Addressing discipline specificity in a multidisciplinary EAP classroom through data-driven learning

Dana Therova *, Durham University, Durham, UK, dana.therova@durham.ac.uk

Andrew McKay, Durham University, Durham, UK, andrew.mckay@durham.ac.uk

Abstract Mastering academic writing is one of the challenges frequently experienced by university students across all levels and disciplines. As writing remains one of the most common ways of demonstrating knowledge in university settings, developing an appropriate academic style is a vital skill for success. In the context of British universities, academic writing skills are generally catered for by English for Academic Purposes (EAP) provision in the form of pre-sessional and in-sessional courses. Ideally, these courses should focus on the characteristics and conventions of the students' specific fields of study to meet their academic needs. This, however, poses a challenge for EAP practitioners, who are usually not specialists in the students' subject domain, amplified by the fact that EAP classes are often taken by a diverse group of learners from a wide range of disciplines. This paper reports on how the issue of discipline specificity in a multidisciplinary EAP classroom in a PhD pre-sessional programme at a British University was addressed by employing a data-driven learning (DDL) approach for the acquisition and development of disciplinary writing conventions including specialised technical vocabulary. After an evaluation of this approach, we conclude that DDL can be usefully implemented in wider EAP contexts to inform students' knowledge of writing in their disciplines.

* Corresponding Author

Background

The UK higher education (HE) climate is characterised by a diverse student body constituting both home (i.e., UK) and international (i.e., non-UK) students. In the academic year 2021-2022, there were 679,970 international students in UK HE institutions, representing approximately 24% of the total student population (Higher Education Statistics Agency [HESA], n.d.). This diversity in the student body inevitably leads to varying levels of academic literacy skills which we define as "the ability to communicate competently in an academic discourse community" (Wingate, 2018, p. 350). In the UK university context, students' academic literacy needs are primarily catered for by English for Academic Purposes (EAP) provision in pre-sessional and in-sessional programmes. Pre-sessional programmes are delivered over several weeks prior to the commencement of degree programmes and are targeted at international students who do not meet the required level of language proficiency. In-sessional courses offer academic support during the academic year alongside the students' degree courses (Pearson, 2020).

These EAP programmes thus play an important role in students' academic life by helping them develop the skills necessary for successful participation in their particular academic context. Of all the literacy skills that EAP programmes support, academic writing has received most attention. This is because writing is one of the main modes of demonstrating knowledge in university settings, in that written assignments constitute one of the principal forms of assessment (Lillis & Scott, 2007; Wingate & Tribble, 2012). It then follows that effective writing is often associated with academic success, whilst ineffective writing can lead to failure (Flowerdew, 2016). Writing is, therefore, regarded as a 'high stakes' activity and students need to acquire the writing conventions used in their specific academic setting and demonstrate the required standard of academic writing if they are to succeed in their studies (Lillis & Scott, 2007).

However, since academic writing practices are not universal, academic literacy needs of students across disciplines also vary reflecting the differences in writing conventions which can occur at word-, phrase-, sentence- or text-level. Examples of such disciplinary

variables include the specialised technical terminology characteristic of individual fields of study (Liu & Lei, 2019; Nation, 2001); the use of hedges and boosters (Jalilifar, 2007; Takimoto, 2015; Xie & Mi, 2023); voice, stance and engagement (Hyland, 2005; Silver, 2012); citation practices (Hyland, 1999, 2006; Pandita & Singh, 2017); self-mention (McGrath, 2016; Mohsen, 2016; Tao, 2021); passive voice (Bada & Ulum, 2018; Leong, 2021); or the variety of genres with a range of structural patterns common in different academic domains (Hyland, 2006). These examples illustrate the challenge that diverse disciplinary backgrounds pose for EAP practitioners who are not specialists in the students' target fields and may thus feel ill-equipped to cope with the differences in disciplinary writing conventions.

In this paper, we describe a data-driven learning (DDL) approach to addressing the unique and individual academic literacy needs of students from various disciplines with a specific focus on developing knowledge of disciplinary writing conventions.

Data-Driven Learning: Introduction

The term data-driven learning (DDL) was coined by Tim Johns (e.g., 1986, 1990, 1991) and is used to describe a student-centred approach to language learning where learners explore large amounts of authentic language data with the aim of self-discovery of linguistic features. DDL is thus an inductive inquiry-based learning approach in which instruction is replaced with discovery learning, or "the attempt to cut out the middleman" (i.e., the teacher) between the learner and the data (Johns, 2012, p. 297) and to bring learners closer to naturally occurring language use. This is enabled through direct access to language data supplied by a corpus (i.e., a large collection of electronic texts), which allows learners to discover language features which are most frequent in their specific contexts, and therefore highly relevant to their learning needs.

This makes DDL a valuable approach to teaching writing in EAP contexts catering for students from various disciplinary backgrounds who need to acquire the writing

conventions and jargon of their field of study (e.g., Charles, 2012, 2018; Smith, 2020; Therova & McKay, 2022). The open-ended supply of authentic language data tailored to specific learner needs which can be obtained from a corpus also addresses the common issue of EAP practitioners not being specialists in the students' target field (Anthony, 2019). In addition, the self-directed exploration of language data incorporating exemplar-based learning promotes independent learning and learner autonomy (Boulton & Cobb, 2017; Gavioli, 2005), supports learning and increases motivation as students are more likely to be engaged with and remember something they actively discover (Hunston, 2002).

Due to its characteristics, DLL has traditionally been regarded as closely aligned with a constructivist learning paradigm where engagement with corpus data leads to construction and consolidation of multiple forms of linguistic information. This allows students to come to meaningful data-driven conclusions about language use rather than them being presented with linguistic 'rules' (Boulton, 2010; Crosthwaite, 2021). However, this emphasis on constructivism as the main learning theory underpinning DDL has been challenged by O'Keeffe (2021), who points out that a focus on constructivist learning at an individual level neglects other paradigms under which learning occurs, namely socio-cultural approaches. O'Keeffe (2021) thus highlights the socio-cultural benefits of DDL when implemented in a classroom where both learners and teachers can be engaged in corpus data consultation as learners take part in guided learning through teacher- or peer-mediated scaffolded activities. Accordingly, DDL should be regarded as an approach to learning with two theoretical underpinnings of constructivist and sociocultural, which can help us not only better understand its underlying principles and benefits, but also inform our pedagogical practices when utilising direct applications of DDL in an EAP classroom.

The next section describes the context in which we implemented DDL, followed by an account of the DDL application including specific examples from the classroom, and finally an evaluation.

Data-Driven Learning: Context

Durham University has a long tradition of offering pre-sessional programmes for international undergraduate and taught postgraduate students across disciplines who, at the time of application, do not meet the requirements for direct entry to the University's degree programmes. These programmes are typically delivered over a period of 6, 10, or 20 weeks. In the summer of 2021, the University piloted a 10-week pre-sessional programme for postgraduate research students aspiring to pursue doctoral study at the University. That year, three students from China and Saudi Arabia representing the disciplines of Physics, Sports Science and Archaeology completed the programme. Following its successful introduction, the programme was further developed and repeated in the summer of 2022 with six students from Saudi Arabia (N=4), China (N=1), and Turkey (N=1) from the disciplines of Computer Science (N=3) and Mathematical Science (N=3).

Considering the diversity in the students' linguistic, cultural, and disciplinary backgrounds, one of the main challenges of the course design was meeting their academic literacy needs. This was of particular importance in the context of the PhD pre-sessional programme given the very specific nature of the students' research area and that submitting a written thesis is a prerequisite of a PhD award. One of the primary aims of the programme was, therefore, to enable students to acquire knowledge of disciplinary writing conventions including the specialised terminology of their respective fields of study. As the DDL approach lends itself to autonomous self-directed exploration of authentic linguistic features, it was deemed suitable for teaching disciplinary writing and vocabulary in a multi-disciplinary classroom.

Data-Driven Learning: Application

The DDL approach was implemented through a series of regular weekly sessions delivered throughout the 10-week programme. In week one, the students were introduced to the concept of 'corpus' and compiled their personal discipline-specific corpus. The guidelines for the compilation of their do-it-yourself (DIY) corpus totalling a minimum of 1

million words recommended the inclusion of relevant reading sources such as journal articles, E-books and E-theses supplied by their individual supervisors and related to the students' research topic. The rationale behind using a DIY corpus as opposed to drawing on widely available ready-made general corpora lies in the direct relevance of a self-compiled corpus to the students' academic needs as a general corpus is less likely to generate the relevant language output. This makes a DIY corpus more suitable in discipline-specific contexts in which students need to familiarise themselves with the conventions of their disciplines (Smith, 2020). The main advantage of using a DIY corpus thus lies in its subject specificity, which makes corpus interrogation a more meaningful way of exposure to the language of the students' specific academic contexts (Boulton, 2011). Additionally, the corpus building process gives students deeper insights into the nature of their corpus data, which can in turn help them interpret their language data more insightfully (Charles, 2012). These advantages of a DIY corpus underline the benefits of utilising a self-complied corpus in the context of a multidisciplinary pre-sessional programme.

Once the students had built their corpus, they were introduced to #LancsBox (Brezina et al., 2020), the corpus software utilised on the programme. This was done through a tutor demonstration of the tool and its functionalities complemented by a series of online tutorials made available by the #LancsBox developers. The main #LancsBox functions used on the programme included Key Word in Context (KWIC) and GraphColl. The KWIC function produces a list of all instances of a search term in the form of a concordance line representing a node with its context, which can subsequently be sorted or filtered to obtain the desired output. In #LancsBox the KWIC function can be used to search not only for individual words and phrases, but also for various grammatical classes such as nouns, pronouns, verbs, adjectives, and adverbs, as well as more complex linguistic structures such as noun phrases or passives. The KWIC searches were complemented by GraphColl output, which generates collocates (i.e., words which systematically co-occur) of the search term, identifies shared collocates of a word, produces colligations (i.e., co-occurrence of grammatical categories), and visualises collocations and colligations.

These two #LancsBox functions were used for the exploration of various aspects of disciplinary conventions in the students' individual fields of study using their self-compiled DIY corpora. As the programme was delivered online (following the Covid-19 pandemic), the students conducted their DDL corpus searches outside of class time in preparation for synchronous 90-minute online sessions, during which they shared their corpus findings. The corpus exploration drew on two types of other weekly sessions, namely Reading and Listening workshops. During these, in addition to practising and further developing their reading, listening skills, and expanding their topic-specific knowledge by engaging with discipline-specific texts supplied by their supervisors, the students were asked to identify 3-5 discipline-specific vocabulary items that were new to them. This corresponds to the notion of 'noticing,' which is a vital first step in vocabulary acquisition that occurs when learners become aware of an unfamiliar word and its usefulness (Nation, 2001). These workshops were, therefore, also designed to enable students to acquire the technical terminology of their disciplines, referring to words relating to a specific topic or subject area (Nation, 2001), which served as a basis for the subsequent DDL.

The exploration of the newly acquired discipline-specific vocabulary using the KWIC and GraphColl #LancxBox functions is closely associated with the various features of word use, which are important aspects of receptive knowledge of a word. These are: 'grammatical function' concerning the patterns in which a word typically occurs; 'collocation' referring to a word which typically co-occurs; and 'constraints on use' relating to how, where, when, and how often one would expect to encounter this word (Nation, 2001). These corpus searches were thus designed to develop the students' knowledge of the newly acquired vocabulary items in their contextual environments, as this is an important aspect of the development of their vocabulary knowledge. The corpus investigation of newly acquired vocabulary items and their usage characteristics was followed by productive use of the words with students practising using the new vocabulary in sentences. Following 'noticing,' this 'generative' use is another key step in the process of vocabulary acquisition (Nation, 2001). It can thus be said that the DDL approach to

vocabulary acquisition addressed not only the different steps of noticing and generative use which are part of the process of vocabulary acquisition, but it also exposed students to other essential aspects of receptive vocabulary knowledge such as grammatical function, collocation, and constraints on use (Nation, 2001).

In addition to acquiring new technical terms, investigating their characteristics and practising their usage, the DDL sessions were also exploited for exploration of other linguistic features important for writing in disciplines. These were conducted by 'smart searches' in the KWIC function in #LancsBox and included academic tone, hedging, passive voice, complex noun phrases, reporting verbs and tense. The corpus-assisted sessions were thus exploited to give students insights into several important aspects of disciplinary academic writing, ranging from the specialised vocabulary of the students disciplines to other features and conventions of disciplinary writing which the students need to become familiar with and demonstrate in their own written production, exemplified below.

Data-Driven Learning: Examples from the Classroom

This section provides examples of the students' corpus searches for acquisition and subsequent deployment of the various linguistic features of their disciplines, reflecting the process of 'noticing' together with examples of 'generative use' in the students' written production. Examples of the students' writing are taken from their final summative assignment completed on the programme comprising a 2,000-word Critical Literature Review relating to individual students' research topics.

Example 1: Example from Computer Science: 'constructivist'

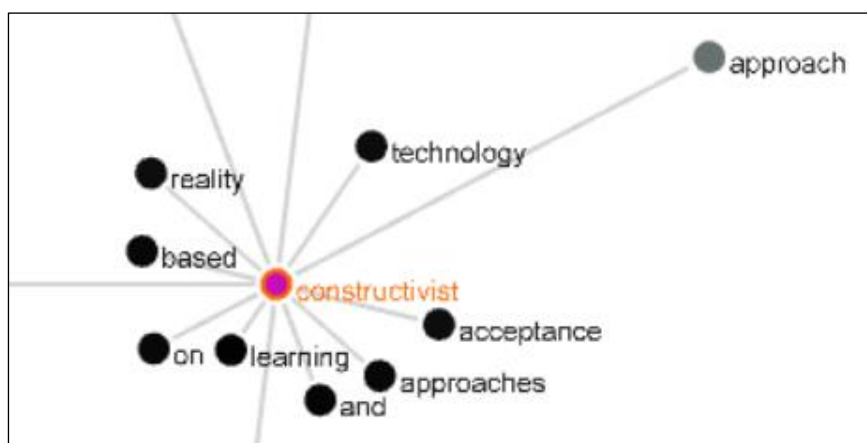**Figure 1:** *KWIC: 'constructivist'*

Figure 2: *GraphColl: 'constructivist'*

The ==constructivist== approach is not based on the repetition of knowledge, but on the transfer of knowledge and the learner's restructuring and application of knowledge (Perkins, 1999). Since the ==constructivist== learning approach aims to avoid rote learning and to ensure active participation of students in learning, technology-supported education is a useful method in concretizing abstract concepts and presenting students with rich learning activities that they can do themselves (Özmen, 2004). … In the ==constructivist== approach, technology plays a vital role in the creation of new learning products and communication in the learning and teaching process, as students take an active role in learning processes (İşman et al., 2002). Laney (1990) states that the use of technology in the ==constructivist== approach is effective in developing high-level thinking abilities that include identifying problems, solving problems, and producing appropriate solutions. … Increasing the number of such applications will be beneficial especially in terms of realizing a teaching with a ==constructivist== approach (Özmen, 2004). … In addition, there is a relatively small body of literature that is concerned with the examination of virtual reality with a ==constructivist== approach in maker activities.

Figure 3: *'constructivist' in student written production*

This example illustrates the student's search of a single vocabulary item (i.e. 'constructivist') using the KWIC and GraphColl functions in #LancsBox. Figure 1 shows

the contextual environment of 'constructivist' highlighting its typical occurrence with 'approach', also identified as a common collocation in Figure 2. These corpus investigations are subsequently reflected in the student's writing where 'constructivist' is used in combination with 'approach', as can be seen from Figure 3.

Example 2: Example from Mathematical Science: 'diagnostic test'

end goal of studying the accuracy of	diagnostic tests	is to apply these tests on future
proposed methods with some 1.1. Accuracy of	diagnostic tests	3 empirical classical methods, including the empirical
given in Section 1.4. 1.1 Accuracy of	diagnostic tests	In two-group classification, accuracy of a diagnostic
and diseased individuals. Measuring the accuracy of	diagnostic tests	is an important goal in medical research.
approaches have been introduced for accuracy of	diagnostic tests	[59, 73]. Test outcomes can be either

Figure 4: *KWIC: 'diagnostic test'*
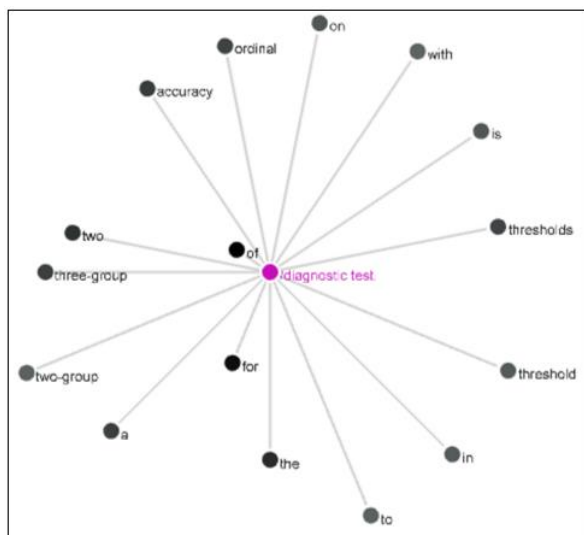
Figure 5:*GraphColl: 'diagnostic test'*

Therova D. & McKay A., (2024). Addressing discipline specificity in a multidisciplinary EAP classroom through data-driven learning. *Enhancing Teaching and Learning in Higher Education* Vol1. pp 21-40. http://doi.org/10.62512/etlhe.9

NPI for assessing the accuracy of ==diagnostic tests== in many fields has been introduced. This demonstrate the value of NPI and an evaluation of NPI framework will be provided in the following sections. … In many application fields including health care, machine learning and finance sector, determining the accuracy of ==diagnostic tests== is crucial. … Also, this approach has been introduced for ==diagnostic test== accuracy considering different types of results such as continuous (real-values), ordinal, and binary data, discussed next. … In the case of real-valued outcomes, Coolen-Maturi et al. (2012b) introduced NPI for measuring the accuracy of ==diagnostic tests== for continuous outcomes. … To compare the accuracy of two ==diagnostic tests==, Alabdulhadi et al. (2021) contrasted two ==diagnostic tests== with continuous outcomes for two or three segments of future individuals. … Together these studies provide important insights into application of NPI method to evaluate the accuracy of ==diagnostic test== results in ordinal data even though they have limitations, which require more investigations in the future.

Figure 6: *'diagnostic test' in student written production*

This example shows the student's search of a multiword expression (i.e. 'diagnostic test') using KWIC and GraphColl. The KWIC search (Figure 4) reveals the combination of 'accuracy of diagnostic tests' further confirmed by the GraphColl search (Figure 5). Figure 6 shows the frequency with which the student used 'diagnostic test(s)' in their writing, in most instances in combination with 'accuracy'.

Example 3: Example from Computer Science: reporting verb and tense

Figure 7: *KWIC: reporting verb ('propose') and tense*

> ... Godard et al. [2] propose a minimal reprojection error to address occlusion and reduces visual artifacts using full-resolution multiscale sampling. … There have been some works [5], [8], [9], [20], [21] that propose solutions to train static and moving objects separately, which aim to solve the problem that moving objects in assumed static environment cannot be reprojected well to the correction position.

Figure 8: *'propose' in student written production*

This example shows how the student used KWIC in #LancsBox to explore the usage of a specific reporting verb (*propose*) in their corpus, followed by examples from the student writing, also reflecting the use of the present tense of the verb commonly found in literature reviews.

The examples of the students' corpus searches above illustrate the different aspects of language for which they utilised #LancsBox and their DIY corpus, ranging from single and multiple-word items to reporting verbs and tense, and how they drew on their corpus findings in their own writing.

Data-Driven Learning: Evaluation

The overarching aim of the DDL sessions was to equip students with the knowledge of disciplinary academic conventions as part of the process of preparing them for their doctoral study by expanding their knowledge of various features of academic writing in their particular fields of study together with discipline-specific terminology. This was achieved through a DDL approach, which proved successful in addressing individual students' academic needs in a multidisciplinary EAP classroom by enabling them to focus on discipline-specific academic conventions through work with their own self-compiled corpora containing texts relevant to their unique academic contexts and needs. This can be seen from the examples of the students' written production above ranging from individual vocabulary items to multiword expressions and other features of academic writing (e.g., reporting verbs and tense), which show the relationship between the corpus

searches utilising the students' own corpora enabling 'noticing' of the selected features leading to 'generative use' in the students' writing.

The value of this approach also became evident from interviews which we conducted with individual students on completion of the programme, during which the students reflected on their use of their personal corpus. Examples of the students' reflections include positive attitudes: "*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). One participant noted the benefits of utilising a DIY corpus in relation to their own writing: "*It was very useful… When I am writing my assignment, I use my own corpus*" (Student 3). Students also reported several advantages of drawing on their DIY corpus for vocabulary acquisition: "*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… 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); and its subsequent usage in writing production: "*I can find lots of specific vocabulary, which means I can use it in my own writing*" (Student 1)*;* (Therova & McKay, 2022). These quotes illustrate the students' perceived value of a DDL approach utilising a specialised self-compiled corpus.

The main benefit of employing a DDL approach in a multidisciplinary EAP classroom thus lies in enhancing individual students' knowledge of disciplinary writing conventions without the tutor being a specialist in the students' disciplines, achieved by utilising a DIY corpus which provides a unique resource for self-directed explorations of authentic language use relevant to the students' area of study. In addition, through discovery learning inherent in DDL, this approach to learning promotes learner autonomy by putting the learner at the heart of the learning process.

Despite these clear benefits of DDL, several drawbacks of this approach need to be acknowledged with regard to utilising a corpus-assisted approach in a classroom in general, and in the context of the present study in particular. Among the obstacles frequently noted by both learners and practitioners as limiting corpus adoption is the amount of time it takes to implement DDL effectively. One of these issues relates to learner and tutor training in employing a DDL approach and the selected corpus software, which may be time-consuming. Other time-related issues include time demands on the teacher outside of the classroom, which involves building a suitable and representative corpus; time concerns relating to in-class DDL activities, specifically the time required to search and analyse the corpus output in order to draw conclusions about language use (Farr, 2008; McCarthy, 2008; Jones, 1991; Poole, 2022); and time which needs to be incorporated into or freed up from an existing curriculum to integrate corpus activities in a manner that complements instruction (Poole, 2022). Some practitioners may thus feel reluctant to adopt a DDL approach as the time and effort required to implement it is more demanding than following "a traditional syllabus" (Poole, 2022). Further, it may leave some practitioners feeling that they do not have the required technical knowledge to not only conduct corpus searches effectively, but also interpret corpus findings accurately and meaningfully (Bridle, 2019; Farr, 2008; Poole, 2022). Some teachers have described this lack of confidence in corpus findings interpretation as "a feeling of fear and losing control" (Breyer, 2009, p. 166), closely relating to concerns regarding learners' confidence in corpus data analysis (Bunting, 2013). Learners' proficiency is also often reported as a potential obstacle in adopting DDL in a classroom as it is often perceived as potentially too difficult for learners with lower proficiency levels, thus more suitable for more advanced learners (Poole, 2022). The diversity in the students' educational backgrounds could also mean that students from deductive learning backgrounds might struggle with the inductive discovery learning rooted in a DDL approach (Crosthwaite, 2021). Another potential disadvantage for the tutor lies in having little control over what happens in terms of the corpus searches carried out by students and the language output they arrive at (Hunston, 2002).

From the above noted challenges of adopting a DDL approach, relevant to the context reported on in this study was the inability of the teacher to monitor the students' corpus

investigations due to the distance mode of learning of the pre-sessional programme in 2021 and 2022. As the students' corpus searches were conducted independently outside class time, it was difficult to track how learners were conducting their corpus consultations or assist them with corpus queries. This is a notable limitation of implementing a DDL approach in an online learning context, and a further study could usefully explore the implementation of a DDL approach in a classroom to assess to what extent the drawbacks of a DDL approach outlined above apply to the context of a PhD pre-sessional programme.

Conclusion

This paper has reported on a DDL approach to teaching writing and vocabulary in a multidisciplinary EAP classroom in the context of a 10-week online PhD pre-sessional programme at a British university. It has described how this approach was used to cater for the varying academic needs of students from different disciplinary backgrounds, particularly in relation to the acquisition and development of disciplinary writing conventions and specialised vocabulary. The evaluation of the application of DDL has highlighted numerous benefits of adopting this approach resulting primarily from the direct relevance of DIY corpora to students' academic needs. We, therefore, argue that a DDL approach utilising self-compiled corpora could be implemented in wider EAP settings, becoming an integral part of EAP provision catering for students from different disciplinary contexts.

References

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.

Bada, E. & Ulum, Ö.G. (2018). Utilization of active and passive constructions in English academic writing. *International Journal of Human Sciences*, *15*(1), 413-421. https://www.j-humansciences.com/ojs/index.php/IJHS/article/view/5256

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language learning*, *60*(3), 534-572. https://doi.org/10.1111/j.1467-9922.2010.00566.x

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. & Cobb, T. (2017). Corpus Use in Language Learning: A Meta-Analysis. *Language learning*, *67*(2), 348-393. https://doi.org/10.1111/lang.12224

Breyer, Y. (2009). Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning*, *22*(2), 153-172. https://doi:10.1080/09588220902778328

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

Bunting, J. (2013). *An investigation of language teachers' explorations of the use of corpus tools in the English for academic purposes class* [Unpublished doctoral dissertation]. Atlanta: Georgia State University.

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

Crosthwaite, P. (2021). DDL is dead? Long live DDL! Studies reviving the field of Data-Driven Learning. *International Perspectives on Corpus Technology for Language Learning - Seminar Series.* Retrieved from https://languages-cultures.uq.edu.au/event/session/7299

dos Santos, E.A., Peroni, S., & Mucheroni, M.L. (2022). The Way We Cite: Common Metadata Used Across Disciplines for Defining Bibliographic References. In: G. Silvello, O. Corcho, P. Manghi, G. M. Di Nuzio, K. Golub, N. Ferro, & A. Poggi (Eds). *Linking Theory and Practice of Digital Libraries* (pp. 120–132). 6th International Conference on Theory and Practice of Digital Libraries. https://doi.org/10.1007/978-3-031-16802-4_10

Farr, F. (2008). Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users. *Language Awareness*, *17*(1), 25-43. https://doi:10.2167/la414.0

Flowerdew, J. (2016). English for Specific Academic Purposes (ESAP) writing: Making the case. *Writing & Pedagogy*, *8*(1), 5-32. https://doi.org/10.1558/wap.v8i1.30051

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

Higher Education Statistics Agency [HESA]. (n.d.). Higher Education Student Data. Retrieved from https://www.hesa.ac.uk/data-and-analysis/students

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge University Press.

Hyland, K. (2006). Disciplinary differences: Language variation in academic discourses. In: K. Hyland, & M. Bondi (Eds.). *Academic discourse across disciplines* (pp. 17-49). Peter Lang Verlag.

Hyland, K. (2005). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies*, *7*(2), 173-192.

Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics*, *20*(3), 341-367.

Jalilifar, A. R. (2007). Hedging as a pragmatic strategy: Variations across disciplines and cultures. *Teaching English Language*, *2*(1), 43-69.

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.

Leong, P. A. (2021). The passive voice in scholarly writing: A diachronic look at science and history. *Finnish Journal of Linguistics*, *34*, 77-102. https://journal.fi/finjol/article/view/103168

Lillis, T., & Scott, M. (2007). Defining academic literacies research: issues of epistemology, ideology and strategy. *Journal of Applied Linguistics and Professional Practice*, *4*(1), 5-32. http://dx.doi.org/doi:10.1558/japl.v4i1.5

Liu, D. & Lei, L. (2019). Technical vocabulary. In: S. Webb (Ed.). *The Routledge handbook of vocabulary studies* (pp. 111-124). Routledge.

McCarthy, M. (2008). Assessing and interpreting corpus information in the teacher education context. *Language Teaching*, *41*(4), 563-574. https://doi:10.1017/S0261444808005247

McGrath, L. (2016). Self-mentions in anthropology and history research articles: Variation between and within disciplines. *Journal of English for Academic Purposes*, *21*, 86-98. https://doi.org/10.1016/j.jeap.2015.11.004

Mohsen, K. (2016). Are we visible? An interdisciplinary data-based study of self-mention in research articles. *Poznan Studies in Contemporary Linguistics*, *52*(3), 403-430. https://doi.org/10.1515/psicl-2016-0017

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

O'Keeffe, A. (2021). Data-driven learning, theories of learning and second language acquisition: In search of intersections. In P. Pérez-Paredes & G. Mark (Eds.), *Beyond Concordance Lines: Corpora in language education* (pp. 35-55). John Benjamins.

Pandita, R. & Singh, S. (2017). Self-citations, a trend prevalent across subject disciplines at the global level: an overview. *Collection Building*, *36*(3), 115-126. https://doi.org/10.1108/CB-03-2017-0008

Pearson, W. S. (2020). Mapping English language proficiency cut-off scores and pre-sessional EAP programmes in UK higher education. *Journal of English for Academic Purposes*, *45*, 1-11. https://doi.org/10.1016/j.jeap.2020.100866

Poole, R. (2022). "Corpus can be tricky": Revisiting teacher attitudes towards corpus-aided language learning and teaching, *Computer Assisted Language Learning*, *35*(7), 1620-1641, https://doi.org/10.1080/09588221.2020.1825095

Silver, M. (2012). Voice and stance across disciplines in academic discourse. In: K. Hyland & C. S. Guinda (Eds.) *Stance and Voice in Written Academic Genres* (pp. 202-217). Palgrave Macmillan. https://doi.org/10.1057/9781137030825_13

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

Takimoto, M. (2015). A corpus-based analysis of hedges and boosters in English academic articles. *Indonesian Journal of Applied Linguistics*, *5*(1), 95-105.

Tao, H. (2021). A diachronic analysis of self-mention and its rhetorical functions in academic writing across four disciplines. *Open Journal of Social Sciences*, *9*, 118-134. https://doi.org/10.4236/jss.2021.97008

Therova D. & McKay A.

Therova, D. & McKay, A. (2022). Introducing data-driven learning on a PhD pre-sessional programme. *Journal of Academic Language and Learning, 16*(1), 91-104. https://journal.aall.org.au/index.php/jall/article/view/819/435435553

Wingate, U. (2018). Academic literacy across the curriculum: Towards a collaborative instructional approach. *Language Teaching, 51*(3), 349-364. https://doi.org/10.1017/S0261444816000264

Wingate, U., & Tribble, C. (2012). The best of both worlds? Towards an English for Academic Purposes/Academic Literacies writing pedagogy. *Studies in Higher Education, 37*(4), 481-495. https://doi.org/10.1080/03075079.2010.525630

Xie, S. & Mi, C. (2023). Promotion and caution in research article abstracts: The use of positive, negative and hedge words across disciplines and rankings. *Learned Publishing, 36*, 249-265. https://doi.org/10.1002/leap.1515