Journal of Academic Language & Learning
Vol. 16, No. 1, 2022, 91-104. ISSN 1835-5196

Introducing data-driven learning on a PhD pre-sessional programme

Dana Therova and Andrew Mckay
Durham Centre for Academic Development, Durham University, Durham, UK
Email: dana.therova@durham.ac.uk and andrew.mckay@durham.ac.uk
(Received 30 January, 2022. Published online 25 September, 2022.)

The aim of this article is to investigate the feasibility and benefits of a data-driven learning (DDL) approach in the context of an English for Academic Purposes (EAP) pre-sessional programme at a British university. It shows how self-compiled corpora can be usefully implemented on a pre-sessional programme to cater for the academic language development needs of international students in a multidisciplinary classroom. The participants were three international students preparing for doctoral study by completing a ten-week online PhD pre-sessional programme, during which they utilised corpus software in regular weekly sessions to expand their discipline-specific academic language knowledge. The usefulness and value of this approach was investigated through semi-structured interviews on completion of the programme, complemented by tutor reflection. The results show that self-compiled corpora can provide a valuable resource for writing in multidisciplinary EAP contexts. This paper thus suggests that a DDL approach could be usefully implemented in PhD pre-sessional programmes, as well as in mainstream EAP classes.

Key Words: English for Academic Purposes (EAP), data-driven learning (DDL), do-it-yourself (DIY) corpora, international students, pre-sessional programme.

1. Introduction

In recent years, the use of corpora (i.e. large collections of electronic texts) to support data-driven learning (DDL) has gained popularity in English for Academic Purposes (EAP) contexts, particularly for the teaching of various aspects of academic writing. The term DDL, first introduced by Tim Johns (e.g., 1986, 1991), refers to an approach to learning in which learners act as researchers drawing on a large amount of language data supplied by a corpus. The main characteristic of this approach is "the attempt to cut out the middleman as far as possible and to give direct access to the data" (Johns, 2012, p. 297). DDL is therefore largely an inductive approach based on the assumption that effective language learning is a form of linguistic research in which learners are brought closer to authentic language use through their own discovery (Johns, 2012).

This approach thus replaces instruction with discovery learning as learners carry out self-directed language-learning exercises through interaction with a corpus which provides an open-ended supply of language data tailored to the specific learners' needs through a learner-centred approach (Anthony, 2019; Hyland, 2006). This makes DDL a valuable approach to teaching writing in EAP contexts catering for students from a wide range of disciplines characterised by varying genres and academic writing conventions, as learners in many academic settings wish to acquire knowledge of their own specialist domain, which will not be obtained from a general corpus (Smith, 2020). It also addresses the issue of EAP practitioners often not being specialists in the

target field, who as a result may find it challenging to design a set of guidelines accurately mapping the various discipline-specific language peculiarities (Anthony, 2016).

2. Literature review

In EAP contexts, the DDL approach to teaching has been investigated by scholars drawing on widely available corpora. These include, for instance, the British National Corpus containing samples of written and spoken language from a wide range of sources (e.g., Bridle, 2019; Chen & Flowerdew, 2018; Gilmore, 2009), the British Academic Written English (BAWE) corpus encompassing university student academic writing representing four broad disciplinary areas across four levels of study (e.g., Vincent & Nesi, 2018; Vincent et al., 2021), and student personal self-compiled do-it-yourself (DIY) corpora reflecting the students' specific disciplinary contexts (e.g., Charles, 2012, 2014, 2018; Dong & Lu, 2020; Lee & Swales, 2006; Smith, 2020).

The use of DIY corpora in EAP contexts catering for various disciplines has been extensively researched by Charles (2012, 2014, 2018), who highlights the value of this approach to teaching writing. In her study, Charles (2012) investigated multidisciplinary graduate EAP students' evaluation of their own discipline-specific DIY corpora and found that most were enthusiastic about working with their corpora. The vast majority of the participants reported that their corpus helped them improve their writing and that they intended to continue drawing on it in the future. The main advantage reported by participants related to subject specificity, particularly discipline-specific vocabulary, including access to examples of the usage of subject-specific terms, and the direct relevance of the corpus to their work. Charles (2012) thus suggests that DIY corpora are a valuable resource for academic writing in discipline-specific contexts which can be usefully implemented in mainstream EAP classes.

In another study, Charles (2014) investigated international graduate EAP students' use of DIY corpora one year after the participants had attended a course introducing them to building and examining their own personal corpus. Her study found that most participants continued to consult their corpus after the course, suggesting that the students were able to work with their corpus independently. The students reported having incorporated their corpora into their writing practices for both composing and revising, particularly for checking grammar and vocabulary, and believed that this approach had improved their academic writing. Charles's (2014) findings thus show that most participants developed a habit of autonomous long-term use of self-compiled corpora for their own language needs, highlighting the value of this approach in EAP contexts.

Charles (2018) also investigated the use of DIY corpora in the context of multidisciplinary doctoral students with a focus on corpus-assisted editing as part of doctoral training with the aim of enhancing the students' writing skills by learning how to use various corpus tools to edit their own writing. This study reports on participants' use of two types of corpora, including expert writing in their field (mainly research articles) and their own writing (primarily thesis chapters), and shows that access to both expert and learner corpora has the advantage of enabling students to make different queries and retrieve different types of information. Charles's (2018) findings thus suggest that corpus-assisted editing is a useful approach to helping students deal with the challenges they experience during the thesis writing process.

Other scholars have also reported on the benefits of using DIY corpora in discipline specific EAP contexts (e.g., Dong & Lu, 2020; Lee & Swales, 2006; Smith, 2020). Of these scholars, Dong and Lu (2020) integrated corpus-based and genre-based approaches to teaching rhetorical structures to engineering master's students. Their study shows that this integrated approach was successful in enhancing learners' genre knowledge, as well as their writing skills, thereby contributing evidence of the potential of this approach for developing learners' discipline-specific genre knowledge and genre-based writing skills. DIY corpora were also explored for Accounting and Finance vocabulary learning in the context of International Business students at a British university (Smith, 2020). The approach to vocabulary learning described in this study resulted in

improvements in technical vocabulary knowledge as it was found to provide a motivating and meaningful way for learners to acquire discipline-specific terminology and familiarise themselves with the usage of these newly acquired specialised vocabulary items. A corpus-assisted approach was also implemented in the context of Chinese doctoral students of pharmacology and biostatistics with a focus on rhetorical consciousness-raising (Lee & Swales, 2006). This study showed that students found the use of corpora confidence-building and empowering, with some participants stating the benefits of access to discipline-specific written discourses and relevant examples at any time. Lee and Swales (2006) believe that this approach to language learning helps learners become more autonomous as it gives them control over their own learning by enabling them to interrogate various linguistic questions that corpora are suited to answering.

The studies reviewed above highlight the benefits of using DIY corpora in EAP contexts catering for students at various levels of academic study as well as in different disciplines. The value of this approach is summarised by Gavioli (2005), who notes two advantages of employing corpus tools in a classroom: first, corpus searches can provide learners with useful information about various aspects of language in their discipline; second, the process of 'search-and-discovery' inherent in a DDL approach promotes learner autonomy in language use. Hence, the main benefit of utilising DIY corpora with EAP learners can be seen primarily in learners' ability to autonomously draw on texts directly relevant to their disciplines, which provide exposure to discipline-specific genres, including examples of writing conventions and terminology characteristic of their disciplines. Considering the direct relevance of DIY corpora to students' academic needs, this can be considered a meaningful way for learners to be exposed to and acquire disciplinary genre knowledge in EAP settings while promoting learner autonomy by placing the learner at the centre of the learning process (Boulton, 2011).

Nevertheless, the value of DIY over ready-made corpora is sometimes questioned, particularly in relation to the amount of time and technical knowledge required to compile them (Bridle, 2019). There are counterarguments to this view, however. First, while widely available corpora offer the advantage of their size compared to DIY corpora, one disadvantage relates to their general content (Boulton, 2016) since a large general corpus is unlikely to provide the relevant data output required by specialist discipline-specific queries. This can be a major drawback in discipline-specific contexts where students need to develop their subject-specific vocabulary and awareness of the writing conventions in their particular disciplines (Smith, 2020). Second, the process of corpus-building enables students to gain deeper insights into the nature of their own corpus data, which can assist them in interpreting the data more insightfully as it serves as a good foundation for corpus interrogation (Charles, 2012). The main benefit of utilising DIY corpora thus lies in their subject specificity, which makes this a suitable and valuable approach to academic writing in multidisciplinary EAP contexts.

However, despite the reported benefits of utilising DIY corpora in various EAP contexts, very little is currently known about the use of DIY corpora in the context of pre-sessional EAP programmes at British universities. The lack of research in this area may be due to the short and intensive nature of these courses, which typically take place in the summer months prior to the commencement of the students' degree study. Hence, it remains unclear whether it is feasible to incorporate a DDL approach into a pre-sessional EAP programme and to what extent this approach benefits the students. This highlights not only a gap in the current body of knowledge, but also in current EAP practices where the use of DIY corpora in the context of pre-sessional EAP programmes seems to be scarce. This paper, therefore, aims to address this important omission in EAP research and practice by investigating the benefits of a DDL approach utilising DIY corpora in the context of a Pathway to PhD pre-sessional EAP programme. Specifically, the aim of this paper is to examine the benefits of this approach in terms of its perceived discipline specificity for students learning how to write within their disciplines in preparation for their doctoral study.

3. Current study

3.1. Context and participants

This study is set in the context of a pre-sessional English for Academic Purposes (EAP) programme at a British university based in the North-East of England. The pre-sessional programme is targeted at international students who do not meet the requirements for direct entry to the University's degree programmes. Typically, the University runs pre-sessional programmes for undergraduate and postgraduate students from various disciplines over a period of 6, 10 and 20 weeks.

In the summer of 2021, the University also offered a separate pre-sessional programme for students aspiring to pursue doctoral study at the University. The programme was developed by the second author and delivered online over a period of 10 weeks by the first author. The goal of the programme was to improve the students' general level of English and academic skills and to raise their awareness of academic culture and conventions at British universities in order to prepare them both linguistically and culturally for the demands of doctoral level study in their chosen field. In addition, the programme aimed to further develop the students' subject-specific content and academic language knowledge relating to their future doctoral research. Therefore, the focus of the programme was on general academic skills as well as discipline-specific concepts and academic writing conventions, to which the students were introduced primarily through discipline-specific texts which were also exploited as a source of general academic and discipline-specific vocabulary. To ensure that the programme reflected the specific needs of doctoral research candidates in their respective disciplines, the programme was designed in consultation with academic departments and supervisors.

Three international students (2 female and 1 male) from China and Saudi Arabia representing three different disciplines (Physics, Sports Science and Archaeology) attended the 10-week Pathway to PhD pre-sessional programme with the aim of preparing for their doctoral study. Due to the different disciplines the students were pursuing, one of the challenges of the programme was to cater for the different disciplinary backgrounds by creating specialist content in a multidisciplinary classroom to meet students' academic needs and personal expectations of the programme. As reported by the participants, expectations included improving their English language, development of their content knowledge relating to their research topics, acquisition of discipline-specific academic writing knowledge, familiarisation with the PhD process, and knowledge of British academic norms and conventions. These needs and expectations were to be met by a data-driven approach to learning (DDL) and the use of do-it-yourself (DIY) corpora, described in more detail below.

3.2. Data-driven learning and do-it-yourself corpora

In the first week of the programme, the students were introduced to the concept of 'corpus' and were asked to build their own corpus containing journal articles, E-books and PhD E-theses from their discipline. The aim of this task was for each student to have their own corpus of authentic discipline-specific texts directly relating to their own topic and reflecting the academic writing conventions in their field of study. As the students had completed their masters in their disciplines, they already had a large number of sources available to them. In addition, they were also provided with a list of useful sources supplied by their PhD supervisors. As a result, the students were able to compile their discipline-specific DIY corpus relatively quickly and effortlessly. The students' corpora ranged from 1–3 million words in size, and contained primarily research articles with some PhD E-theses and a small number of E-books.

The corpus-building was followed by a practical DDL workshop in week 1 of the programme. In preparation for this workshop the students had watched several online tutorials made available by Sketch Engine (Lexical Computing, n.d.) and #LancsBox (Brezina et al., 2020), these being the two software packages utilised on the programme. These two corpus tools were selected for their

different functionalities and interface (one being web-based and the other desktop-based) to introduce the students to different tools available to enable them to choose their preferred software or to use them in a complementary manner. Both tools enable exploration of language using existing corpora (e.g., the British National Corpus or the Brown Corpus), as well as self-compiled corpora. Some of the tools' functionalities used on the programme include the Concordance function in Sketch Engine and the Key Word In Context (KWIC) function in #LancxBox, both generating a list of all instances of a search term in a corpus in the form of a concordance, which can subsequently be sorted or filtered to obtain the desired output. For example, these functionalities can be used to find the frequency of a word or phrase or grammatical categories such as nouns, verbs or adjectives. The search terms can also include 'smart searches' in #LancxBox for complex linguistic structures (e.g., PASSIVES, NOUNS for instance). In addition, collocations were explored using the WordSketch function in Sketch Engine and the GraphColl function in #LancsBox, generating collocates (i.e. words which systematically co-occur) of the search term. As well as identifying collocations, GraphColl enables identification of colligation (i.e., co-occurrence of grammatical categories), visualises both collocations and colligations, and identifies shared collocates of a word or phrase.

Following this introductory DDL workshop, during which the programme tutor did a demonstration of the selected tools and their different functionalities relevant to the purposes of the programme, the DDL approach was applied by subsequently utilising the DIY corpora in recurring 90-minute-long weekly online sessions. These sessions were facilitated by the first author over eight weeks, running from week 2 to week 9 of the programme (with week 1 being the introductory week and week 10 the assessment week) and included: Reading Support Sessions, Listening Support Sessions, and Vocabulary Building Sessions. The overarching aims of these sessions were two-fold: (i) to develop the students' discipline-specific content knowledge, and (ii) to expand their repertoire of discipline-specific vocabulary items and how they are used in their specific disciplinary contexts.

These sessions promoted facilitative learning whereby the students provided the lesson content and the instructor's primary role was to give feedback on their learning and guidance on future development and application of newly acquired knowledge, discussed in more detail below. The design and delivery of these sessions is also in line with the primary characteristic of the DDL approach to learning in which learners act as researchers drawing on language data supplied by a corpus (Johns, 1986, 1991, 2012).

3.2.1. The reading and listening support sessions

Whilst focusing primarily on content knowledge relating to the students' research topic, these sessions were also exploited for the development of their language. To achieve this, the students were asked to read one of the sources in their self-compiled corpus (i.e., a journal article, a chapter of an E-book, or a chapter of an E-thesis) for the Reading Support Session, and listen to a pre-recorded online lecture for the Listening Support Session. Both reading and listening materials were supplied by the students' PhD supervisors to ensure that the content reflected the students' subject specialism and were thus directly relevant to the students' research topics.

In preparation for the synchronous sessions delivered on Microsoft Teams, the students were given the following guidelines:

1. Provide a summary of the lecture / article.
2. What have you learnt from it?
3. How can it inform your own research?
4. What did you find particularly interesting and / or challenging (e.g., in terms of the content, language use, structure etc.)?
5. Note any new vocabulary or functional academic phrases.

3.2.2. Vocabulary building sessions

The Vocabulary Building Sessions followed the Reading and Listening Support Sessions and focused on the vocabulary and academic phrases that the students noted during the Reading and Listening Support sessions. In preparation for the Vocabulary Building Sessions, the students typically investigated various aspects of the newly met vocabulary noted during their reading and listening. These included the meaning, word class and pronunciation which they would look up using an online dictionary such as the *Longman Dictionary of Contemporary English Online* (Pearson, n.d.), *Oxford Learner's Dictionaries* (Oxford University Press, n.d.) or *Cambridge Dictionary* (Cambridge University Press, n.d.) for general academic words and discipline-specific online dictionaries for specialist terms. Additionally, the students explored the concordance lines, collocations and other frequent word combinations, particularly relating to the usage of prepositions, together with relevant frequency information using Sketch Engine and / or #LancxBox, as can be seen from Figure 1 extracted from one of the students' work.

Solid: Firm and stable in shape; not liquid or fluid (adj); A substance or object that is solid rather than liquid or fluid (noun).

As adj x354; As noun x75; As adv x2

nouns modified by "solid"

Line: solid line

Commun: Solid State Commun

Curve: solid curve

Drive: solid state drives

Theorem: a general proposition not self-evident but proved by a chain of reasoning; a truth established by means of accepted truths.

As noun x46

modifiers of "theorem"

Ehrenfest

Kramers

force

prepositional phrases

"theorem" of ...

"theorem" for ...

"theorem" to ...

"theorem" with ...

Concordances

1. The fundamental theorem of the theory of fine ferromagnetic particles.
2. Starting from the Kramers theorem for spin split bands it is shown that the spin surface.
3. We first illustrate the application of Ehrenfest's theorem to a spinless particle.
4. Applications of Ehrenfest's theorem with operator A equal to the three components of the position operator.

Figure 1. Example from Physics.

In addition to the development of the students' vocabulary, the Vocabulary Building Sessions were used for exploration of other important concepts relating to academic writing in the students' specific disciplines, which complemented the focus of the programme on the development of academic writing skills. For instance, the students interrogated their self-complied corpora for the usage of different reporting verbs (e.g., state, claim, demonstrate, suggest, find, explain etc.), and the frequencies with which these verbs occur to establish how common they are in their disciplines, together with the form (i.e. active versus passive), as shown in Figure 2.

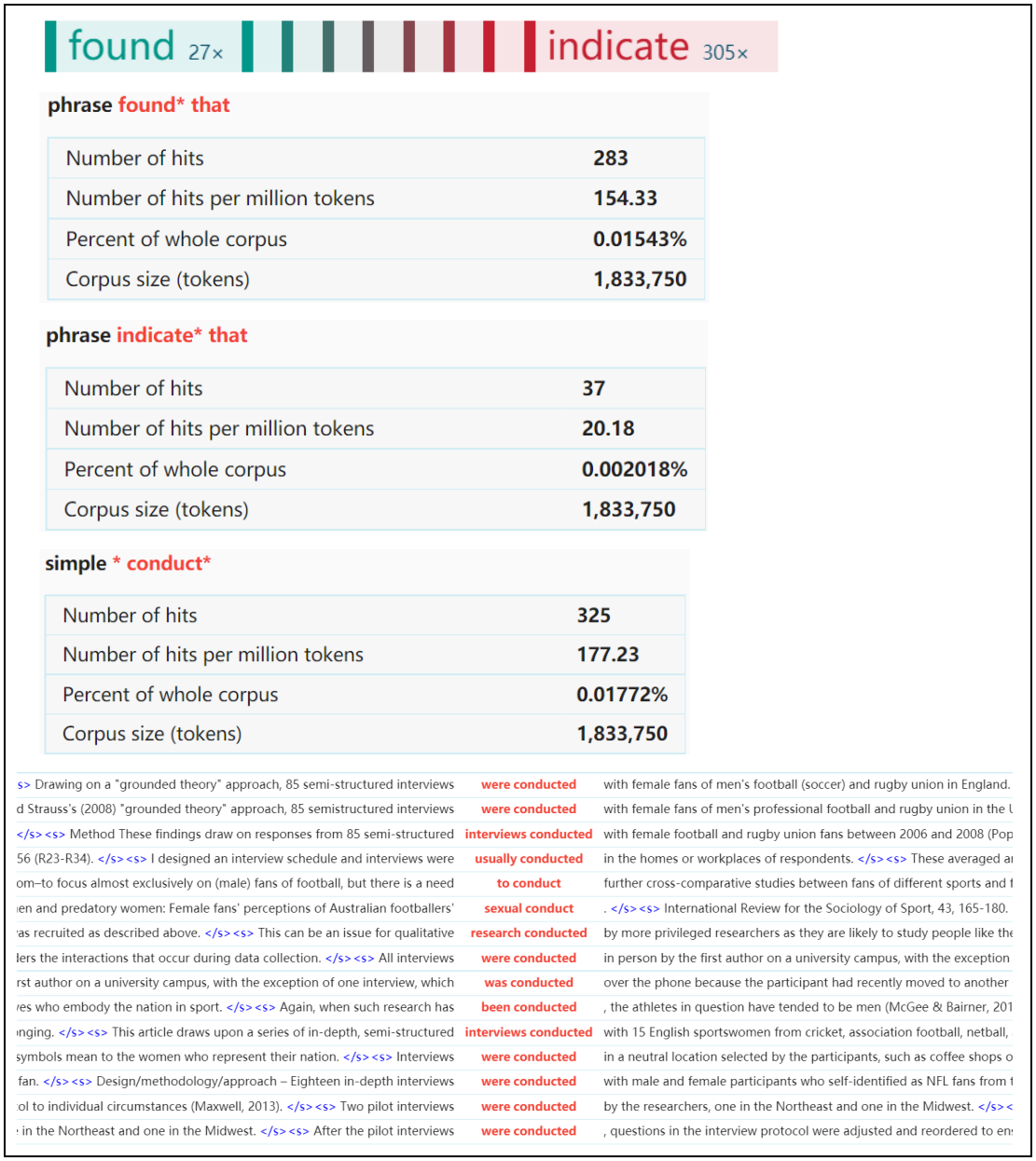

Figure 2. Example from Sport Science.

The tense of the reporting verbs was another aspect for which the students interrogated their corpora (Figure 3).

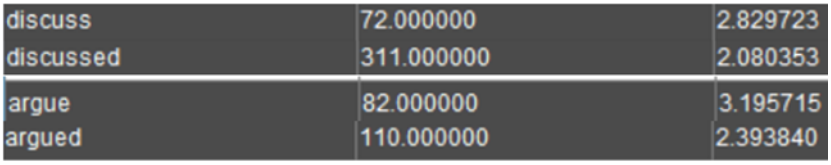

Figure 3. Example from Archaeology.

As can be seen from the examples above, the Vocabulary Building Sessions served multiple purposes, including the acquisition of discipline-specific vocabulary and their typical usage in the students' specialist contexts, as well as for the exploration of other concepts vital in academic writing, such as the usage of reporting verbs.

3.3. Research method

The effectiveness of the above-described DDL-focused sessions was investigated through semi-structured interviews relating to the students' reflections that were complemented by the programme instructor's reflections. The semi-structured format of interviews was selected to enable extensive follow-up of the participants' responses achieved by a loose set of guidelines allowing flexibility (Hyland, 2016). The interviews were conducted with individual participants online via Microsoft Teams at the end of the pre-sessional programme to enable them to comment on their learning experience, particularly their reflections on the use of self-compiled discipline-specific corpora during the DDL sessions.

To address the potential issue of reactivity relating to the effects of the researcher on the nature of the collected data (Hammersley & Atkinson, 2007), the interviews were conducted by the second author who was not involved in the delivery of the programme. We believe that this minimised the problem of reactivity and thereby increased the objectivity of the student responses. It is also noteworthy that eliminating the issue of reactivity is not always a prime consideration provided that the researcher understands how their presence may have influenced the collected data, which ought to be interpreted accordingly (Hammersley & Atkinson, 2007).

The interviews lasted approximately 20 minutes and covered a range of topics relating to the students' reflections on the Pathway to PhD pre-sessional programme, including the students' expectations of the programme and their perceptions of the most and least beneficial aspects of the programme. However, since a DDL approach had not previously been implemented on the University's pre-sessional programme, the main focus of the interviews was the students' perceived benefits of the Reading and Listening Support Sessions and Vocabulary Building Sessions, particularly their reflections on the use of self-compiled discipline-specific corpora during these sessions.

These reflections were explored through the following set of questions relating to:

a. the benefits of the programme: *In your opinion, what was the most beneficial aspect of the pre-sessional programme for you? Why? In your opinion, what was the least beneficial aspect of the pre-sessional programme for you (if any)? Why?*
b. programme content relating to the Reading and Listening Support Sessions and Vocabulary Building Sessions: *Can you tell me about the Reading Support Sessions / Listening Support Sessions / Vocabulary Building Sessions: Did you find them useful / beneficial? In what way? What did you learn from these sessions? What was the main benefit of these sessions for you?*
c. the use of discipline-specific self-compiled corpora: *How would you reflect on your use of your discipline-specific corpus? Did you find it useful / beneficial? How / why / in what way? What did you mostly use your discipline-specific corpus for? What was the size of your corpus / what did it contain?*

Through these questions, this study sought to explore the benefits of a DDL approach in terms of its discipline specificity on a PhD pre-sessional programme, as perceived by the students, as well as assessing the feasibility of this approach on a 10-week pre-sessional programme. The collected interview data were analysed using thematic analysis (Braun & Clarke, 2006). Following this approach, the interview data were interrogated using a deductive approach with the aim of identifying various themes closely related to the phenomenon under investigation. This led to the production of several initial themes. This stage was followed by further refining of these themes, resulting in themes relevant to the aims of the present study (Braun & Clarke, 2006).

The semi-structured interviews were accompanied by the tutor's reflection with the aim of providing the instructor's perspectives on the success of the DDL approach in the programme. These reflections were concerned primarily with the students' ability to use the corpus tools, the discipline-specific language output their DIY corpora generated, and how successful this was in informing the students' knowledge of discipline-specific genres and academic writing conventions.

In sum, the above outlined methods were employed with the aim of exploring the effectiveness and feasibility of a DDL approach utilising DIY corpora in the timeframe of a 10-week PhD pre-sessional EAP programme from both the students' as well as tutor's perspective. Specifically, the methods employed in this study were intended to investigate whether students from diverse disciplines can develop and draw on their self-compiled corpora to address various discipline-specific language queries.

4. Results and discussion

As will be detailed below, from the interviews with the three pre-sessional students, two main themes relating to the implementation of the DDL approach were identified. These include the value of self-compiled corpora (Sub-section 4.1.1), and their direct application (Sub-section 4.1.2), which also served as a basis for the tutor reflection (Sub-section 4.2).

4.1. Students' perspectives

4.1.1. Value of DIY corpora

The usefulness of DIY corpora was noted by all three participants, particularly in relation to their own work:

> *It's amazing … interesting and important ... I can see my work, which is also the most important place to use my corpus*. (Student 1)
>
> *This [the use of own corpus] was what I preferred most on the course.* (Student 2)
>
> *It was very useful … We could build our own corpus and search something in our own corpus ... When I am writing my assignment, I use my own corpus*. (Student 3)

This result shows the students' enthusiasm and perceived benefits of using their own self-compiled discipline-specific corpora on the pre-sessional programme. In particular, the usefulness of this approach in relation to their own work was noted. This is in line with other studies into DDL in multidisciplinary EAP contexts (e.g., Charles, 2012, 2014, 2018), as well as in discipline specific EAP settings (e.g., Dong & Lu, 2020; Lee & Swales, 2006; Smith, 2020) which report on the value of this approach.

The value of this approach was further emphasised by one participant, who noted that s/he was going to continue consulting their corpus in the future:

> *I'll keep it [my corpus] throughout my PhD life, throughout my academic life.* (Student 1)

This reflection indicates that the student was developing a long-term corpus habit. That is, s/he planned to use their personal corpus independently and autonomously after the programme finished by incorporating it into their writing practice to respond to their language needs. This ongoing use points to the student's commitment to their personal corpus. It also suggests that the participant regarded their own corpus as a valuable resource. This is consistent with Charles's (2014) study, which also found the students' commitment to long-term use of corpora.

The interview data thus suggest that the participants regarded their DIY corpus as a valuable resource, likely due to its direct relevance to their work. Learner autonomy is a further important point that emerged from the interview data.

4.1.2. Applications of DIY corpora

One of the reported applications of the students' self-compiled corpora related to the acquisition of discipline-specific vocabulary:

> *I can find lots of specific vocabulary, which means I can use it in my own writing.* (Student 1)
>
> *It's fantastic because at the beginning when I tried to look for some new vocabulary in a general corpus I didn't find results but when I built my own corpus I found it more useful; I could find more results.* (Student 2)
>
> *If we find unfamiliar vocabulary and if we just search in a dictionary we will forget how to use this vocabulary directly and when we compare those words in the Sketch Engine or #LancsBox we'll deepen our understanding of those words.* (Student 3)

In addition to acquisition of discipline-specific vocabulary, the participants reported drawing on their self-compiled corpora to learn about the usage of the newly acquired vocabulary items as well as to check correctness of usage in their own language production:

> *I can search the specific word and how to use it, what is the preposition to use with it and whether my usage is correct.* (Student 1)
>
> *In the past, I didn't think there was another way to understand a word, like how I can use the preposition. This is a big issue for me and I think for all international students so this helped me a lot to think about new vocabulary not just the meaning but also I can use it correctly … I used the corpus for many things like the classification of the word, how I can use the word, the modifiers of the word and prepositions.* (Student 2)

One participant also stated that s/he was going to continue drawing on their discipline-specific corpus to address their specific vocabulary-related language needs:

> *In future, I will use my corpus when I am confused about prepositions.* (Student 2)

This result indicates that the students utilised their self-compiled corpora for very specific language-related queries, such as the usage of prepositions. A further reported application of the DIY corpora related to the checking of correctness of usage. This finding corroborates other studies, which have also found the benefits of drawing on a corpus as a reference tool in error correction or editing (e.g., Bridle, 2019; Charles, 2018; Gilmore, 2009). These findings hence underline the benefits of DIY corpora in relation to the students' specific self-identified language queries, thereby further developing learner autonomy.

The interview data thus revealed the value of DIY corpora for the acquisition of discipline-specific vocabulary items as well as for checking the correctness of their own vocabulary usage in an autonomous manner.

4.2. Tutor's perspectives

4.2.1. Value of DIY corpora

From the tutor's perspective, the DDL approach proved effective for several reasons. First, the students showed great enthusiasm about working with their DIY corpora and engaged well with all sessions based on their own corpus, which generated language output of direct relevance to their research area. Students' enthusiasm about working with their own corpora, as observed by the tutor, is confirmed by the interview data (as discussed in Sub-section 4.1), and has also been reported by others (e.g., Charles, 2012). It would thus appear that the discipline-specificity and direct relevance of a DIY corpus are important characteristics of a DDL approach for graduate students who aim to further improve their discipline-specific language knowledge. This effectiveness was also reflected in the quality as well as volume of the work which the students produced

on a weekly basis in preparation for the Reading and Listening Support Sessions and Vocabulary Building Sessions (described in Sub-section 3.2).

Second, during the Reading and Listening Support Sessions and Vocabulary Building Sessions, the tutor was also able to observe a gradual development of learner autonomy by all three students over the course of the programme. This observation is in line with other studies (e.g., Gavioli, 2005; Boulton, 2011) which also underline the link between a DDL approach and the development of autonomous learning. On this programme, this was evident from the students' increasingly autonomous selection of words from the Reading and Listening Support Sessions, which they had chosen to focus on in the Vocabulary Building Sessions, and the different aspects of language use they explored through their corpus searches. In addition, the students started to use their own initiative to explore other functions of the two corpus tools without having received explicit instructions to do so.

Thus, based on the tutor reflection, the usefulness and value of using a DDL approach and drawing on DIY corpora in a relatively short multidisciplinary pre-sessional PhD programme was demonstrated by the learners' positive attitude towards using their own discipline-specific DIY corpora as well as the output they produced.

4.2.2. Applications of DIY corpora

In terms of direct applications of DIY corpora by the students, the tutor saw the value of the DDL approach in relation to the acquisition and development of vocabulary and academic writing conventions in the students' respective fields of study, and a subsequent deployment of the vocabulary items in their written production. These direct applications were enabled by specific authentic examples of language use which the learners were able to extract from texts relevant to their research topics. From the tutor's point of view, this text relevance was regarded as one of the primary benefits and direct applications of this approach due to the fact that the instructor, like many EAP practitioners, was not an expert on the students' varied topics and disciplines. As suggested by Anthony (2016), a DDL approach thus has the potential to address the issue of EAP instructors often not being specialists in the target field. Hence, it can be said that on this programme, the DDL approach utilising DIY corpora successfully replaced instruction with discovery learning as learners effectively conducted self-directed language searches reflecting their own specific linguistic needs, thus reflecting the 'search-and-discovery' inherent in a DDL approach (Boulton, 2011; Anthony, 2019; Hyland, 2006).

In addition to gaining insights into the academic writing conventions in their disciplines, owing to the representation of three different disciplines on the programme, the participants were able to learn about disciplinary differences relating to the various linguistic features explored on the programme, such as the different reporting verbs together with their frequencies, tense and form across disciplines (see Sub-section 3.2.2). These disciplinary differences became evident from the discussions the corpus findings presented by the students generated, which showed that the students started to notice differences between their disciplines. Hence, a multidisciplinary classroom can also be seen as beneficial in informing the students' knowledge of academic conventions as was the case on this programme where the insights from their own corpus searches, complemented by their colleagues' searches, subsequently informed the students' own written production on the programme.

The tutor reflection on the applications of the students' DIY corpora thus suggests that students can draw on self-compiled corpora in their written production and that this approach has direct applications relevant to the students' academic needs.

5. Conclusion

This study set out to investigate the feasibility and benefits of a data-driven learning (DDL) approach in the context of an EAP PhD pre-sessional programme at a British university. The

findings show not only the feasibility of DDL on a relatively short EAP pre-sessional programme, but also highlight several benefits of this approach. These benefits include the direct relevance of a self-compiled corpus to the students' work, the acquisition of discipline-specific vocabulary together with specific language queries (e.g. usage of prepositions) often used for editing purposes (i.e. checking of correctness of usage). The success of this approach has been underlined by one participant's reported intention of long-term autonomous use of their own corpora, and the tutor reflection on the implementation of the DDL approach on the programme which proved successful in informing the students' knowledge of discipline-specific vocabulary and academic writing conventions in their field of study, while at the same time promoting learner autonomy.

Despite the above reported benefits, the limitations of the current study need to be acknowledged. First, the findings are based on only a small number of participants on one pre-sessional programme. In addition, since the students were only interested in determining the conventions of their specific disciplines, as opposed to how their discipline might vary from other disciplines, comparisons in the students' corpus findings were not explicitly focused on. Also, as the interviews were conducted on completion of the programme, other implementations of the DIY corpora than those we have discussed may not have been reported by the participants due to the time that had elapsed.

Further research would, therefore, benefit from a replication of this approach on a greater number of pre-sessional programmes with a larger number of participants. To gain deeper insights into the different ways in which students draw on their corpora, it would be useful to interview students throughout the programme. Next, given that previous corpora research has identified disciplinary differences in academic writing conventions, further research could usefully explore disciplinary differences in a multilingual EAP classroom and how these results compare with what applied linguists have found, which would generate valuable findings that would benefit EAP practitioners without expertise in the relevant disciplines. Nonetheless, as this study has highlighted the value and success of this approach in the context of a multidisciplinary classroom represented by the three diverse disciplines investigated in the present study, it can be argued that a DDL approach utilising DIY corpora is likely to be successful in wider EAP settings. We thus propose that DDL could become an integral part of PhD pre-sessional programmes as well as more widely on EAP programmes catering for students from different disciplinary contexts.

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: Theory into practice* (pp. 162–180). Routledge.

Anthony, L. (2019). Tools and strategies for Data-Driven Learning (DDL) in the EAP writing classroom. In K. Hyland & L.C.W. Lillian (Eds.), *Specialised English: New directions in ESP and EAP research and practice* (pp. 179–194). Routledge.

Boulton, A. (2011). Data-driven learning: The perpetual enigma. In S. Goźdź-Roszkowski (Ed.), *Explorations across Languages and Corpora* (pp. 563-580). Peter Lang.

Boulton, A. (2016). Integrating corpus tools and techniques in ESP courses. *Concepts and Frameworks in English for Specific Purposes*, *69*, 113–137. https://doi.org/10.4000/asp.4826

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/https://doi.org/10.1191/1478088706qp063oa

Brezina, V., Weill-Tessier, P., & McEnery, A. (2020). #LancsBox v. 5.x. [software]. Available at: http://corpora.lancs.ac.uk/lancsbox

Bridle, M. (2019). Learner use of a corpus as a reference tool in error correction: Factors influencing consultation and success. *Journal of English for Academic Purposes*, *37*, 52–69. https://doi.org/10.1016/j.jeap.2018.11.003

Cambridge Dictionary (n.d.). Cambridge University Press. Retrieved from: https://dictionary.cambridge.org/

Charles, M. (2012). "Proper vocabulary and juicy collocations": EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes*, *31*(2), 93–102. https://doi.org/10.1016/j.esp.2011.12.003

Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for Specific Purposes*, *35*(1), 30–40. https://doi.org/10.1016/j.esp.2013.11.004

Charles, M. (2018). Corpus-assisted editing for doctoral students: More than just concordancing. *Journal of English for Academic Purposes*, *36*, 15–25. https://doi.org/10.1016/j.jeap.2018.08.003

Chen, M., & Flowerdew, J. (2018). Introducing data-driven learning to PhD students for research writing purposes: A territory-wide project in Hong Kong. *English for Specific Purposes*, *50*, 97–112. https://doi.org/10.1016/j.esp.2017.11.004

Dong, J., & Lu, X. (2020). Promoting discipline-specific genre competence with corpus-based genre analysis activities. *English for Specific Purposes*, *58*, 138–154. https://doi.org/10.1016/j.esp.2020.01.005

Gavioli, L. (2005). *Exploring corpora for ESP learning.* John Benjamins.

Gilmore, A. (2009). Using online corpora to develop students' writing skills. *ELT Journal*, *63*(4), 363–372. https://doi.org/10.1093/elt/ccn056

Hammersley, M., & Atkinson, P. (2007). *Ethnography: Principles in practice* (3rd ed.). Routledge.

Hyland, K. (2006). E*nglish for academic purposes: An advanced resource book.* Routledge.

Hyland, K. (2016). Methods and methodologies in second language writing research. *System*, *59*, 116–125. https://doi.org/10.1016/j.system.2016.05.002

Johns, T. (1986). Micro-concord: a language learner's research tool. *System*, *14*(2), 151–162.

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

Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. *ELR Journal*, *4*, 1–16.

Johns, T. (2012). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 293–313). Cambridge University Press.

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, *25*(1), 56–75. https://doi.org/10.1016/j.esp.2005.02.010

Lexical Computing (n.d.). Sketch Engine [software]. Available at: https://www.sketchengine.eu/

Oxford Learner's Dictionaries (n.d.). Oxford University Press. Retrieved from: https://www.oxfordlearnersdictionaries.com/

Longman Dictionary of Contemporary English Online (n.d.). Pearson. Retrieved from: https://www.ldoceonline.com/

Smith, S. (2020). DIY corpora for Accounting & Finance vocabulary learning. *English for Specific Purposes*, *57*, 1–12. https://doi.org/10.1016/j.esp.2019.08.002

Vincent, B., & Nesi, H. (2018). The BAWE Quicklinks Project: A New DDL Resource for University Students. *Lidil*, *58*. https://doi.org/10.4000/lidil.5306

Vincent, B,. Nesi, H., & Quinn, D. (2021). Exploiting corpora to provide guidance for academic writing: The BAWE Quicklinks project. In M. Charles & A. Frankenberg-Garcia (Eds.), C*orpora in ESP/EAP writing instruction: preparation, exploitation, analysis* (pp. 13–31). Routledge.