Autonomous corpus use by graduate students

Charles, Maggie; Hadley, Gregory

DOI:
10.1016/j.jeap.2022.101095

Document Version
Peer reviewed version

Citation for published version (Harvard):
Charles, M & Hadley, G 2022, 'Autonomous corpus use by graduate students: a long-term trend study (2009–2017)', *Journal of English for Academic Purposes*, vol. 56, 101095.
https://doi.org/10.1016/j.jeap.2022.101095

Journal of English for Academic Purposes

Autonomous corpus use by graduate students: A long-term trend study (2009-2017)

--Manuscript Draft--

Manuscript Number:	JEAPJ-D-21-00410R1
Article Type:	Full Length Article
Keywords:	Do-it-yourself corpora; student corpus use; EAP learner autonomy; long-term corpus use; graduate EAP; data-driven learning
Corresponding Author:	Maggie Charles Oxford University Oxford, UNITED KINGDOM
First Author:	Maggie Charles
Order of Authors:	Maggie Charles
	Gregory Hadley, PhD
Abstract:	Corpus use by EAP students has reportedly increased over the last decade, with considerable optimism about the future of this approach (Chen & Flowerdew, 2018a). However, much research employs data from short classroom courses; little is known about how student corpus use has varied over a span of multiple years. This paper uses long-term trend data from a corpus-based course for graduates which ran 50 times (2009-2017) at a UK university. The course taught students to build do-it-yourself corpora based on their research topic and promoted autonomous consultation of this resource. Questionnaires on corpus use were administered at three stages: pre-course (544 students), immediate post-course (343) and delayed post-course, after one year (221). The data show that pre-course corpus use was constant (mean 24%), while immediate post-course use (mean 87%) and delayed post-course use rose only slightly (mean 62%) from 2009 to 2017. The lack of appreciable growth in corpus use over nine years does not support the expectation of increased take-up in future. However, the means for regular autonomous use (≥1/week) at 61% (immediate post-course) and 37% (delayed post-course), show the success of the do-it-yourself corpus approach in fostering the autonomous use of corpora by graduates.
Suggested Reviewers:	Ana Frankenberg-Garcia, PhD Reader, University of Surrey a.frankenberg-garcia@surrey.ac.uk specialist in use of corpora with students
	Karin Whiteside, PhD Head of EAP, University of Reading k.whiteside@reading.ac.uk specialist in use of corpora with graduates
Response to Reviewers:	

Autonomous corpus use by graduate students: A long-term trend study (2009-2017)

Maggie Charles[a,*], Gregory Hadley[b]

[a]University of Oxford, Oxford, UK

[b] Faculty of Humanities, Humanities Building A, Room 503, Niigata University, 8050 Ikarashi 2 Nocho, Nishi Ward, Niigata, 950-2181, Japan

Corresponding author: Maggie Charles

Email addresses: maggiecharles_oxford@yahoo.com (M. Charles)
ghadley@human.niigata-u.ac.jp (G. Hadley)

Autonomous corpus use by graduate students: A long-term trend study (2009-2017)

Abstract

Corpus use by EAP students has reportedly increased over the last decade, with considerable optimism about the future of this approach (Chen & Flowerdew, 2018a). However, much research employs data from short classroom courses; little is known about how student corpus use has varied over a span of multiple years. This paper uses long-term trend data from a corpus-based course for graduates which ran 50 times (2009-2017) at a UK university. The course taught students to build do-it-yourself corpora based on their research topic and promoted autonomous consultation of this resource. Questionnaires on corpus use were administered at three stages: pre-course (544 students), immediate post-course (343) and delayed post-course, after one year (221). The data show that pre-course corpus use was constant (mean 24%), while immediate post-course use (mean 87%) and delayed post-course use rose only slightly (mean 62%) from 2009 to 2017. The lack of appreciable growth in corpus use over nine years does not support the expectation of increased take-up in future. However, the means for regular autonomous use (≥1/week) at 61% (immediate post-course) and 37% (delayed post-course), show the success of the do-it-yourself corpus approach in fostering the autonomous use of corpora by graduates.

Keywords: Do-it-yourself corpora; student corpus use; EAP learner autonomy; long-term corpus use; graduate EAP; data-driven learning

1. Introduction

Recent surveys have reported that the use of corpora by EAP students has increased markedly over the last decade. For example, Tribble (2015) notes a rise in the number of instructors using corpora in their teaching, and since the majority of respondents to his survey are active in university settings, most are likely to be engaged in some form of EAP. Similarly, Chen and Flowerdew

(2018a) found a substantial increase in the number of papers on corpus use in EAP published between 2010 and 2016 in comparison with the previous 10 years, again suggesting that more students are being exposed to data-driven learning (DDL). There is also mounting evidence that the informed application of corpus tools and techniques promotes both vocabulary learning (Lee et al., 2018) and language learning more generally (Boulton & Cobb, 2017). Such findings support the view that student corpus use is increasing and the expectation that it will continue to do so in future. Prospects for the further adoption of corpus approaches in EAP therefore seem bright, a view underlined by Chen and Flowerdew's (2018a, p. 358) claim that DDL is becoming 'more mainstream'.

Other researchers, however, sound a note of caution with regard to what we might term the 'rhetoric of optimism'. Both Chambers (2019) and Pérez-Paredes (2019) draw on the work of Bax (2003, 2011) concerning the normalisation of computer-assisted language learning (CALL) in order to cast some doubt upon the extent to which corpus use has become a widespread and accepted technique in second language teaching and learning. Bax (2003) distinguishes seven stages in the normalisation of CALL, the last of which refers to the point at which the technology becomes so integrated into the language learning and teaching environment that it is no longer the focus of attention. Bax (2003) argues that it is only when the technology has become 'invisible' in this way that we can consider its use to be fully normalised. Building upon Bax's work, Chambers (2019) notes the existence of a gap between the published accounts of research using corpora in the classroom and most language teachers' everyday teaching reality. Thus she questions the extent to which corpus use has become normalised as an accepted part of teachers' practice. Pérez-Paredes (2019) deals with the issue of the normalisation of corpus use from a slightly different standpoint. He reviews papers on pedagogic corpus use that appeared in the top five research journals in the field of CALL over the period 2011-2015 and analyses the factors likely to impede the wider adoption and normalisation of DDL. Among these, he points to the fact that most of this research dealt with corpus use in class, i.e. not independently, and analysed only the short-term impact of

DDL on language gains; less attention was paid to students' cognitive abilities and their long-term autonomous corpus use. Like Chambers, he perceives a disconnect between research and teaching, noting that the main stake-holders in these papers are researchers rather than teachers or students. These two studies suggest that, although the reported increase in student corpus use provides some grounds for optimism, there is nevertheless a long way to go before DDL can be considered as normalised.

The research focus identified by Pérez-Paredes (2019) on brief, one-off class-based courses does not enable enquiry into either students' autonomous corpus use outside class or their long-term corpus use after the course has ended. These two issues are closely intertwined, since students who are in the habit of independent corpus consultation have the means to apply these skills over the entire duration of their academic, professional and personal lives. Seen in this light, it is clear that developing a greater understanding of the two issues is of prime importance if corpus use is to become more firmly embedded within EAP teaching and learning.

Boulton's (2017) research timeline shows that during the early years of the twenty-first century much research on student corpus use was concerned with establishing the extent to which the practice was effective. Thus many reports were limited in scope, examining the performance of a single group of students during or after a single course or intervention, which itself was often a one-off phenomenon. As noted in Author (in press), such experimental classes allowed little, if any, autonomous corpus use and often lasted only a matter of weeks (e.g. Boulton, 2010; Cresswell, 2007; Pérez-Paredes et al. 2013). While some accounts included the suggestion that the research should be replicated using an extended time-frame (Gaskell & Cobb, 2004; Huang, 2011; Pérez-Paredes et al., 2011), such work seems rarely, if ever, to have been reported. Autonomous corpus use has been a particular focus for studies of graduate learners (Author, 2012, 2014, 2015; Cortes, 2007; Cotos et al., 2017; Lee & Swales, 2006). It has often been researched through detailed case studies of small groups of learners, typically using a longer time-frame of several months (e.g. Chang, 2014; Park & Kingerer, 2010; Yoon, C., 2016; Yoon, H., 2008). More recently, however,

there have been large-scale investigations of whole cohorts following a specific academic programme with an element of autonomous use; these include Chen and Flowerdew's (2018b) research on graduate workshops and Crosthwaite et al.'s (2019) report on graduates' corpus work over the course of four to five months.

It would seem, then, that Flowerdew's (2015) call for more longitudinal research is currently being addressed and with it, the investigation of autonomous corpus use. Undoubtedly, the studies mentioned above have provided much valuable information about students' reactions to corpus work, their corpus enquiry behaviour and degree of engagement with the approach. However, the time period examined still tends to be relatively short (i.e. a matter of months), which means that the findings cannot give us a sense of the development of student corpus use over the years. In order to do so, this paper takes a different approach: it provides a long-term trend study of data from a single course which ran 50 times from 2009 to 2017. The course taught students to build do-it-yourself (DIY) corpora based on their research topic and promoted autonomous consultation of this resource.

Following Holec (1979, p. 3), we define autonomy in language learning as 'the ability to take charge of one's own learning'. Accordingly, we examine the students' use of their corpus outside the classroom and with no input from corpus experts. Based on responses to questionnaires administered at three stages, pre-course, immediate post-course and delayed post-course, this study investigates the extent to which this autonomous corpus consultation was adopted by students over the nine years. It is guided by the following research questions (RQs):

RQ1. How prevalent was pre-course corpus use for English language learning (ELL) 2009-2017?

RQ2. How prevalent was immediate post-course use of the DIY corpus outside class (i.e. autonomously) 2009-2017?

RQ3. How prevalent was delayed post-course use of the DIY corpus (i.e. after one year) 2009-2017?

RQ4. How prevalent was regular corpus use (≥1/week) at the pre-course, immediate post-course and delayed post-course stages 2009-2017?

To the best of the authors' knowledge, no other study of EAP corpus pedagogy has taken this long-term view, spanning multiple iterations and years. Such research enables us to see the degree to which corpus use has become an integral and normalised part of student writing practices over an extended period of time.

2. Context of the research

The course in question, entitled 'Writing in your Field with Corpora', ran once a year at a UK research university and formed part of the in-sessional provision on academic writing for graduate students. The course consisted of one two-hour session for six weeks and was held in a computer laboratory, although many students preferred to use their own laptops and were encouraged to do so. During the nine years of the study, four to seven parallel multi-disciplinary classes were offered each year and the course was taught by four different tutors. As the course was open-access and non-credit-bearing, attendance was irregular, but an average of nine students attended each class (Range: 3-14). Following the approach of Lee and Swales (2006), students built their own DIY discipline- and topic-specific corpora from research articles (RAs) they had already downloaded and saved for their research purposes. The aim of the course was for students to create a tailor-made resource for individual ongoing use. In-class, they explored realisations of specific discourse functions within their own discipline using a set of search terms provided by the tutor, while outside class they were encouraged to consult their corpora autonomously for help with their research writing. See Author (2012, 2014, 2015) for further details on the course and materials. The software used was AntConc (Anthony, 2018). Over the research period, the course underwent several name changes; improvements to the tasks were also introduced and timetabling issues sometimes dictated

changes in the sequence of topics. These were minor modifications, however, because the aims, structure, content, and approach of the course remained essentially unchanged.

3. Method and Data

The data for this paper derive from three questionnaires which were administered to students who attended the course: pre-course (Qre1), immediate post-course (Qre2) and delayed post-course (Qre3). Qre1 and Qre2 were paper-based questionnaires completed in class, Qre1 during the first class of the course and Qre2 at the end of the final class. Qre3 was distributed approximately one year after the end of the course and was initially supplied as an email attachment (2009-10); thereafter it was administered through a link to an on-line survey hosted at the SurveyExpression website (SurveyExpression.com). Student responses were entered into spreadsheets and statistical procedures were carried out using Excel and StatPages.org (https://statpages.info/anova1sm.html).

3.1 Questionnaire Data

Further details about the nature and collection of the questionnaire data will now be considered. Not all students completed all questionnaires and not all students gave responses to all items on the questionnaires. However, in order to provide the fullest possible account of the data, the figures given in this paper cover all responses given to any item discussed.

Qre1 consisted of 19 closed questions; here we focus on prior corpus use. This question was addressed by four items on the pre-course questionnaire, given here as a) to d) below.

a) Had you USED a corpus or corpora before this term's classes? Yes _____ No _____

If students answered affirmatively, they were directed to three further questions:

b) If yes, which corpus/corpora have you used?

c) How often do you use a corpus?

In response to question c), students were invited to choose one of the following options: *several times a day, about once a day, about five times a week, about once a week, about once a month,*

seldom, other. Students who chose the option '*other*' were asked to specify how often they used the corpus. A fourth question asked students to give the purpose for which they used the corpus:

d) What have you used a corpus for?

Students were asked to select one of the following options: *composing written work, revising written work, both composing and revising written work, other*. Students who chose the option '*other'* were asked to specify their purpose in more detail. The responses to question d), together with the name of the corpus used, enabled ELL purposes to be distinguished from other research uses.

Qre2 included 27 open and closed questions; the present paper reports on responses for frequency of autonomous use. The data used to answer RQ2 come from the following item, e), on the immediate post-course questionnaire completed at the end of the last session of the corpus course:

e) How often do you use your corpus outside class?

Students were offered the same options as for question c) in Qre1, with the addition of *'never'*, a choice which was not covered by a prior item on this questionnaire. It should be noted that the form of question e) was designed to obtain only data on students' autonomous use of their own DIY corpus. In-class use and use of other corpora were excluded.

The email version of Qre3 consisted of 12 open and closed questions, while the on-line version was expanded, with 27 open and closed questions. This paper focuses on data concerning the frequency of autonomous use. The form of the questions was virtually identical in the email and online versions of the questionnaire; the two items, f) and g) read as follows:

f) Have you used your own corpus AT ANY TIME since the academic writing course ended?

Yes/No

g) If you answered 'yes', how often do/did you use your own corpus?

The wording of item f) was chosen to reflect the fact that many students had used their corpus regularly for only a limited period of time; they stopped because they were no longer doing any

academic writing. This particularly affected Master's students, who tended to use their corpus while writing their dissertations, but stopped after they graduated as they no longer needed to write academic texts. It was considered important to include data on such respondents, since they will have used their DIY corpora autonomously according to their own needs. The options offered for g) were the same as those for question c) from Qre1. This consistency enabled accurate comparisons to be made between pre-course, immediate post-course and delayed post-course responses.

A summary of the research questions and the questionnaire items used to answer them are provided in Table 1.

Research question	***Questionnaire & item(s)***	***Response required***
RQ1: How prevalent was pre-course corpus use for ELL 2009-2017?	**Qre1 pre-course**	
	a) Had you USED a corpus or corpora before this term's classes?	Yes/No
	b) If yes, which corpus/corpora have you used?	Name of corpora
	c) How often do you use a corpus?	Select 1 option *several times a day, about once a day, about five times a week, about once a week, about once a month, seldom, other*

	d) What have you used a corpus for?	Select 1 option *composing written work, revising written work, both composing and revising written work, other*
RQ2 How prevalent was immediate post-course use of the DIY corpus outside class 2009-2017?	**Qre2 immediate post-course**	
	e) How often do you use your corpus outside class?	Select 1 option *several times a day, about once a day, about five times a week, about once a week, about once a month, seldom, never, other*
RQ3 How prevalent was delayed post-course use of the DIY corpus (i.e. after one year) 2009-2017?	**Qre3 delayed post-course**	
	f) Have you used your own corpus AT ANY TIME since the academic writing course ended?	Yes/No
	g) If you answered 'yes', how often do/did you use your own corpus?	Select 1 option *several times a day, about once a day, about five times a week, about once a week, about once a month, seldom, other*

RQ4 How prevalent was regular corpus use (≥1/week) at the pre-course, immediate post-course and delayed post-course stages 2009-2017?	**Qre1 c), Qre2 e) and Qre3 g) above**	Select 1 option *several times a day, about once a day, about five times a week, about once a week*

Table 1 Summary of research questions and questionnaire items

The number of questionnaires received fluctuated over the years because administrative issues meant that the number of parallel classes varied considerably (ranging from four in 2015 and 2017, to seven in 2011). Given the elective nature of the course, there was a considerable drop-off in the number of questionnaires received between pre- and post-course stages, with 544 responses for Qre1 and 343 for Qre2. Furthermore, contacting participants one year after the course had finished proved challenging, since many students had completed their studies and no longer had university email addresses. This led to a further decrease in the response rate for the delayed post-course questionnaire (Qre3), which amounted to 221. Nonetheless, Qre2 data are available for roughly 60% and Qre3 data for about 40% of Qre1 respondents. Figure 1 shows the number of responses received for each questionnaire by year. These data are sufficient to allow a long-term view of the take-up of corpus use over 2009-2017 to be seen. Thus they provide a novel way of assessing the extent to which the corpus approach was accepted by learners and incorporated into their autonomous language learning practices.

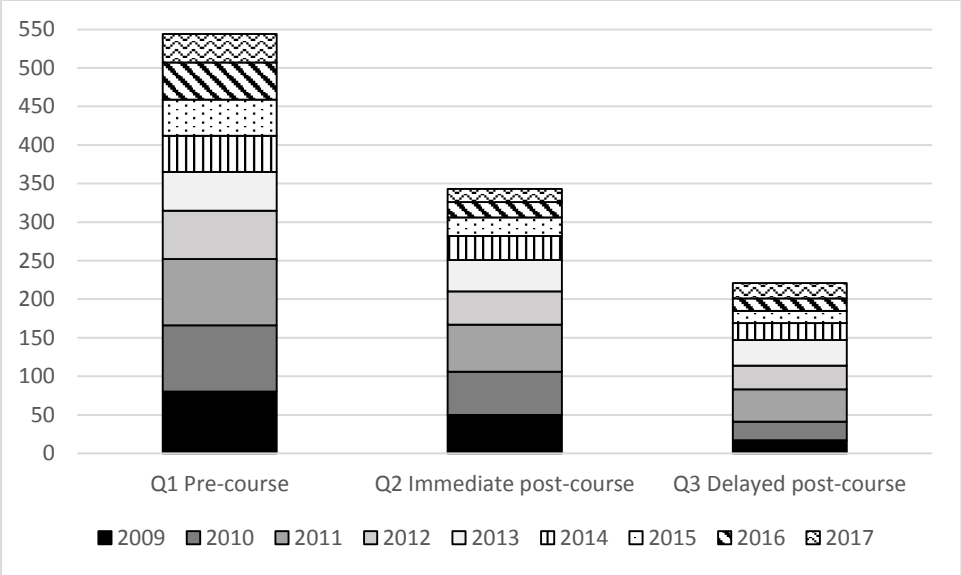

Figure 1 Numbers of questionnaires received by year

3.2 Participant Data

Most of the students took the course during the third term of their first year of graduate studies. In total, doctoral students made up the majority of participants at 57% as against 32% at Master's level and 11% of other status (e.g. postdoctoral or diploma students). These percentages remained stable over the period 2009-2017. Students belonged to over 80 different disciplines, 41% of which were natural sciences, 35% social sciences and 24% humanities. The percentage of natural science students remained fairly constant over the nine years, but that of social sciences showed a tendency to decrease, and that of humanities to increase. Participants reported 51 different L1s, with the most frequent being Chinese (28% of the total), Spanish (9%), German (7%), Italian and Japanese (6%). These five L1s accounted for over 50% of all responses to Qre1 and there was little change in their relative occurrence over the study period. In total, there were 53% female and 47% male participants, with a slight increase in the participation of males during the nine years. The make-up of the participant groups in terms of degree level, discipline, LI and gender showed little variation over the study period.

3.3 Data on students' DIY corpora

Each student built an individual DIY corpus based on the RAs that they had already downloaded and saved for their research purposes. Since these papers were usually in pdf format, they had to be converted to plain text in order to be readable by AntConc (Anthony, 2018). Initially (2009-2014), files were converted individually, but from 2015, the AntFileConverter (Anthony, 2017) became available, which enabled batch conversion of files and greatly speeded up the process. Most corpora consisted solely of RAs in the topic area of the student's research, but those who wished to include other genres or who were working on inter-disciplinary topics, created sub-corpora of different genres or disciplines. Students were advised to select papers from respected journals and to choose a range of different writers. Although tutors explained how to clean the corpus files (by removing items that are not part of the running text e.g. references and tables), most students did no further file preparation and the corpora remained 'dirty'. Although this may mean that the number of words in the corpora is somewhat inflated, these 'quick and dirty' corpora proved adequate to the needs of these students, as shown in their evaluations (Author, 2012). Students were asked to save their corpora on the institutional server and to update them as they added or deleted files; these records provide figures for the size and number of DIY corpora. However, as some students failed to upload their corpora at all, while others did not keep their corpus folders up to date, some figures are likely to be under-estimates. The data are supplemented where available by students' responses to queries from Qre2 and Qre3 asking them to record the size of their corpus. Data on corpus size are available for 333 participants. The number of files in each corpus spans a very wide range from five to 2053, with the majority (46%) consisting of between ten and 19 files. The corpus size in words by the percentage of students is presented in Figure 2.

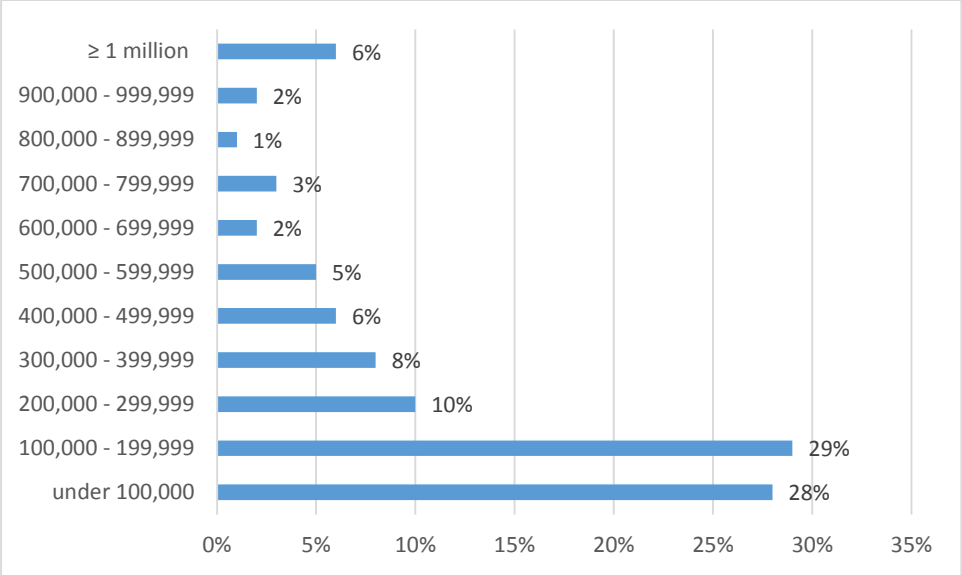

Figure 2 Corpus size in words by percentage of students (n = 333)

As can be seen, the majority of these corpora are very small, with 67% under the 300,000 word size often recommended for a specialised corpus (see e.g. Flowerdew, 2012). While accepting that such small corpora would not be considered adequate for corpus linguistic research, it should be noted that these DIY corpora are specialised not just by discipline, but even more narrowly by the student's research topic. For example, the research of one doctoral student in the discipline of geography/environment concerned the role of transport and travel behaviour in shaping subjective wellbeing and health levels. His corpus consisted of 14 RAs on this topic and amounted to 129,738 words. Such a narrowly focused corpus is likely to contain many examples of lexicogrammatical features relevant to the topic, although it would not be adequate for more general queries in the field of geography/environment. Despite their small size, we would argue that such corpora are valuable in achieving the specific purposes for which they have been compiled. The geography/environment student referred to above found his corpus helpful for language learning, as demonstrated in Qre2 by his strong agreement that using his corpus helped him improve his writing.

4. Results and Discussion

This section is organised by research question (RQ) and includes a discussion of each result in turn.

4.1 RQ1: How prevalent was pre-course corpus use for ELL 2009-2017?

Students were asked whether they had used a corpus before the course and if so, about their frequency and purpose in doing so. The data for RQ1 on all corpus users appear in Figure 3. Based on the 544 responses recorded, it can be seen that the use of corpora for ELL remained stable over the study period with a mean of 24% (SD 5%). This consistent trend shows that there was no increase in acceptance or take-up of the corpus approach among these EAP students, irrespective of their prior ELL background or exposure to corpora. Thus the increased research interest in corpus-based language learning noted by Chen and Flowerdew (2018a) does not seem to have translated into a rise in the take-up of the corpus approach by the students themselves.

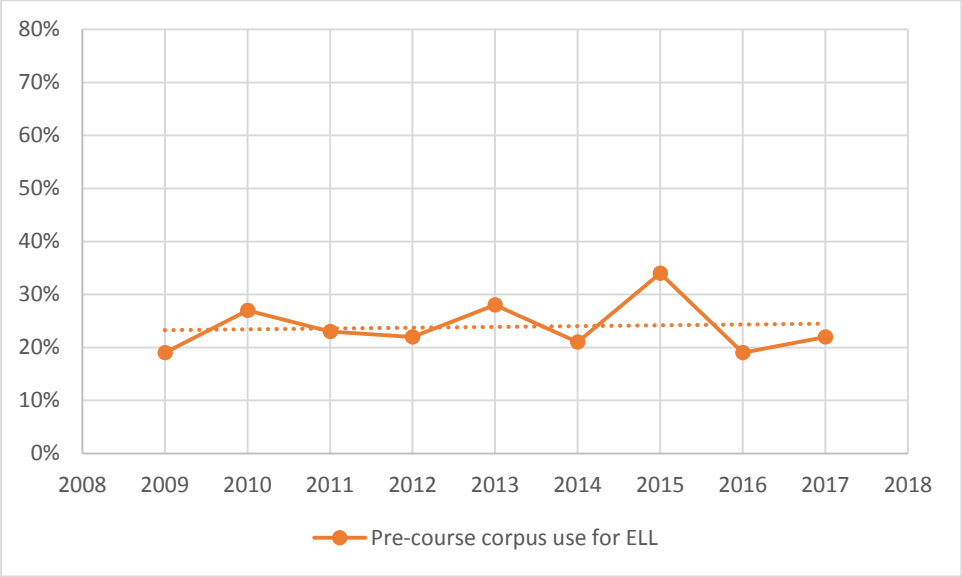

Figure 3 Percentages of students' pre-course corpus use for ELL

In answer to question b) about the corpora they had used, 146 students (27%) mentioned a total of 166 corpus resources, with some students noting more than one. Most students gave the names of well-known resources, including for example, the British National Corpus (76 mentions;

14%), Just the Word (38; 7%) and the Corpus of Contemporary American English (5; 1%). A small number of students (12; 2%) mentioned using a corpus for their own research purposes, with about half of these mentioning Greek and Latin resources (7; 1%). Only a small minority (8; 2%) erroneously mentioned non-corpus resources such as online dictionaries (e.g. The Oxford English Dictionary, Lexico) and other databases (e.g. The Manchester Phrasebank, JSTOR). Thus the majority of the students who reported on the corpora they had used showed an accurate understanding of what a corpus is.

4.2 RQ2 How prevalent was immediate post-course use of the DIY corpus outside class (ie autonomously) 2009-2017?

Students were asked about the frequency of their corpus use outside class. The number of responses received amounted to 343 and Figure 4 shows the results for all users. The data show a high take-up of corpus use overall, which increased only very slightly over the study period, from 84% in 2009 to 88% in 2017, with a peak at 92% in 2015 (Mean 87% SD 2.7%). This suggests that one effect of the corpus course was to stimulate these students to try out the approach autonomously for their own purposes. The fact that this was a consistent effect provides a validation of the efficacy of the DIY corpus approach for these students. However, the slight increase seen here in the adoption of the corpus approach over time provides scant support for the notion that corpus-based work is gaining increasing acceptance by EAP students. In fact, it may reflect a more general increase in students' willingness to make use of technological tools to assist with writing, as noted by Crosthwaite (2017). See Strobl et al. (2019) for a review of such tools.

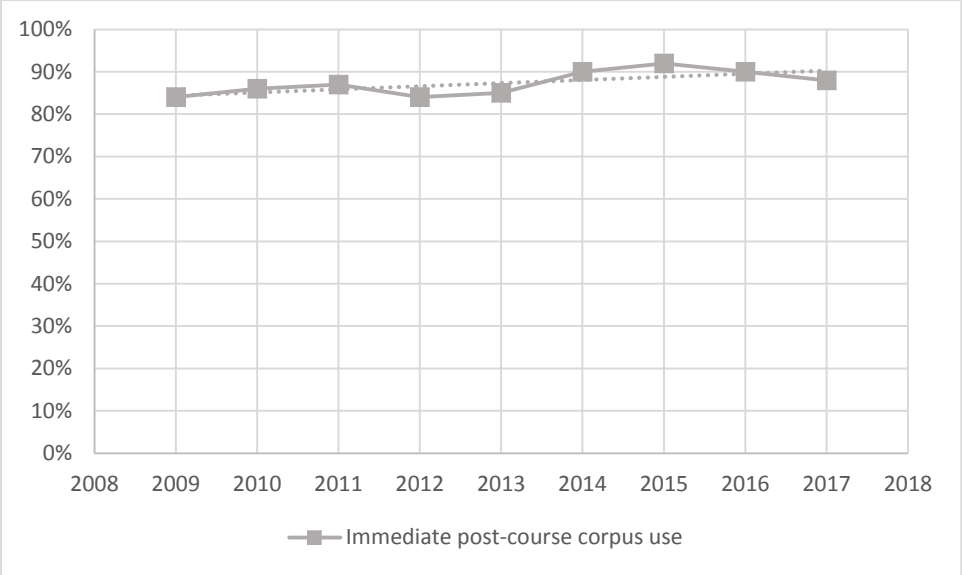

Figure 4 Percentages of students' immediate post-course corpus use outside class

4.3 RQ3 How prevalent was delayed post-course use of the DIY corpus (i.e. after one year) 2009-2017?

In Qre3, students were asked if they had used their corpus at any time since the course and if so, how frequently. There were 221 responses and the data on all users appear in Figure 5. The results show a predominantly stable trend, which rises only slightly over the research period. The mean percentage of users is 62% (SD 11.3%). These results indicate the success of the course in encouraging the adoption of the DIY corpus approach and underline the importance of a narrowly focused discipline- and topic-specific corpus for graduate students. However, they provide little, if any, support for the suggestion that corpus use among students is increasing.

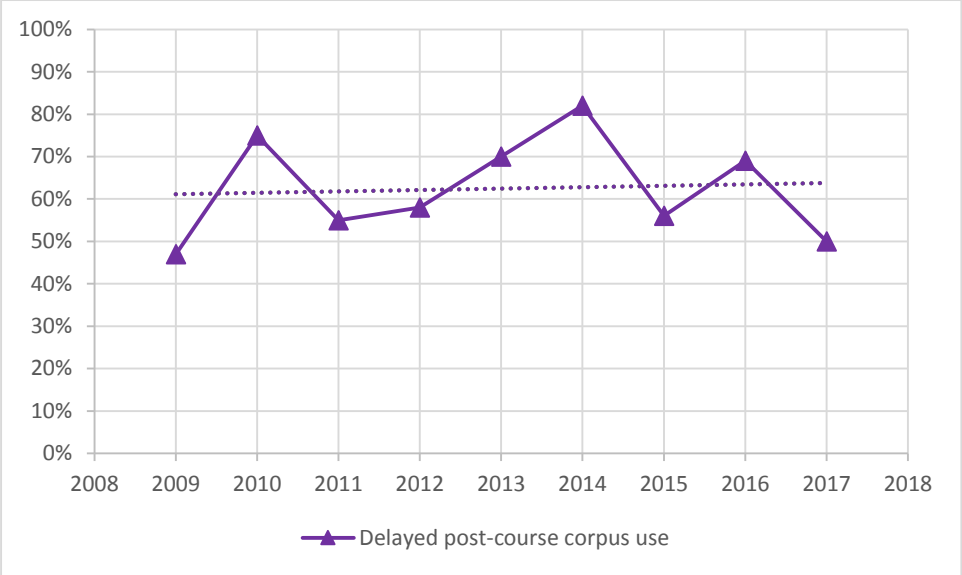

Figure 5 Percentages of students' delayed post-course corpus use

In sum, the results for student corpus use from 2009 to 2017 at pre-course, immediate post-course and delayed post-course do not provide convincing evidence for an upsurge in corpus use by students.

4.4 RQ4 How prevalent was regular corpus use (≥1/week) at the pre-course, immediate post-course and delayed post-course stages 2009-2017?

As noted above, the results given so far cover all use, including students who reported only consulting their corpus occasionally. In order to shed further light on the extent to which corpus use has become embedded as a customary component in students' writing practice, it is necessary to examine regular corpus use, since it is possible that this has indeed increased over time.

A regular user was defined as one who reported using the corpus at least once per week. Accordingly, we compared the three sets of data for regular users at the pre-course, immediate post-course and delayed post-course stages (Figure 6). Results are based on responses to questions c), e) and g) (see Table 1); the data cover the following student responses: *several times a day, about once a day, about five times a week, about once a week*. In Figure 6, the bottom data series shows pre-course data; the middle series gives the delayed post-course data, while the top data series refers to

immediate post-course data. For pre-course respondents a mean value of 11% (SD 4%) was found and the trend remained constant over the nine years. The data on immediate post-course users exhibit a somewhat different pattern: there was a clear upward trend in regular use over the nine years, from 58% in 2009 to 71% in 2017, peaking at 75% in 2016 (Mean 61% SD 6.8%). At the delayed post-course stage, the mean percentage of regular users stands at 37% (SD 10.3%) and there was just a slight rise over the nine-year period. A one-way ANOVA run on these three sets of data shows that all differences are significant at the $p = < 0.05$ level.

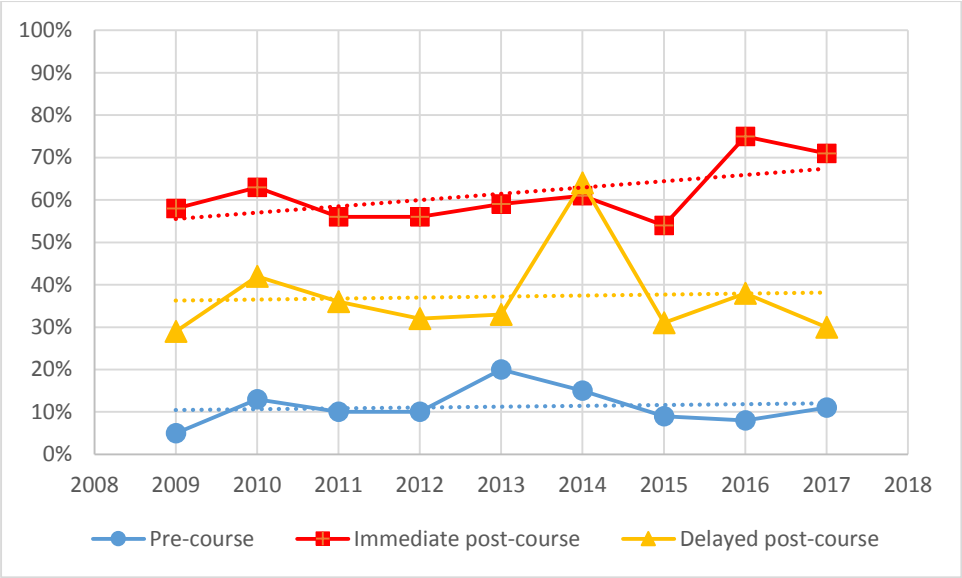

Figure 6 Percentages of students' regular corpus use at pre-course, immediate post-course and delayed post-course stages

Comparing the data from a long-term perspective, it can be seen that the rising trend at the immediate post-course stage is much more pronounced than that at delayed post-course. Thus it is only these immediate post-course data that provide good evidence for the suggestion that there has been an increase in student corpus use over time. One possible reason for this rising trend at the immediate post-course stage might be that it reflects increasing demands upon graduate students over the years and that such pressures may be particularly acute at the end of students' first year of graduate studies. However further research would be necessary to verify this conjecture.

We turn now to individual comparisons, examining first the immediate and delayed post-course data. Results show that there was a considerable reduction in the percentages of those who used their corpus regularly between these two stages. Thus, the very high levels of corpus use achieved immediately after the course were not maintained one year later. There are several possible reasons for this. First it is necessary to consider the academic stage of the student's development, the research tasks they need to accomplish and the type of texts they need to write. In this regard we may distinguish between Master's and doctoral students. Responses to question g), which asked about delayed post-corpus use, would include data from Master's students who had to submit dissertations by the beginning of the subsequent academic year. However, it would exclude responses from those who had submission dates at the end of the summer term, whose data would only appear in the immediate post-course results. This factor may account for some of the discrepancy between regular immediate and delayed post-course use among these students.

The academic situation of doctoral students is rather different. At the end of their first year, they are required to produce an 'upgrade paper', which is a text of considerable length and importance, since its successful completion enables the student to move from probationer to full research student status. Thus, the demands of the research and writing process are likely to lead to regular corpus use reported at the immediate post-course stage. However, in their second year, many doctoral students are engaged in field or experimental work, which is likely to mean a reduction in the amount of academic writing carried out and thus less need for corpus consultation at the delayed post-course stage.

Of course, in accounting for the difference between regular use reported in the immediate and delayed post-course questionnaires, other factors may also come into play: the novelty effect of a new learning resource may have worn off and it is possible that the lack of expert advice and weekly practice make it less likely that students will persevere with corpus use. However, despite these disincentives, the mean percentage of regular users at the delayed post-course stage (37%)

still shows that a considerable number of students find regular corpus consultation a valuable addition to their writing techniques and tools.

Moving on to a comparison of regular use at the pre-course stage with the data at immediate and delayed post-course stages, we see that the mean percentage is considerably higher at both post-course stages (61% and 37% respectively) than the pre-course mean of 11%. This shows the success of the course in encouraging autonomous DIY corpus use and underlines the importance of an individualised discipline- and topic-specific corpus for graduate students.

Finally, when we compare the data for regular users (Figure 6) with those for all users (Figures 3, 4 and 5), it is clear that there is a considerable disparity, which persists throughout the nine years. Some discrepancy is to be expected, but it is worth noting that the reduction in use amounts to a large number of potential users who decreased corpus consultation subsequent to the immediate post-course stage.

Further evidence of this decline in use can be seen in Table 2, which presents the mean percentages of non-use, all use and regular use. When compared to the immediate post-course stage, the mean percentage of both all use and regular use decreased, while that for non-use increased substantially from 6% to 39%, an indication that students have discontinued corpus use. The differences in the group means as determined by one-way ANOVA were found to be statistically significant ($p = < .001$).

Stage	Mean % of non-use for ELL (SD)	Mean % of all ELL use (SD)	Mean % of regular ELL use (SD)
Pre-course	76% (SD 5%)	24% (SD 5%)	11% (SD 4%)
Immediate post-course[1]	6% (SD 5%)	87% (SD 2.7%)	61% (SD 6.8%)
Delayed post-course	39% (SD 10%)	62% (SD 11.3%)	37% (SD 10.3%)

Table 2 Mean percentages of non-use, all use and regular use at pre-course, immediate post-course and delayed post-course stages

[1] Figures do not include unclear responses (7%).

As noted earlier in this section, there are many possible factors which could lead to a decline or discontinuation of corpus use, including dissertation submission dates and the student's academic stage and activities. Although the mean percentages of regular users still represent a substantial number of students who are committed to using their corpora, the disparities between non-use, all corpus use and regular use are still a cause for concern. One possible reason for the decline seen in corpus use is that some students have already developed a writing procedure which is successful for them and which they do not wish to modify (Author, 2014). Student motivation is another factor that may well have an impact (Liou & Liu, 2021). It may also be that a corpus approach is not well-suited to the learning preferences of certain students; for example, Dudley Evans and St John (1998, p. 208) argue that 'technophiles' are more likely to benefit than 'technophobes', while Boulton (2009) suggests that learning style preferences may affect DDL use. It is likely, then, that a combination of multiple factors, personal, academic and contextual, comes into play when students decide whether to use a corpus or not. Further research is necessary to clarify the reasons for the observed decline in corpus use and to identify the traits that characterise the regular user.

5. Pedagogical implications

This research leads both to a set of specific recommendations that can be made for the type of course provision that is likely to benefit graduate students, as well as to wider implications for the adoption of corpus work in EAP more generally.

As Kwan (2013) points out in the Hong Kong context, many doctoral students have access to little or no writing support during their studies, a situation which may well also affect Master's students. Given such limited help, the long-term trend results presented here suggest that both doctoral and Master's students would benefit from a DIY corpus course towards the end of their first year of graduate studies. One way in which such corpus work could be integrated into graduate

studies is through the use of on-line platforms such as those described by Crosthwaite et al. (2019) and Cotos (2016). Such systems allow students ready access to the resources they require at the point of need, an important consideration for busy graduates.

After an initial introductory corpus course, the needs of doctoral and Master's students diverge, which implies that different follow-up approaches are necessary for the two types of students. Master's students would find it helpful to have extended practice in using large online corpora, which would enable them to address any later, more general language learning needs. Doctoral students, on the other hand, would profit from short just-in-time refresher courses on using their DIY corpora as they move into the 'writing-up' phase of their studies and particularly as they prepare their first academic research articles as early career researchers. Online support groups could also play an important role in re-activating doctoral students' knowledge of corpus use so that they attain or exceed the level of corpus competence achieved at the end of the original DIY corpus course.

However, such interventions, although necessary, may not be sufficient to persuade students to adopt the new writing practice of using corpora. One strategy which might have some success, especially with graduate students, who are being trained to evaluate evidence and arguments, could be to introduce them to some of the relevant empirical data which supports the claim that corpus pedagogy is successful, for example by presenting selected findings from Boulton & Cobb (2017) or Pérez-Paredes (2019). Another strategy which could complement the discussion of data is suggested by Rogers (2003) in his work on the diffusion of innovations. He notes the importance of success stories in persuading users to adopt novel behaviours. Instructors could adopt such an approach, encouraging current and former students to recount their successes to their peers, thereby opening up the corpus class to a discussion of difficulties overcome, writing improvements endorsed by supervisors and theses/dissertations successfully completed. Incorporating elements of both data and stories occasionally into corpus classes might well prove motivating to novice and/or hesitant corpus users. In the end, however, it is practical considerations of utility and particularly

the individualised discipline- and topic-specific nature of the DIY corpus that is most likely to drive students' continued use of their corpus.

A further pedagogical suggestion emerges from the observation that at all stages (pre-course, immediate and delayed post-course) there is a wide and persistent discrepancy between all corpus users and regular users. This points to the existence of a sizeable number of students who use the corpus only occasionally or not at all. It is noticeable that this gap remains considerable throughout the nine years of the study. Similar student groupings have been noted by Kaszubski (2010) who termed them 'adopters' and 'minimalists' and by Crosthwaite et al. (2019) who called them 'persistent users' and 'search gurus'. If we seek to argue that using a DIY corpus would benefit the majority of graduate students, then we need to ask why some students become regular users, while others remain occasional users or non-users, 'refusers' in Kaszubski's terms or 'quitters' for Crosthwaite et al.

Rather than employing such simple three-part categorisations, however, it may be more enlightening to consider Bax's (2003, pp. 24-25) seven stages in the adoption of CALL as a means of better understanding the possible process that students may be experiencing (Table 3).

Stage	Explanation
1 Early adopters	A few adopt it out of curiosity
2 Ignorance & scepticism	Most ignorant or sceptical
3 Try once	Tried, but rejected due to early problems; no relative advantage
4 Try again	Tried again, its relative advantage is recognised
5 Fear & awe	More use, but fear alternating with exaggerated expectations
6 Normalising	Gradually seen as normal
7 Normalisation	So integrated it is invisible, 'normalised'

Table 3 Stages in the adoption of CALL

Most EAP students, who have had no exposure to the use of corpora at all, are likely to be at stages 1 and 2. However, after their six-week corpus course, the students in the present study can be situated at stages 3-6, depending on their level of use. Some non-users are at stage 3, rejecting corpus use as too time-consuming or not as useful as Google, while others are simply not doing any writing that warrants the use of their DIY corpus. Occasional users are likely to be distributed between stages 4-6, and regular users are probably at stage 6, perceiving corpus use as a normal procedure. Indeed for some, corpus use may already have become fully normalised. The pattern of adoption of DIY corpus use seen in these students supports the point made by Rogers (2003) and Bax (2003, 2011) that the diffusion of new technologies tends to be uneven in populations. This poses particular challenges for instructors who wish to promote corpus use, because it implies that different strategies are likely to appeal to different groups of students, who could be within a single class, but at varying stages of adoption. We consider therefore that instructors should attempt to take a more individualised and flexible approach as far as possible; multiple strategies should be at their disposal, which can be applied to different groups at different times and in different pedagogical contexts. This view is also advocated by Ackerley (2021) in relation to undergraduate corpus users.

Finally, there is a more general point to be made regarding the place of technology, in this case corpus use, and the process of its normalisation within EAP and more broadly within language learning. As Bax (2011, p. 13) argues, ‘normalisation depends on far more than the attributes of the technology itself… it involves a host of social and cultural elements operating together in complex ways.’ This statement implies the necessity of a shift in emphasis away from a narrow concentration on the technology and tools themselves and towards a wider focus on the overall context of learning. Thus, although improving the technological tools, i.e. corpus construction and text analysis software, is undoubtedly important, it does not necessarily follow that the existence of better tools will lead to an increase in more regular and committed corpus users. For this to occur, instructors and designers of software and corpora need constantly to bear the socio-cultural context

in mind, not only by asking what technology the context requires, but more radically whether there is a need for technology in this context at all. Such a re-orientation in our thinking about the pedagogical applications of corpora has been advocated by several authors, (Author, 2021; Chambers, 2019; Frankenberg-Garcia, 2012, 2016; Römer, 2009). It is the needs of the student end-user that must always be at the centre of future developments.

6. Conclusions

To summarise, this paper has traced the development of graduates' autonomous use of their DIY corpora over nine years from 2009 to 2017. Taking a long-term perspective, it showed that prior to the corpus course, corpus use for ELL was stable, with all use at around 24% and regular use at around 10% of students. On completion of the course, immediate use was stable at 80-90% of students, while regular use took an upward trajectory, increasing from 58% to 71% from 2009 to 2017. Finally, data recorded one year after the end of the course showed that 62% of students had used their corpus since the course finished, while 37% were regular users, trends which showed only a slight rise over the nine year period. Although these findings reveal some grounds for optimism, the lack of significant growth in corpus use over the nine years does not support the expectation that there will be increasing take-up in the future. There is still much work to be done in order to achieve this goal.

This study has a number of limitations which need to be borne in mind when interpreting the findings. First, the number of questionnaires completed showed a substantial drop-off from pre-course through immediate post-course to the delayed post-course stage. While some reduction in participation is probably unavoidable due to the one-year time span of the study for each cohort, it would be particularly valuable if future research could achieve more consistent participation rates across the three stages. Second, the present study provides data on students from a single course at one university in the UK; sociocultural differences, especially in the pedagogical context of courses mean that generalisations should be made with caution. In order to gain a fuller picture of

autonomous graduate take-up of the corpus approach over time, it is necessary to carry out further research on a range of different graduate corpus courses in a number of institutions and countries. Finally, the course under study involved graduates in building their own individualised DIY corpora; it would be useful to know whether the results shown here are replicated when different types of corpora are used. In terms of assessing and possibly improving the level of corpus adoption, it would also be valuable to examine the impact of other variables such as the timing and length of the course, the amount of instructor support available and the target writing assignments. Taking a long-term view of such attributes of corpus pedagogy would provide a new perspective on the development of student corpus use, thereby enabling instructors to tailor their interventions to the needs of their students more closely and with greater confidence.

Such research, though, will not of itself lead to greater corpus use by students in the future. There is a pressing need for those who are committed to the corpus approach to redouble their efforts to reach out directly to both students and their teachers. To realise the optimistic vision of corpus use as a mainstream activity will require active engagement within the everyday practices of EAP.

References

Author (2012).

Author (2014).

Author (2015).

Author (2021).

Author (in press).

Ackerley, K. (2021). Exploiting a genre-specific corpus in ESP writing: Students' preferences and strategies. In Author.

Anthony, L. (2017). *AntFileConverter* (Version 1.2.1) [Computer software]. Tokyo, Japan: Waseda University. http://www.laurenceanthony.net

Anthony, L. (2018). *AntConc* (Version 3.5.7) [Computer software]. Tokyo, Japan: Waseda University. http://www.laurenceanthony.net

Bax, S. (2003). CALL - past, present and future. *System*, *31*(1), 13–28. https://doi.org/10.1016/s0346-251x(02)00071-4

Bax, S. (2011). Normalisation revisited: The effective use of technology in language education. *International Journal of Computer-Assisted Language Learning and Teaching*, *1*(2), 1–15. https://doi.org/10.4018/ijcallt.2011040101

Boulton, A. (2009). Corpora for all? Learning styles and data-driven learning. In M. Mahlberg, V. González-Díaz, & C. Smith (Eds.), *Proceedings of the Corpus Linguistics Conference CL2009*.

Boulton, A. (2010). Learning outcomes from corpus consultation. In M. Moreno Jaén, F. Serrano Valverde, & M. Calzada Pérez (Eds.), *Exploring new paths in language pedagogy: Lexis and corpus-based language teaching* (pp. 129–144). Equinox.

Boulton, A. (2017). Corpora in language teaching and learning. *Language Teaching*, *50*(4), 483–506. https://doi.org/10.1017/S0261444817000167

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

Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching*, *52*(4), 460–475. https://doi.org/10.1017/S0261444819000089

Chang, J.-Y. (2014). The use of general and specialized corpora as reference sources for academic English writing: A case study. *ReCALL*, *26*(2), 243–259. https://doi.org/10.1017/S0958344014000056

Chen, M., & Flowerdew, J. (2018a). A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. *International Journal of Corpus Linguistics*, *23*(3), 335–369. https://doi.org/10.1075/ijcl.16130.che

Chen, M., & Flowerdew, J. (2018b). 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

Cortes, V. (2007). Genre and corpora in the English for academic writing class. *ORTESOL Journal*, *25*, 9–16.

Cotos, E. (2016). Computer-assisted research writing in the disciplines. In S. A. Crossley & D. S. McNamara (Eds.), *Adaptive Educational Technologies for Literacy Instruction* (pp. 198–210). Routledge.

Cotos, E., Link, S., & Huffman, S. (2017). Effects of DDL technology on genre learning. *Language Learning and Technology*, *21*(3), 104–130. http://llt.msu.edu/issues/october2017/cotoslinkhuffman.pdf

Cresswell, A. (2007). Getting to 'know' connectors? Evaluating data-driven learning in a writing skills course. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 267–287). Rodopi. https://doi.org/10.1163/9789401203906_018

Crosthwaite, P. (2017). Retesting the limits of data-driven learning: Feedback and error correction. *Computer Assisted Language Learning*, *30*(6), 447–473. https://doi.org/10.1080/09588221.2017.1312462

Crosthwaite, P., Wong, L., & Cheung, J. (2019). Characterising postgraduate students' corpus query and usage patterns for disciplinary data-driven learning. *ReCALL*, *31*(3) 255–2275. https://doi.org/10.1017/S0958344019000077

Dudley-Evans, T., & St John, M.-J. (1998). *Developments in English for specific purposes: A multi-disciplinary approach*. Cambridge University Press.

Flowerdew, L. (2012). *Corpora and language education*. Palgrave Macmillan. https://doi.org/10.1057/9780230355569

Flowerdew, L. (2015). Data-driven learning and language learning theories: Whither the twain shall meet. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language*

corpora for data-driven learning (pp. 15–36). John Benjamins. https://doi.org/10.1075/scl.69.02flo

Frankenberg-Garcia, A. (2012). Integrating corpora with everyday language teaching. In J. Thomas & A. Boulton (Eds.), *Input, process and product: Developments in teaching and language corpora* (pp. 36–53). Masaryk University Press.

Frankenberg-Garcia, A. (2016). Corpora in the classroom. In G. Hall (Ed.), *Routledge handbook of English language teaching* (pp. 383–398). Routledge.

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, *32*, 301–319. https://doi.org/10.1016/j.system.2004.04.001

Holec, H. (1979). *Autonomy and foreign language learning*. Pergamon Press.

Huang, L.-S. (2011). Language learners as language researchers: The acquisition of English grammar through a corpus-aided discovery learning approach mediated by intra- and interpersonal dialogues. In J. Newman, H. Baayen, & S. Rice (Eds.), *Corpus-based studies in language use, language learning and language documentation* (pp. 89–122). Rodopi. https://doi.org/10.1163/9789401206884_007

Kaszubski, P. (2011). IFAConc—A pedagogic tool for online concordancing with EFL/EAP learners. In A. Frankenberg-Garcia, L. Flowerdew, & G. Aston (Eds.), *New trends in corpora and language learning* (pp. 81–104). Continuum. https://doi.org/10.5040/9781474211925.ch-005

Kwan, B. S. C. (2013). Facilitating novice researchers in project publishing during the doctoral years and beyond: A Hong Kong-based study. *Studies in Higher Education*, *38*(2), 207–225. https://doi.org/10.1080/03075079.2011.576755

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

Lee, H., Warschauer, M., & Lee, J.H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, *40*(5), 721–753. https://doi.org/10.1093/applin/amy012

Liou, H.-C., & Liu, S.-Y. (2021). Exploring the relationships between English writing motivation and uptake of corpus-aided corrective feedback: A longitudinal study. In Author

Park, K., & Kingerer, C. (2010). Writing/thinking in real time: Digital video and corpus query analysis. *Language Learning and Technology*, *14*(3), 31–50. http://llt.msu.edu/vol14num3/parkkinginger.pdf

Pérez-Paredes, P. (2019). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning*, 1–26. https://doi.org/10.1080/09588221.2019.1667832

Pérez-Paredes, P., Sánchez-Tornel, M., Alcaraz Calero, J., & Aguado Jimenez, P. (2011). Tracking learners' actual uses of corpora: Guided vs non-guided corpus consultation. *Computer Assisted Language Learning*, *24*(3), 233–253. https://doi.org/10.1080/09588221.2010.539978

Pérez-Paredes, P., Sánchez-Tornel, M., & Alcaraz Calero, J. (2013). Learners' search patterns during corpus-based focus-on-form activities. *International Journal of Corpus Linguistics*, *17*(4), 482–515. https://doi.org/10.1075/ijcl.17.4.02par

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.

Römer, U. (2009). Corpus research and practice: What help do teachers need and what can we offer? In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 83–97). John Benjamins. https://doi.org/10.1075/scl.33.09rom

Strobl, C., Ailhaud, E., Benetos, K., Devitt, A., Kruse, O., Proske, A., & Rapp, C. (2019). Digital support for academic writing: A review of technologies and pedagogies. *Computers & Education*, *131*, 33–48. https://doi.org/10.1016/j.compedu.2018.12.005

Tribble, C. (2015). Teaching and language corpora: Perspectives from a personal journey. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 37–62). John Benjamins. https://doi.org/10.1075/scl.69.03tri

Yoon, C. (2016b). Individual differences in online reference resource consultation: Case studies of Korean ESL graduate writers. *Journal of Second Language Writing*, *32*, 67–80. https://doi.org/10.1016/j.jslw.2016.04.002

Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning and Technology, 12*(2), 31–48. https://doi.org/10125/44142

Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning and Technology, 12*(2), 31–48. https://doi.org/10125/44142

Vitae

Maggie Charles taught EAP at Oxford University, UK for many years. Her research interests lie in the analysis of academic discourse and corpus use in EAP writing pedagogy. She has published widely in these fields, recently co-editing the volume *Corpora in ESP/EAP Writing Instruction* with Ana Frankenberg-Garcia (Routledge, 2021).

Gregory Hadley is a Professor of Cultural Studies and Applied Linguistics at Niigata University, Japan. A Visiting Fellow at Oxford University, he is the author of *English for Academic Purposes in Neoliberal Universities: A Critical Grounded Theory* (Springer, 2015) and *Grounded Theory for Applied Linguistics: A Practical Guide* (Routledge 2017).

Author statement

Declarations of interest: none

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.