality. Such constructions tend to be underused by L2 learners, and in our case, the students felt that they were important constructions to know. For instance, the verbs *find*, *predict*, *generalize*, and *observe* could be often located in BAWE and COCA within the impersonal statement pattern *it* + modal + *be* + past participle.

A different but also interesting discussion was motivated with the use of COCA for the examination of CAN and MAY, the two most frequent modals occurring in the passive voice in the academic section of the corpus. Most students could see the differences in stance when using this construction, e.g., they thought that epistemic and evaluative purposes were achieved better (i.e., distancing one's voice from the data to substantiate the evidence). A student reported that even though he felt that first person pronouns in L2 research writing make him feel "uneasy" and "uncertain", he could not really "muster enough courage" to widen the scope with different structures and be equally clear at certain points in the text, especially at discussing results. This observation parallels Hyland's (2002b: 1108) in that

> academic literacy is a 'foreign culture' to students of all backgrounds, where they find their previous understandings of the world challenged, their old confidences questioned, and their ways of talking modified. For students struggling to gain control of their discipline and master its content, this can lead to a sense of powerlessness and uncertainty.

This type of *it* + *can /may*+ passive constructions was regarded as "insipid" and "less direct" by some students, who wondered "why not construct more active personal references" in the research process narrative. In Spanish, in fact, a frequent parallel device is the inclusive (author + readership) *we* + modal + infinitive, in accordance, at least partially, with Sheldon's (2009: 254) "typological variation of six possible identities realized by first-person pronouns". Our graduate learners were somehow referring, in their classroom talks, to their own writing practices in the L1 and interlanguages.

4. CONCLUDING REMARKS

Some ideas about DDL integration in EAP courses may be discussed, based on our own findings and other authors' results:

1.   A higher language command and better and more widespread use of technologies may play in our favor. As demonstrated by the 2017 group, more faculty and graduate students tend to have a B2 level nowadays. This variable is important because the students at B2 were able to more easily reflect on the corpus-driven information, connecting linguistic items to authorial stance. The 2013 group understood that native-like writing could serve as a good example for their examination of language choices, and this comparative knowledge could benefit their own academic writing. In 2010 and 2017, all the students came from areas in Hard and Social Sciences. Thus, no bias could be alleged in 2017 in

167

DDL INTEGRATION IN DIFFERENT EAP SCENARIOS OF ACADEMIC WRITING
Curado Fuentes, A.   *Encuentro 27*, 2018, ISSN 1989-0796, pp. 157-172

terms of their being language-focused learners. Easier corpora interfaces and clearer / more abundant results also favored the 2013 group's understanding and identification of data for their academic aims.

The question is that those students with lower language levels may not become attracted to the DDL resources as the more advanced learners do. DDL seems to work better if there is a language background that underscores lexico-grammatical items. Thus, offered to the academic community are simplified interfaces and concordance displays—cf. Chujo et al. 2015—but for EAP, when those students are starting at A2 or B1, in the case of making the connections between lexis and grammar in specialized discourses, much more attention and effort is needed. Such efforts may not only be devoted to language, but also disciplines / sub-disciplines.

2.   The specific text type. EAP courses usually include the study of academic genres and sub-genres. An example is the research paper introduction. However, for the writing assignments described, other (less structurally rigid) types were chosen so that the students could be given more freedom in their writing style. Thus, a text-based variation of linguistic choices (e.g., usage of I-pronouns in argumentative essays) could be explored in the DDL activities. The wide array of linguistic items that emerged in the DDL sessions belonged to the so-called lexico-grammatical aspect of language learning (e.g., a key verb is frequently used in a given tense and voice within a given text type). This correlation was less easily comprehended by most learners, and only in 2013, and partially, the graduate students coming from Philology and Education majors could really make the connections and reproduce such insights onto their own dissertation papers.

The exploration of Philology versus non-Philology students' reactions to DDL for academic writing is also key. Perhaps, one way of examining such variations (if there are any) is the approach to more current and updated resources where the learners may make the connections better. Indeed, it seems that corpora such as HKCPE have become outdated, and others like BAWE and COCA may not disclose the association of crucial linguistic items within specific text types (as specialized ad-hoc corpora may achieve). More recent discipline-based corpora (e.g., the Elsevier corpus of research—Kwary, 2018) may be a good alternative.

3.   The "detective" work. Except for the 2017 group (who did not undergo any direct evaluations), the case studies seem to demonstrate that both inductive and deductive strategies are well developed to correct or clarify lexico-grammatical items (e.g., searching for alternative verbs within a given structure, or the opposite, identifying a grammatical pattern and then the verbs in it). For the discovery of more optional linguistic items and frequent native-like expressions, the students have always indeed valued the tools as positive (explicitly in 2013 with the questionnaires, but also implicitly in discussions and after-the-course surveys in 2010 and 2017).

It seems that more research is needed in the form of longitudinal studies that focus on post-evaluation over a one-year course or, even more, over a whole study program / degree, and / or professional career. For example, an optimal scenario in bilingual programs could be that the faculty /

168

DDL INTEGRATION IN DIFFERENT EAP SCENARIOS OF ACADEMIC WRITING
Curado Fuentes, A.   *Encuentro 27*, 2018, ISSN 1989-0796, pp. 157-172

instructors' academic competences in L2 be evaluated every x number of months for an x-year period. The long-time effect of DDL on their linguistic progress may thus be effectively assessed.

REFERENCES

Anthony, L. 2016. «Introducing corpora and corpus tools into the technical writing classroom through Data-Driven Learning (DDL)». In J. Flowerdew & T. Costley (Eds.), *Discipline Specific Writing: From Theory to Practice* (pp. 163-181). Abingdon, UK: Routledge.

Anthony, L. 2018. *Introducing English for Specific Purposes*.London: Routledge. http://www.routledge.com/9781138936652.

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. 2011.«Data-driven learning: The perpetual enigma». In S. Goźdź-Roszkowski (Ed.), *Explorations across languages and corpora* (pp. 563–580). Frankfurt, Germany: Peter Lang.

Boulton, A. 2012.«Corpus consultation for ESP: A review of empirical research». In A. Boulton, S. Carter-Thomas, & E. Rowley-Jolivet (Eds.), *Corpus-informed Research and Learning in ESP: Issues and Applications* (pp. 261–291). Amsterdam: John Benjamins.

Boulton, A., and Cobb, T. 2017. «Corpus use in language learning: A meta-analysis». *Language Learning. A Journal of Research in Language Studies,* 67/2, 348-393. DOI: 10.1111/lang.12224.

Brtitish Academic Written English Corpus (BAWE). Retrieved from https://www.sketchengine.eu/bawe-corpus/

Burgess, S. 2017. «Accept or contest: A life-history study of humanities scholars' responses to research evaluation policies in Spain». In M. Cargill & S. Burgess (Eds), *Publishing Research in English as an AdditionalLanguage: Practices, pathways and potentials* (pp. 13-33). Adelaide: University of Adelaide Press. DOI: https:/ / doi.org/ 10.20851 / english-pathways.

Charles, M. 2015. **«**Same task, different corpus: The role of personal corpora in EAP classes**»**. In A. Leńko-Szymańska& A. Boulton (Eds.), *Multiple Affordances of Language Corpora for Data-driven Learning* (pp. 129-154). Amsterdam: John Benjamins.

Cheung, Y. L., and Low, T. H. 2017.«Pre-university students' voice construction in argumentative essays».*RELC Journal*, 1-16. DOI: 10.1177/0033688217716508.

Chi-Square Test Calculator. 2018, November 18. Retrieved from http://www.socscistatistics.com/tests/chisquare2/Default2.aspx.

Chujo, K., Oghigian, K., and Akasegawa, S. 2015. **«**A corpus and grammatical browsing system for remedial EFL learners**»**.In A. Leńko-Szymańska& A. Boulton (Eds.), *Multiple Affordances of Language Corpora for Data-driven Learning* (pp. 109-128). Amsterdam: John Benjamins.

Coxhead, A. 2000.«An Academic word List».*ELI Occasional Publications* #18. School of Linguisticsand Applied Language Studies, Victoria University of Wellington.

Curado-Fuentes, A. 2010.«A Comparison of field and L1 influence on non-native writers».In *EuropäischenGesellschaftfürFachsprache*(pp. 48-53). Hamburg: University of Hamburg. https://www.amazon.de/Internationalen-Europäischen-Fachsprachen-Linguistische-Kommunikation/dp/toc/392722653X.

Curado-Fuentes, A. 2014.«Learning the integration of DDL in Secondary Education EFL teaching: The importance of ESP».In A. Floyd (Ed.).*A Roadmap for Languages for Specific Purposes: Multilingualism and ICT's* (pp. 135-140). A Coruña: Galebooks.

Davies, Mark. 2008-. *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present*. Available online at https://corpus.byu.edu/coca/.

Ellis, R. 2010. «A framework for investigating oral and written corrective feedback».*Studies in Second Language Acquisition,* 32, 249-335.

Gea-Valor, M.L., Rey-Rocha, J, and Moreno, A.I. 2014.«Publishing research in the international context: An analysis of Spanish scholars' academic writing needs in the social sciences». *English for Specific Purposes*, 36, 47–59.

Gil-Salom, L., and Soler-Monreal, C. 2014.«Writers' positioning in literature reviews in English and Spanish computing doctoral theses».*Journal of English for Academic Purposes,* 16, 23-39.

Granger, S. (Ed.) 1998. *Learner English on Computer*. London and New York: Addison Wesley Longman.

Heuboeck, A., Holmes, J, and Nesi, H. 2010. *The BAWE Corpus Manual FOR THE PROJECT ENTITLED 'An Investigation of Genres of Assessed Writing in British Higher Education'*. Funded by the ESRC [project number RES-000-23-0800] Version III.

Hong Kong Polytechnic Corpus of Engineering. Retrieved from http://rcpce.engl.polyu.edu.hk/HKEC/

Hood, S. 2010. *Appraising Research Evaluation in Academic Writing*. New York: Palgrave MacMillan.

Hyland, K. 1994. «Hedging in academic writing and EAP textbooks».*English for Specific Purposes,* 13, 239–256.

Hyland, K. 1999. «Disciplinary discourses: Writer stance in research articles». In C.N. Candlin& K. Hyland (Eds.), *Writing: Texts, Processes and Practices* (pp. 99-122). London: Routledge.

Hyland, K. 2002a. «Options of identity in academic writing»..*ELT Journal,* 56, 351–358.

Hyland, K. 2002b. «Authority and invisibility: Authorial identify in academic writing». *Journal of Pragmatics,* 34, 1091–1112.

Hyland, K. 2004. *Genre and Second Language Writing*. Ann Harbor: The University of Michigan Press.

Hyland, K. 2016. *Academic Publishing: Issues and Challenges in the Construction of Knowledge*.Oxford: Oxford University Press.

Ibérica.*The Journal of the European Association of Languages for Specific Purposes.*Edited by C. Sancho Guindo. Retrieved from http://www.aelfe.org/?s=presentacio

Johns, T. 1990. «From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning». *CALL Austria,* 10, 14–34.

Kwary, D. A. 2018. «A corpus and a concordancer of academic journal articles».*Data in Brief,* 16, 94-100. DOI: doi.org/10.1016/j.dib.2017.11.023.

Martín, P., and León-Pérez, I.K. 2017. «Disseminating research internationally: Intra-subdisciplinary rhetorical structure variation in immunity and allergy research articles».In M. Cargill & S. Burgess (Eds), *Publishing Research in English as an AdditionalLanguage: Practices, pathways and potentials* (pp. 151-169). Adelaide: University of Adelaide Press. DOI: https:/ / doi.org/ 10.20851 / english-pathways.

Michigan Corpus of Upper Llevel Student Papers. 2009. Ann Arbor, MI: The Regents of the University of Michigan

Mizumoto, A., and Chujo, K. 2015. «A meta-analysis of data-driven learning approach in the Japanese EFL classroom».*English Corpus Studies,* 22, 1–18.

Mizumoto, A., and Chujo, K. 2016. «Who is data-driven learning for? Challenging the monolithic view of its relationship with learning styles».*System*, 61, 55-64. http://dx.doi.org/10.1016/j.system.2016.07.010.

Moon, S., and Oh, S.Y. 2018. «Unlearning overgeneratedbe through data-driven learning in the secondary EFL classroom». *ReCALL*, 30/1, 48-67. https://doi.org/10.1017/S0958344017000246.

O'Keeffe, A., McCarthy, M., and Carter, R. 2007.*From Corpus to Classroom*. Cambridge: Cambridge University Press.

Paltridge, B. 2015.«Referees' comments on submissions to peer-reviewed journals: When is a suggestion not a suggestion?»*Studies in Higher Education*, 40/1, 106-122.

Pérez-Llantada, C. 2014. «Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage». *Journal of English for Academic Purposes*, 14, 84–94.

Ruivivar. J., and Lapierre, C. 2018. «Learning outcomes and learners' impressions of parallel and monolingual concordancers». In P. Taalas, J. Jalkanen, I. Bradley & S. Thouesary (Eds.), *Future-proof CALL: Language Learning as Explorations and Encounters – Short Papers from EuroCALL 2018* (pp. 278-283). Research-publishing.net. DOI: https://doi.org/10.14705/rpnet.2018.26.850.

Sancho-Guinda, C. 2012.«Proximal Positioning in Students' Graph Commentaries».In K. Hyland & C. Sancho-Guinda (Eds.), *Stance and Voice in Written Academic Genres* (pp. 166-183). London: Palgrave Macmillan.

Schneer, D. 2013. «Rethinking the argumentative essay».*TESOL Journal,* 5/4, 619-653.

Sheldon, E. 2009.«From one I to another: Discursive construction of self-representation in English and Castilian Spanish research articles». *English for Specific Purposes,* 28, 251-265.

Sinclair, J. 1991. *Corpus, Concordance, Collocation.*Oxford, UK: Oxford University Press.

Tribble, C. 2001. «Small corpora and Teaching writing; towards a corpus-informed pedagogy of writing».In M. Ghadessy, A. Henry & R. L. Roseberry (Eds.), *Small Corpus Studies and ELT: Theory and Practice* (pp. 381-401). Amsterdam: John Benjamins

WordSmith Tools. 2004. [Software]. Available from http://www.lexically.net/wordsmith/downloads/

APPENDIX

Translate the following into English (or Spanish):

Más de X
Cada vez más
Con el fin de + verbo
Dicho de otro modo,
In terms of + concept
En el ámbito de + área de estudio
En el caso de
En lo que respecta a
De acuerdo con los datos
En relación con
En lo concerniente a
Por todo ello,
As far as X is concerned…
As well as...
Se podría definir como
El hecho de que
Es difícil + verbo
This is found to be…
Parece que
Se observa que

Table 2. Academic lexical bundles / items for pre- and post-tests.

171

DDL INTEGRATION IN DIFFERENT EAP SCENARIOS OF ACADEMIC WRITING
Curado Fuentes, A.   *Encuentro 27,* 2018, ISSN 1989-0796, pp. 157-172

Survey Questions	Mean
11. Using corpora for language exploration is a good way to learn more about that language	4,4889
2. Learning how to use corpora tools is important for my career as teacher	4,4444
10. I would like to use corpora in the future and learn more about their use	4,2955
8. Corpora are good to improve my writing knowledge	4,2444
9. Corpora are good for error correction	4,1778
7. Google can also work as a positive corpus tool	4,1111
1. I enjoy using the computer for L2 improvement	4,0889
4. Using corpora gives me more chances to approach authentic English	4,0667
14. Using small corpora is better for academic purposes	4,0444
5. Free on-line corpora are more convenient and useful	4,0000

Table 2. Items from the 2013 group's questionnaire ranked in order of means (high values)

Survey Questions	Mean
3) I am intimidated by computers	2,0889
6) Corpora confuse me and I don't see the point	2,3556
13) Using corpora is not worth the time and effort.	2,6136
12) Corpora are frustrating to work with.	2,8444

Table 3.Items from the 2013 group's questionnaire ranked in order of means (low values)