Lidil

Revue de linguistique et de didactique des langues

58 | 2018

**L'enseignement et l'apprentissage de l'écrit
académique à l'aide de corpus numériques**

The BAWE Quicklinks Project: A New DDL Resource for University Students

Le projet « BAWE Quicklinks » : une ressource nouvelle de DDL pour les étudiant.e.s d'université

Benet Vincent and Hilary Nesi

Electronic version

URL: <http://journals.openedition.org/lidil/5306>

DOI: 10.4000/lidil.5306

ISSN: 1960-6052

Publisher

UGA Éditions/Université Grenoble Alpes

Printed version

ISBN: 978-2-37747-064-8

ISSN: 1146-6480

Electronic reference

Benet Vincent and Hilary Nesi, « The BAWE Quicklinks Project: A New DDL Resource for University Students », *Lidil* [Online], 58 | 2018, Online since 02 November 2018, connection on 01 May 2019. URL : <http://journals.openedition.org/lidil/5306> ; DOI : 10.4000/lidil.5306

This text was automatically generated on 1 May 2019.

© Lidil

The BAWE Quicklinks Project: A New DDL Resource for University Students

Le projet « BAWE Quicklinks » : une ressource nouvelle de DDL pour les étudiant.e.s d'université

Benet Vincent and Hilary Nesi

Acknowledgements

We would like to acknowledge the helpful and constructive comments provided by the anonymous reviewers of this paper as well as comments from attendees at the BAAL Corpus Linguistics Special Interest Group event “New Directions in DDL” held at Coventry University on 8 June 2018. We are also grateful to John Higgins, a former colleague of Tim Johns, and pioneer in DDL (see e.g. Higgins, 1991) whose presentation at the event drew our attention to DDL resources and literature we were not previously aware of.

1. Introduction

- 1 Data-driven learning (DDL) is a method of language teaching and learning which encourages learners to investigate corpus data and apply it in their own language learning. While the use of corpora in language teaching has a history that reaches back to the late 1960s (McEnery & Wilson, 1997; Chambers, 2010), DDL as an approach was formulated by Tim Johns in the 1980s to help international university students improve their academic writing. Since then, it has been used in a variety of ways in an attempt to meet the needs of students (and teachers) in a range of contexts and with different levels of access to hardware, software and corpora.
- 2 Tim Johns did not typically expect his students to manage their corpus investigations on their own; he used DDL in one-to-one consultation sessions where the tutor and student worked together to resolve specific English for Academic Purposes (EAP) writing problems and in his “reactive” classroom materials which responded to student questions

(Johns, 1986, 1991a, 1991b). Since then, however, some approaches to DDL have treated it as a more autonomous activity (e.g. Bernardini, 2004). This has the advantage of freeing learners to explore corpus resources wherever and whenever they need to, but it places heavy demands on them, requiring the formulation of appropriate corpus queries and the interpretation of the results therefrom. For this reason, this “hard” form of DDL (Gabrielatos, 2005) has tended to be confined mainly to higher education contexts where learners have advanced language knowledge and good study skills.

- 3 Johns reported the results of some of his DDL consultation sessions and answers to student queries in a series of “kibbitzers”.¹ Each kibbitzer outlines an authentic writing issue, such as the difference between *by* and *using*, and presents a solution (and in some cases activities) based on corpus evidence. This kind of approach has been described as “soft” DDL (Gabrielatos, 2005), in that kibbitzers do not require access to a corpus, and are accessible to learners even if they have no corpus query skills. Some EAP practitioners and materials writers have continued to develop this soft approach, because it requires fewer resources and places fewer demands on learners, whether working alone or in a larger class without one-to-one support (Boulton, 2010). “Soft” corpus-informed lessons can be created relatively easily by printing out some relevant data, for example in the form of concordance lines, and devising accompanying paper-based tasks (see, for example, Tribble & Jones, 1990; Johns, 1991a, 1991b). Nevertheless, the creation of such materials requires both corpus knowledge and sensitivity to student needs; students also need to be fairly carefully introduced to such tasks and activities to avoid potential frustration or bafflement (Quinn, 2015).
- 4 Corpora and corpus-compilation processes have of course greatly changed since Johns first developed DDL. He used very small homemade corpora, mostly derived from newspapers and popular science magazines, and a simple non-networked programme that generated Keyword in Context (KWIC) concordance lines but little else. Most current corpora are hundreds of times bigger than the ones Johns used, making it possible to ascertain usage patterns for rarer language items that do not occur sufficiently frequently in small corpora. Many general corpora are now available online via software which offer a range of options to search for lexico-grammatical patterns as well as single words or word combinations. Online publishing and the resulting increased availability of computer-readable text have also made it easier for teachers and learners to compile their own corpora. EAP practitioners such as Charles (2007, 2011, 2015), Lee & Swales (2006) and Yoon (2011), for example, have asked students to compare language choices made in their own writing with those made by professionals, using research articles gathered from discipline-specific online journals.
- 5 However, even though DDL can be approached in a variety of ways to suit different learning contexts, it still continues to meet with resistance from teachers and learners. This resistance is probably due to the difficulty of finding the right balance between individual freedom and supportive guidance. Autonomous approaches encourage learners to consult corpus resources to solve their own individual production problems as the need arises. Without a teacher there to advise them, however, they may not know exactly what to look for in the corpus, and they may have difficulty working with complex corpus interfaces such as those provided by BNCweb and Sketch Engine (see, for example, Pérez-Paredes, Sánchez-Tornel & Alcaraz Calero, 2013). In contrast, a more guided approach to DDL is less likely to be tailored to individual needs, and may be more limited in terms of both quantity and quality because most learners do not have

permanent one-to-one access to a corpus-literate language tutor. Kibbitzers and classroom activities provide experience in, and help with, the interpretation of corpus data, but usually address a small number of common problems that are not necessarily experienced by every learner. Paper-based DDL activities also reduce the opportunity for the “serendipitous learning” that becomes possible when interacting freely with a corpus using corpus query tools (Bernardini, 2000, 2004).

- 6 The success of all types of DDL must also depend on the appropriacy of the chosen corpus (Aston, 1997; Krishnamurthy & Kosem, 2007). Multi-million word general corpora can generate an overwhelming amount of data in response to a simple query, much of which may not be relevant to an EAP learner’s personal academic context. Resources such as *SkELL*,² *Just the Word*³ or *Word and Phrase*⁴ are more user-friendly, but also tend to draw on general corpora, and therefore do not provide information on the discipline-specific academic usage that is useful to EAP learners. For some learners, in particular postgraduates, small, homemade corpora compiled from academic journals may be a good choice for DDL activities, but research articles are not always the best model for EAP learners when they are aiming to produce other academic genres, such as essays, case studies or reflective pieces (Durrant, 2013). Certain lexical items and turns of phrase are appropriate in some genres but not in others.
- 7 This paper reports on a DDL initiative which aims to build a database of hyperlinks to concordances, providing those who teach university students with a resource that the students can use to help them overcome problems of expression. The aim is also that students will be encouraged to interact with corpus data autonomously, in their own time and at their own pace. We also report on some initial responses from a small group of EAP students whose work has been corrected using the links.

2. Procedure for creating the DDL materials

2.1. Inspiration for the project

- 8 The technique adopted for this initiative was developed for *Writing for a Purpose* (WfAP), a collection of online EAP writing resources created in association with the British Council and available on their “Learn English” website (see Nesi, Gardner & Kightley, 2015). The idea in WfAP is to illustrate important features of student academic genres and thereby provide writing guidance to EAP learners. One aspect of this guidance is to demonstrate typical phraseologies of specific genres by using concordances drawn from samples of relevant texts. The corpus used was the British Academic Written English (BAWE) corpus, which contains over 6.5 million words of proficient student writing in a wide range of genres and disciplines (Nesi & Gardner, 2012⁵). This corpus was an obvious choice for two main reasons. The first of these is that the written genres WfAP focuses on were the ones identified in the BAWE corpus project, and were therefore relevant to undergraduates studying in a wide range of disciplines; some BAWE assignment genres (e.g. research reports and literature surveys) are intended as preparation for research at higher academic levels, but by no means all university students go on to pursue academic careers, and other assignment types (e.g. case studies and problem questions) aim to introduce students to the type of writing they will have to produce in other walks of life. A second compelling reason for using BAWE is that, unlike other possible choices—the most obvious being the Michigan Corpus of Upper-Level Student Papers—this corpus is

open access on Sketch Engine, meaning that it is possible to make use of the powerful query language and sophisticated filtering of results to create lines like those shown in Figures 1 and 2.⁶ Moreover, Sketch Engine provides an easy-to-use link creation facility (“get short permanent link”) which allows users to make hyperlinks and direct learners to corpus outputs—concordance, frequency lists, etc.—exemplifying a particular language feature.

Figure 1. – KWIC concordance lines used in *Writing for a Purpose* to illustrate conclusion patterns in Humanities essays.

beginning of the play. Again, in the real world of this age, **it can be assumed** that it is normally men who are predominant and make million slave imports during the same period.” Therefore, **it can be surmised** that they were being distributed elsewhere. Nevertheless, it would have been found. </p><p> Ibid., p. 54 </p><p> Thus, **it can be concluded** that, during an era defined by economic and : 28). If the injection does result in temporary paralysis **it can be concluded** that the side injected is dominant in language function: an endorsement of it’. If it was a Leveller proposal, then **it can be assumed** that it would have been suggested far earlier in the debate were real alternatives to a Church service’, from which **it can be inferred** that religious practice did not appeal to a minority of

- 9 Using this permalink facility, some parts of the *Writing for a Purpose* resources refer learners to carefully selected concordance outputs from the BAWE corpus which show lexical and phraseological features typical of particular types of assignment and with specific rhetorical functions. For example, the pattern *it can / could be assumed / inferred / concluded that*⁷ is common in the concluding sections of essays in the Humanities; to illustrate this pattern, a hyperlink was created to take learners to the concordance lines from the BAWE corpus shown in Figure 1. Another example is the pattern *it is important / recommended / suggested / vital that*, which is frequently used by Business students in making recommendations in Case Study assignments. Learners interested in viewing realisations of this pattern can click on a link which takes them to the lines shown in Figure 2. In both cases, these lines were the result of a painstaking query construction and then concordance filtering procedure to ensure not only that the lines in question were relevant, but also that there were not so many of them that learners would be overwhelmed.

Figure 2. – KWIC concordance lines used in *Writing for a Purpose* to illustrate Business case study recommendation patterns.

past. However, since it is their jobs that will be changing,	it is important that	there is no negative reaction to the project. The changes i
formal influence over the changes made. However, since	it is important that	the firm meets it needs they should be kept well informed
when the lifespan increases to ten years. Therefore,	it is important that	an accurate estimation of the duration of each investment
, operation and eventual project outcome. Therefore,	it is important that	the key project stakeholders are identified and
: <p> Having decided on the most profitable product mix,	it is important that	the firm now develops a clear production schedule in order
across the full range of its services is AOL. However,	it is important that	Yahoo focuses its attention not only on like-for-like
t implementation should take six months. </p><p> Hence,	it is recommended that	the project pace should be of a medium one. This shall allc
by £ 1 per metre of standard width cloth equivalent,	it is recommended that	a negotiation with the supplier should be done in order to
all levels but to a limited operational area) slice. For CFS	it is recommended that	a progressive pattern with a vertical slice is used because
in advance whether it is worth considering. However	it is recommended that	other methods, such as Payback, ARR are used at the same
not directly encourage continuous improvement. Hence,	it is recommended that	there should be a system for learning and continuous
eventually results in diminished performance. Hence,	it is recommended that	sales and manufacturing units are treated as follows; </p>
duction to ensure units' performance. </p><p> Therefore,	it is suggested that	manufacturing unit is treated as a pseudo-profit centre.
whether loyal or satisfied customers are profitable,	it is suggested that	the manager may use 'average duration of customer
the management group. For further and concrete analysis,	it is suggested that	interview should be conducted with lower level manager
ve been somewhat passive. </p><p> However in this case	it is vital that	they are well-informed of the changes that will be taking

- 10 Learners who are directed to these sorts of corpus outputs are being encouraged to notice for themselves (Rutherford, 1987; Schmidt, 1990) how the typical realisations that are associated with the meanings involved are patterned in ways that may not be simply summarised by an abstract “rule” (Boulton & Cobb, 2017). That is, their attention can be drawn to the types of verbs that occur in the frame *it can be* that*, or to the types of words/constructions that precede it (Fig. 1). This means of presentation via authentic

examples may give learners the confidence to use the phraseologies themselves as well as helping them spot when a writer is drawing a conclusion.

- 11 This use of hyperlinks to specially selected concordance lines to illustrate common patterns, as in Figures 1 and 2, need not of course be confined to exemplifying particular rhetorical functions. This facility can also be used to help learners address issues in their own writing as and when they occur by inserting hyperlinks and directing students to concordance lines showing more appropriate usage. By writing the corpus queries ourselves and selecting relevant lines from those retrieved, we can circumvent the perennial barriers to effective student use of corpora: lack of knowledge of complex corpus query language (CQL), and lack of awareness of phraseological features. This approach can also be of use to EAP practitioners without corpus expertise, since we can share hyperlinks used to address common student writing issues.
- 12 The exploitation of hyperlinks to give students written guidance is suggested as a means of feedback by Hyland (2003) and explored in more detail by Gaskell and Cobb (2004), who investigated the efficacy of inserting hyperlinks to concordances to guide students to correct grammar errors. The concordances used in Gaskell and Cobb's (2004) study were retrieved from the Brown corpus using the Lextutor interface;⁸ they were interested in the extent to which students benefited from the intervention and whether students then went on to use the concordancer independently. Our proposal, we feel, improves on Gaskell and Cobb's approach in a number of ways. Firstly, the issues we address are phraseological rather than simply grammatical; they are generally what Ferris (2002) refers to as "untreatable" in that one cannot simply look up the answer in a reference work. Secondly, the Sketch Engine interface is considerably more versatile than the Lextutor one. We might add that the corpus we use is more appropriate for the target students, as it is more than six times larger, is more contemporary and is composed of texts that EAP students themselves will write. Finally, our aim is to create a database of re-usable links that can be shared with other teachers and lecturers.

2.2. Steps in creating the DDL materials

- 13 The first step to creating the DDL materials discussed in this paper involves examining written coursework produced by students studying at undergraduate and Master's level in a university in the UK in order to identify recurrent problems that affect intelligibility and/or communicative force. For each such problem, we then create queries on the Sketch Engine open-access interface to exemplify more appropriate usage. The corpus output from these queries is progressively refined until each query leads to a page containing a relatively manageable number⁹ of concordance lines illustrating typical uses of the relevant pattern. The intention in doing so is that the corpus data will draw learners' attention to differences between their own wordings and those of proficient student writers (often in the same or similar disciplines, producing the same or similar academic genres), in line with Johns' original conception of DDL realised in the form of the kibbitzer (Johns, 1986, 1991b). In order to demonstrate this procedure, we will now consider an example of an issue addressed, starting with the instance in a student paper.
- 14 In this example, the student has written the following clause (emphasis added): *grammar teaching could **bring** a variety of negative impacts on English study*. For proficient users of English the use of the verb *bring* here is a very unusual choice when the verb *HAVE* is so strongly associated with *IMPACT on*.¹⁰ Having found an apparent phraseological issue, the

next consideration is whether this is worth bringing to this student's attention, and, for the purposes of the project, whether this is likely to be an issue that other learners may encounter—since an important aspect of the initiative is the re-usability of the links we create. This question can be considered from different perspectives. One factor is the relative complexity of the phrase in question. Detailed discussion of how complexity may be judged is beyond the scope of this paper (see e.g. Sinclair, 2004; Gries, 2008), but clearly this phrase contains a number of elements which contribute to its potential complexity: HAVE + optional determiner + optional adjective + IMPACT + *on*. Another factor worthy of consideration is the frequency of the phrase in academic use. A search for IMPACT (as a noun) followed by *on*¹¹ in the BAWE corpus on Sketch Engine retrieves 714 instances, giving a normalised frequency of around 86 hits per million words.¹² This relatively high normalised frequency provides evidence that it is worth paying some attention to this phrase (see e.g. Vincent, 2013).

- 15 Having decided that a particular issue is worth bringing to the attention of a learner or learners, the next step is to create concordance lines that will make the normal wording of the pattern or phrase salient. Clearly, this involves considering which element(s) of the phrase or pattern in this specific case has caused difficulty, i.e. here, the choice of verb. However, bearing in mind the variability that may be observed in phrases (Francis, 1993; Philip, 2009), decisions here should also be made regarding which of the possible variations will be included with respect to the error or issue that has been identified. Do we view the realisation **bring a variety of negative impacts on ...** in terms of the abstract phrase HAVE + determiner + adjective + IMPACT + *on* or consider specifically instances of *negative IMPACT* or even simply *negative impacts*? Such decisions may depend on a range of factors (considered in Section 4 below), not least initial findings in the corpus.
- 16 Frequent users of the Sketch Engine interface will be aware that there are a number of query options which allow this process to be approached from different angles. The option chosen will depend on decisions made regarding which elements of the phrase in question are considered as particularly salient or important. In this case, as noted already, IMPACT *on* gives a large number of lines; to limit this number, a “simple” search for *negative impact on* was used. To make the pattern of usage more salient, the 45 lines retrieved in this manner were “left sorted” (i.e. so that words to the left of the node are in alphabetical order); a sample of these lines is shown in Figure 3.

Figure 3. – Sample of lines retrieved from BAWE using the simple search *negative impact on*.

reduced purchasing power of the masses has a negative impact on the growth of the domestic market and on the
<p> Surprisingly, the level of urbanisation has a negative impact on child mortality at the ten percent level of
en decreasing on the global market which has a negative impact on balance of payments of developing countries
divisions in the union movement can have a negative impact on trade union membership (Bean 1994:41;
/pe of heat exchanger. The capital costs have a negative impact on the payback time, hence only the lower power
/pe of heat exchanger. The capital costs have a negative impact on the payback time, hence only the lower power
capacity leakage. Again, this could have a negative impact on staff motivation, as it would remove a great
(determinant), it will inevitably have a negative impact on the financial performance (result). Hence,
, generally, in senior market spouses have a negative impact on travel, which suggests that tourism marketing
iol and coming from the UK are found to have a negative impact on qtmark. </p><p> Amongst all the variables, ability w
at different points of time, it is said to have a negative impact on job satisfaction, commitment and motivation
factor being 'unhelpful' for it to have a negative impact on the regression. The coefficient on revision
in imports of Polish meat, as one way to have a negative impact on Polish economy. Russia has in a number of cases
of the state and that this would have a negative impact on local producers who would lose business to

- 17 Having retrieved suitable examples from the corpus, the next stage is to create a permanent link to the concordance lines using the Sketch Engine hyperlinks facility.¹³

This link is then inserted directly into the learner's assignment in an appropriate place, with a short accompanying comment;¹⁴ the font for the verb *bring* is also changed to red to draw attention to its inappropriacy. The idea, then, is for the student to notice the added comment (in bold typeface), click on the hyperlink and then correct the issue marked on the assignment, or at least note the issue for future reference.

- 18 In order to ease the introduction of the links, it seems important that students have some awareness of concordances and, ideally, the nature of the corpus that is used. Accordingly, the students who have received feedback in this way have been shown how the links to Sketch Engine were included in their coursework feedback, and have been introduced to some of the features of Sketch Engine and the BAWE corpus. This was preceded earlier in the course with exposure to other corpus-based tools in a session in which they investigated the potential of *SkELL*, *Just the Word* and *Word and Phrase* to address similar issues.
- 19 Assignments we have returned to these students have generally contained between two and five hyperlinks. In some cases these address all the phraseological problems in the assignment, but in other cases, where assignments contain a great many errors, we decided to select only a small number for the hyperlink treatment, so as not to overwhelm the writers but instead to give them more time for in-depth interaction with the corpus data, following the approach of Gaskell and Cobb (2004).

3. Advantages of the approach

- 20 An advantage of this intervention over fixed, paper-based approaches is that the links lead to interactive Sketch Engine pages. This means that a number of view options are available: users can sort to the left and right, view full sentences and/or filter for uses particularly relevant to their own discipline and chosen genre. To show how this works, we have first provided a sample of lines showing realisations of the pattern *HAVE + effect (on)* (which is similar in meaning to *have an impact on*) in student academic writing in the sciences, allowing for between one and three intervening words.
- 21 The lines presented in Figure 4 can help raise awareness of the uses of the pattern in student writing by showing items that commonly occur between *HAVE* and *effect* and also by indicating the prevalence of *on* after *effect*. Figure 5 is a screenshot of some of the same corpus output as Figure 4 but with a larger context provided for the last concordance line. This context is revealed by clicking on the node item; it is useful when concordance lines which are of interest to the learner do not show full sentences or when a larger context is necessary to decipher the meanings of certain items, e.g. instances of anaphoric reference.

Figure 4. – Sample of lines retrieved from BAWE using the simple search HAVE * effect.

</p><p> In cauliflower the increase of C0 2	has also a beneficial effect	. Higher concentrations of C0 2 lead to an
and another earl (An 1 and D8) occur. These	have the effect	of stamen abortion by blocking normal specific
quad tree encoding, as can readily be seen,	has little effect	. </p><p> In the example of the region above,
, provide further evidence that some genes	have a major effect	on ageing. </p><p> One genetic theory of agei
stresses and to resist fracture. Age also	has an effect	on bone-cross sectional area, where the older
is generated for the independent variable	having a causal effect	on the dependent variable, which can be used a
s the reduction in the model appear to have	had any effect	on the linear dependence of the residuals. Both
. Maternal smoking does not appear to	have any effect	on test of motor impairment score. </p><p> TI
in modelling, despite having been shown to	have some effect	on reading comprehension score. </p><p> Cons
n included in the model when it was seen to	have the largest effect	on ToMI score individually. This may be due to
productivity of the oceans through this may	have a positive effect	on reducing global warming, as it can be shown
change. Variation in the solar output can also	have an effect	on our climate. The sun's output has been linke
changing of the earths orbit is also believed to	have some effect	on global temperatures. The effect, known as
in emphasis during encoding, for example, can	have a profound effect	on how, or indeed whether, information is
, p.05. However, birth order does appear to	have a significant effect	on how overprotective fathers are of their

- 22 Figure 6, meanwhile, shows some of the same corpus output with details of the level and first language of the author (one of the student contributors to the BAWE corpus), and the discipline and genre of the assignment. These details are revealed by clicking on the bar to the side of the concordance lines.
- 23 The information provided in Figures 5 and 6 helps students to contextualise the phraseological information that is available in the concordance lines by providing extra information that may help them make decisions when using the patterns. However, this does not exhaust the interactive possibilities for the students who follow the hyperlinks to Sketch Engine. As noted above, they may choose to sort the lines, either to the left, the right or by the node itself, which in this case helps show the most common patterns of determiner and adjective use with HAVE + effect. Having entered the SketchEngine site, it is also possible that some learners will go on to explore the corpus using other related searches, although, as Gaskell and Cobb (2004) point out, this will likely depend on the training they have received in using the interface.

Figure 5. – A selection of KWIC concordance lines for HAVE + effect with context provided for the last concordance line.

. Maternal smoking does not appear to	have any effect	on test of motor impairment score. </p><p>
from modelling, despite having been shown to	have some effect	on reading comprehension score. </p><p> Co
not been included in the model when it was seen to	have the largest effect	on ToMI score individually. This may be due to
productivity of the oceans through this may	have a positive effect	on reducing global warming, as it can be shov
change. Variation in the solar output can also	have an effect	on our climate. The sun's output has been linl
changing of the earths orbit is also believed to	have some effect	on global temperatures. The effect, known as
in emphasis during encoding, for example, can	have a profound effect	on how, or indeed whether, information is
, p.05. However, birth order does appear to	have a significant effect	on how overprotective fathers are of their

⌵

< previous significantly more emotional warmth towards their youngest child, $F(2,80)=0.74$, $p.05$. Likewise, fathers are not significantly more emotionally warm towards their eldest offspring, $F(2,80)=.19$, $p.05$. However, birth order does appear to **have a significant effect** on how overprotective fathers are of their offspring, $F(2,80)=3.85$, $p<.05$. </p><p> For this particular experimental design, a multivariate analysis of variance (MANOVA) produced results that were identical to those obtained when a one-way [next >](#)

Figure 6. – A selection of KWIC concordance lines for **HAVE + effect** showing details of the author and the text.

Mathematic...	from modelling, despite having been shown to	have some effect	on reading comprehension score. </p><p>																		
Mathematic...	not been included in the model when it was seen to	have the largest effect	on ToMI score individually. This may be due																		
Meteorolog...	productivity of the oceans through this may	have a positive effect	on reducing global warming, as it can be sl																		
Meteorolog...	change. Variation in the solar output can also	have an effect	on our climate. The sun's output has been																		
Meteorolog...	changing of the earths orbit is also believed to	have some effect	on global temperatures. The effect, known																		
Psychology...	in emphasis during encoding, for example, can	have a profound effect	on how, or indeed whether, information is																		
Psychology...	<table><tr><td>Discipline group (area)</td><td>PS</td></tr><tr><td>Discipline</td><td>Mathematics</td></tr><tr><td>text.educ</td><td>UKA</td></tr><tr><td>Text genre</td><td>Exercise</td></tr><tr><td>Grade</td><td>D</td></tr><tr><td>First language of author</td><td>English</td></tr><tr><td>Level of studies</td><td>4</td></tr><tr><td>text.sex</td><td>m</td></tr><tr><td>text.studentage</td><td>25-</td></tr></table>			Discipline group (area)	PS	Discipline	Mathematics	text.educ	UKA	Text genre	Exercise	Grade	D	First language of author	English	Level of studies	4	text.sex	m	text.studentage	25-
Discipline group (area)	PS																				
Discipline	Mathematics																				
text.educ	UKA																				
Text genre	Exercise																				
Grade	D																				
First language of author	English																				
Level of studies	4																				
text.sex	m																				
text.studentage	25-																				

- 24 Most students at British universities study in the medium of English, and do not receive explicit English language or writing instruction. Moreover, their writing is usually assessed primarily for content rather than language accuracy/appropriacy. In this environment lecturers tend to concentrate on imparting disciplinary knowledge rather than providing language support, so they welcome any means of enabling students to deal with language issues in their own time, outside class. However, although our procedure is designed for use by lecturers who cannot spare class time to discuss linguistic choices with their students, we ran an initial trial with international students who were taking a course in academic writing, and were receiving a greater than usual amount of language support. These students were highly motivated learners, as evidenced by their willingness to study as exchange students in the UK, and they were sufficiently interested in corpus-informed methods of language teaching and learning to provide us with thoughtful feedback. Volunteers from this group (n=8) participated in a focus group where the instructor (one of the authors) asked them whether or not they found the hyperlinks useful, and to elaborate on any difficulties they faced.
- 25 The feedback from the students was very positive. The aspects of the intervention that were consistently mentioned as being helpful were the contextualisation of the phrases presented, the ease of contrast with their own writing and the numbers of examples provided. They also claimed that this approach helped their retention of the target usage. Perhaps the strongest argument in favour of the use of corpora in general, and the hyperlink approach in particular, was that rather than relying on memorisation or simply knowing the answer, access to the corpus was seen as providing a tool that the students could refer to in future. This final point indicates the potential of the approach; this group of students at least became interested in exploring Sketch Engine further. It should be borne in mind, however, that the students had already been introduced to *SkELL*, *Just the Word* and *Word and Phrase*, as part of their academic writing course; this sort of pre-exposure might well have made them more inclined to appreciate the benefits of corpus consultation. Clearly, further research is needed to explore in more depth the responses and attitudes of a wider range of potential student users.

4. Discussion: creation and classification issues

- 26 This section discusses some of the issues that have arisen so far, both relating to the creation of the hyperlinked concordances and in their subsequent classification for re-use.

4.1. Style vs Grammar

- 27 Our aim in creating these links is to address phraseological problems in student writing that affect intelligibility and/or communicative force. A further important focus is on problems that are likely to recur in the writing of other students, and ones that can be recognised by students if they refer to the hyperlinked corpus data we supply. The hyperlinks we create to help with such problems are being categorised and stored, so that they can be re-used by ourselves, other EAP practitioners, and subject tutors.
- 28 Not all the problems we encounter can be classed as “errors” in the sense of grammatical mistakes; some are simply unidiomatic or in an inappropriate register, often because of the writers’ collocation choices. In the commentaries that accompany the hyperlinks we therefore have to decide whether to label a usage as “wrong”, and present a correction, or whether to simply offer suggestions which might enhance readability and give the impression of greater academic literacy. For example, one student wrote “In this short essay, there isn’t room to further discuss this important aspect”. In our view this is grammatically correct but rather unwieldy, or unidiomatic in the context, as evidenced by the fact that the form “there isn’t room” is not found in the BAWE corpus. Our reusable hyperlink, <ske.li/ewj>, leads to KWIC concordance lines which suggest a more appropriate choice of wording, shown in Figure 7.

Figure 7. – A selection of KWIC concordance lines for “beyond the scope of”.

motivations for joining particular trade unions are **beyond the scope** of this essay. Nevertheless, it is recognised that ‘ideological
ory of elliptic curves, although this is conveniently **beyond the scope** of this discussion. It is involved in the proof of the Mordell-Weil
ore common with the visual approach as that goes **beyond the scope** of this work, but it does go to reiterate the significance of c
The issue of genocide is a complex one, and it is **beyond the scope** of this paper to tackle it fully, but suffice it to say here that
al perspectives have been formulated which go far **beyond the scope** of this discussion. However, it is possible to pick out recurri
es surrounding child soldiers are complex and it is **beyond the scope** of this report to consider each of the above issues compreh
retations and a complex level of analysis reaching **beyond the scope** of this essay. The ‘transformation problem’ is important in a
nen and the environment, is discussed below. It is **beyond the scope** of this essay to provide a comprehensive outline of all instai

In cases like this, where it is a matter of idiomaticity or appropriacy to context, it might be sufficient simply to highlight the problematic words and supply a link offering an alternative usage.

- 29 On the other hand, students sometimes use prepositions in ways that could be described as ungrammatical or non-standard. The form *comprising of*, for example, was marked as an error in feedback provided to one student; the accompanying hyperlink <ske.li/ep5> leads to grammatically correct concordance lines, a sample of which are shown in Figure 8.

Figure 8. – A selection of KWIC concordance lines for *comprising*.

is a component of reality. Within that part of reality **comprising** the transitive production of knowledge, one can find t
 ovides stability to your shoulder. One of the muscles **comprising** this cuff is relatively avascular and prone to rupture if
 ults. </p> Materials <p> A structured questionnaire **comprising** three sections was used (see Appendix A). These sect
 nal. </p><p> There are small amounts of vegetation **comprising** trees and bushes at the top of faces A and C but the n
 tudy participants were provided with questionnaires **comprising** two sections. The first section contained 35 items sel
 antitatively the most important component of meat, **comprising** up to 75% of its weight. The water is inversely related

- 30 The lines shown in Figure 8 indicate just one way of addressing the non-standard *comprising of*. It would also be possible to take a different approach to this mistake, e.g. by suggesting an alternative verb such as *consist* which is followed by *of*¹⁵ or indicating when the verb *comprise* may be followed by *of*.

4.2. Student proficiency

- 31 Not all the grammatical errors we have found in our data are produced by early-stage EAP learners; for example, English was the first language of the student who wrote *comprising of*. So far, we have found that our approach can be used to provide guidance to a wide range of novice academic writers, regardless of their first language and their English language knowledge. At the same time, this range raises the question of the extent to which it is necessary to adjust the presentation of the feedback and choice of instances to meet the needs of learners at different proficiency levels.
- 32 In terms of presentation of lines, it would seem generally advisable to reduce the size of the concordance sample when creating links for less proficient students, since otherwise there is a danger they will be overwhelmed by the quantity of data. In our experience, more proficient students are more tolerant of and even welcome the provision of greater numbers of lines as they provide greater evidence of usage. A related issue (discussed above in Section 2.2) is that of the degree of variability in the pattern that is presented. It seems reasonable to limit the number of different forms that lower proficiency students are exposed to; a student who uses BRING with *negative impacts on* may struggle if too many variations of the pattern HAVE + IMPACT + *on* are shown in the concordance lines. However, a more proficient student may be able to handle a greater degree of variability in the lines, to the extent that they may even find it useful to see lines including both IMPACT and EFFECT, such as those shown in Figure 9. In order to make informed choices regarding which lines to present, it may also, of course, be necessary to carry out a considerable amount of informal research on the typical realisations of the given pattern.

Figure 9. – Sample of lines showing the pattern HAVE + EFFECT/IMPACT + *on*.

around a third of Europe's population,	had a dramatic impact on	the course of religion in the
. The European Economic and Monetary Union	has a notable effect on	trade, boosting bilateral trade among EMU
However, it does appear that birthweight ratio	has a strong effect on	reading comprehension score, while
the backwards exponential smoothing	has still had little effect on	the series. A few of the more extreme values have
thought that the innovation of the railways did	have a considerable impact on	the growth of the economy in 1965. The problem is
nothing and there are worries that that would	have a detrimental effect on	society. </p><p> Yet means tested benefits come up
another hotel from this site. Customer reviews	have a positive impact on	decision-making. </p><p> Destination Asia- return
score. Both birthweight and gestational age	have a significant effect on	ToMI score at the 5% level. While the result for
detail about demarcation as I feel it does not	have any effect on	my final conclusion. It is concerned with
as the fact that Draba was a black slave, would	have had a great impact on	the film audience. As Draba hurls the trident,

- 33 A further issue relating to proficiency levels is one which we have mentioned already: the number of errors that it is advisable to address in any one piece of work. In student

writing where there are many errors it is not practical or even possible to provide links to exemplify every error (Gaskell & Cobb, 2004). On the other hand, limiting the number of hyperlinks to around five may give the student the impression that there are no further issues to address. Our approach, like other DDL interventions, does not make any claims to comprehensiveness and does not set itself up as in opposition to other means of providing feedback on errors. It is meant to be understood and used as a complementary approach for specific errors and infelicities that are amenable to a phraseological treatment.

4.3. Generalisability and abstraction

- 34 Another consideration when composing corpus queries relates to the generalisability of the outputs which are chosen. Although we intend our links to be reusable by other practitioners, we also want them to address the writing problems experienced by individual students. For this reason, it is necessary to find a balance between feedback that is over-generalised and that which may only apply in one specific context. This is a question not only of usefulness to students and practitioners but also of a sensible ratio of time spent creating concordances to the number of uses they may have. This can again be exemplified by the example of HAVE + IMPACT/EFFECT + *on*. The lines retrieved to help the student who used *brings* instead of *has* (see Fig. 3) are quite specific in that they all include *negative impact on*, but we know that more generally other adjectives and even other nouns can enter into this pattern. The question here is whether it is better to create a set of lines such as those shown in Figure 9 or whether to stick to the more specific and targeted link shown in Figure 3. The answer to this question may depend on a number of factors, such as student level, student familiarity with concordance lines and the delicate balance between making a correction obvious and encouraging the student to look more deeply into a range of possible corrections, requiring more cognitive processing, an important aspect of DDL (Boulton & Cobb, 2017). There is no definitive answer to this question, but it is one which we will need to consider when creating further links.
- 35 Careful use of CQL should result in concordance lines which show appropriate usage but (notwithstanding points raised in Section 4.2) allow for intervening words, and for variations of tense, aspect, voice and number inasmuch as these are found in the pattern concerned. The lines can then aim to show the function of a pattern in a range of different sentence types. A related issue which has not received much attention in this paper, but which is of importance to learners, is that of genre and discipline specificity. Since the Sketch Engine interface for the BAWE corpus provides the facility to select specific genres and/or disciplines, there is an argument for limiting searches in this way when frequently occurring patterns are exemplified.
- 36 A final aspect relating to generalisability and abstraction concerns the outputs that are provided in links. So far in this paper we have only discussed the use of concordance lines but these are of course not the only corpus outputs that might be useful to show patterns of usage; other possibilities include collocation lists, word sketches or frequency lists. One useful feature available on Sketch Engine (and indeed many corpus interfaces) is “Frequency: Node forms”, which lists the types retrieved by a search in frequency order.

Figure 10. – Screenshot of Frequency: Node forms for *ability to* followed by infinitive verb.

word		Frequency	Items: 469 Total frequency: 1,010	
P N	ability to produce	31		
P N	ability to use	16		
P N	ability to make	15		
P N	ability to create	14		
P N	ability to control	14		
P N	ability to understand	13		
P N	ability to communicate	13		
P N	ability to influence	12		
P N	ability to learn	11		
P N	ability to express	10		
P N	ability to work	9		
P N	ability to identify	9		

- 37 An example of the sort of output the node form frequency facility provides is shown in Figure 10; the full list is available through this link: <skel.li/e2s>. The advantage of providing this output over a sample concordance of *ability to* is that it is immediately obvious which verbs most commonly occur with *ability to*; one can even start to classify types of verb (e.g. verbs relating to *producing/making*, *understanding* and *communicating*). Moreover, one can easily access the original concordance as well as concordances for individual verbs from this screen if desired. At present, outputs of this sort (i.e. those which are not concordance lines) are somewhat under-utilised in our project, which remains a limitation of the approach, but we are interested in exploiting them more in future (see next section).

4.4. Towards a classification of hyperlinks created

- 38 As we have pointed out, the vision for this intervention is not just to help students address phraseological issues but also to create a searchable bank of hyperlinks for re-use by us, our colleagues and other interested parties. This is in the spirit of Johns' (1991b, p. 36) proposal to create "ready-made DDL materials that [teachers] could adapt with their own students in mind"¹⁶ but it should also be noted that these links are not solely for teachers of English, but are also intended to be used by subject lecturers. The foregoing discussion has raised issues relating not only to the creation of hyperlinked concordances but also to their classification for ease of retrieval and re-use. This is an ongoing task, but our category names will need to enable us to identify the style or grammar issue, and possibly also the student proficiency level, the degree of abstraction or generalisability, and even perhaps the discipline and/or genre. Our progress with this classification process is documented on the website of the project, where issues and their related hyperlinks are stored in entries of the directory, which practitioners are welcome to consult.¹⁷

5. Conclusion

- 39 We believe that, as well as being within the spirit of DDL, the approach introduced in this paper has a number of advantages. The first of these is the use of the BAWE corpus, which contains the type of writing that the students concerned aspire to produce, rather than web-scraped mega-corpora of "general English" (*SkELL*) or academic corpora composed largely of research articles (*Word and Phrase* – academic component). A second advantage

is ease of access—the multitude of choices and the daunting nature of corpus interfaces is largely overcome because we conduct the searches in advance and provide links to pre-filtered concordance lines. The fact that the links are live and allow for further exploration does not, of course, preclude further “serendipitous learning” (Bernardini, 2004) and may in fact pique the curiosity of some students to find out more about how to use the corpora. Finally the fact that the links are provided in individual feedback allows students to access and learn from them at their own pace.

BIBLIOGRAPHY

Websites referred to

BAWE: <www.coventry.ac.uk/bawe>.

Just the Word: <www.just-the-word.com>.

SkELL: <<https://skell.sketchengine.co.uk/run.cgi/skell>>.

Sketch Engine Open Corpora: <<https://the.sketchengine.co.uk/open/>>.

Tim Johns' kibbitzers (Wang Lixun): <<http://ec-concord.ied.edu.hk/TimJohns/>>.

Tim Johns' kibbitzers (WordSmith Tools): <<http://lexically.net/TimJohns>>.

Word and Phrase: <www.wordandphrase.info>.

Writing for a Purpose: <<https://learnenglish.britishcouncil.org/en/writing-purpose/>>.

References

ASTON, Guy. (1997). Small and Large Corpora in Language Learning. In B. Lewandowska-Tomaszczyk & P. J. Melia (eds), *PALC '97: Practical Applications in Language Corpora* (pp. 51–62). Łódź: Łódź University Press.

BERNARDINI, Silvia. (2000). Systematising Serendipity: Proposals for Concordancing Large Corpora with Language Learners. In L. Burnard & T. McEnery (eds), *Rethinking Language Pedagogy from a Corpus Perspective* (pp. 225–34). Frankfurt am Main: Peter Lang.

BERNARDINI, Silvia. (2004). Corpora in the Classroom: An Overview and some Reflections on Future Developments. In J. Sinclair (ed.), *How to Use Corpora in Language Teaching* (pp. 15–38). Amsterdam/Philadelphia: John Benjamins.

BOULTON, Alex. (2010). Data-Driven Learning. On Paper, in Practice. In T. Harris & M. Moreno Jaén (eds), *Corpus Linguistics in Language Teaching* (pp. 17–52). Bern: Peter Lang.

BOULTON, Alex & COBB, Tom. (2017). Corpus Use in Language Learning: A Meta-Analysis. *Language Learning*, 67(2), 348–93.

CHAMBERS, Angela. (2010). What Is Data-Driven Learning? In A. O'Keeffe & M. McCarthy (eds), *The Routledge Handbook of Corpus Linguistics* (pp. 345–58). London: Routledge.

CHARLES, Maggie. (2007). Reconciling Top-Down and Bottom-Up Approaches to Graduate Writing: Using a Corpus to Teach Rhetorical Functions. *Journal of English for Academic Purposes*, 6(4), 289–302.

- CHARLES, Maggie. (2011). Corpus Evidence for Teaching Adverbial Connectors of Contrast: *however, yet, rather, instead and in contrast*. In N. Kübler (ed.), *Corpora, Language, Teaching, and Resources: From Theory to Practice* (pp. 113–31). Bern: Peter Lang.
- CHARLES, Maggie. (2015). Same Task, Different Corpus: The Role of Personal Corpora in EAP Classes. In A. Leńko-Szymańska & A. Boulton (eds), *Multiple Affordances of Language Corpora for Data-Driven Learning* (pp. 131–54). Amsterdam: John Benjamins.
- DURRANT, Philip. (2013). Discipline and Level Specificity in University Students' Written Vocabulary. *Applied Linguistics*, 35(3), 328–56.
- FERRIS, Dana. (2002). *Treatment of Error in Second Language Student Writing*. Ann Arbor, MI: The University of Michigan Press.
- FRANCIS, Gill. (1993). A Corpus-Driven Approach to Grammar: Principles, Methods and Examples. In M. Baker, G. Francis & E. Tognini-Bonelli (eds), *Text and Technology: In Honour of John Sinclair* (pp. 137–56). Amsterdam: John Benjamins.
- GABRIELATOS, Costas. (2005). Corpora and Language Teaching: Just a Fling or Wedding Bells? *Teaching English as a Second Language – Electronic Journal*, 8(4), 1–35. Available at <<https://files.eric.ed.gov/fulltext/EJ1068106.pdf>> (accessed 18 December 2017).
- GASKELL, Delian & COBB, Thomas. (2004). Can Learners Use Concordance Feedback for Writing Errors? *System*, 32(3), 301–19.
- GRIES, Stefan. (2008). Phraseology and Linguistic Theory: A Brief Survey. In S. Granger & F. Meunier (eds), *Phraseology: An Interdisciplinary Perspective* (pp. 3–26). Amsterdam/Philadelphia: John Benjamins.
- HIGGINS, John. (1991). Looking for Patterns. In T. Johns & P. King (eds), *English Language Research Journal*, 4, 63–70.
- HYLAND, Ken. (2003). *Second Language Writing*. Cambridge: Cambridge University Press.
- JOHNS, Tim. (1986). Micro-Concord: A Language-Learner's Research Tool. *System*, 14(2), 151–62.
- JOHNS, Tim. (1991a). Should You Be Persuaded: Two Examples of Data-Driven Learning. In T. Johns & P. King (eds), *English Language Research Journal*, 4, 1–16.
- JOHNS, Tim. (1991b). From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data-Driven Learning. In T. Johns & P. King (eds), *English Language Research Journal*, 4, 27–45.
- KRISHNAMURTHY, Ramesh & KOSEM, Iztok. (2007). Issues in Creating a Corpus for EAP Pedagogy and Research. *Journal of English for Academic Purposes*, 6(4), 356–73.
- LEE, David & SWALES, John. (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.
- MCENERY, Anthony & WILSON, Andrew. (1997). Teaching and Language Corpora (TALC). *ReCALL*, 9(1), 5–14.
- NESI, Hilary & GARDNER, Sheena. (2012). *Genres across the Disciplines*. Cambridge: Cambridge University Press.
- NESI, Hilary, GARDNER, Sheena & KIGHTLEY, Adam. (2015). Writing for a Purpose. In T. Pattison (ed.), *IATEFL 2014: Harrogate Conference Selections* (pp. 145–6). Faversham, Kent: IATEFL.

PÉREZ-PAREDES, Pascual, SÁNCHEZ-TORNEL, Maria & ALCARAZ CALERO, Jose. (2013). Learners' Search Patterns During Corpus-Based Focus-on-Form Activities. *International Journal of Corpus Linguistics*, 17(4), 482–515.

PHILIP, Gill. (2009). Reassessing the Canon: “Fixed” Phrases in General Reference Corpora. In S. Granger & F. Meunier (eds), *Phraseology: An Interdisciplinary Perspective* (pp. 95–108). Amsterdam/Philadelphia: John Benjamins.

QUINN, Cynthia. (2015). Training L2 Writers to Reference Corpora as a Self-Correction Tool. *ELT Journal*, 69(2), 165–77.

RUTHERFORD, William. (1987). *Second Language Grammar: Learning and Teaching*. London: Longman.

SCHMIDT, Richard. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics*, 11(2), 129–58.

SINCLAIR, John. (2004). *Trust the Text*. London: Routledge.

TRIBBLE, Chris & JONES, Glyn. (1990). *Concordances in the Classroom: A Resource Guide for Teachers*. London: Longman.

VINCENT, Benet. (2013). Investigating Academic Phraseology through Combinations of Very Frequent Words: A Methodological Exploration. *Journal of English for Academic Purposes*, 12(1), 44–56.

YOON, Choongil. (2011). Concordancing in L2 Writing Class: An Overview of Research and Issues. *Journal of English for Academic Purposes*, 10(3), 130–9.

NOTES

1. Examples of kibbitzers can be found at <<http://lexically.net/TimJohns>> and <<http://ec-concord.ied.edu.hk/TimJohns/>>.
2. See SkELL: <<https://skell.sketchengine.co.uk/run.cgi/skell>>.
3. See Just The Word: <www.just-the-word.com>.
4. See Word and Phrase: Frequency Lists <www.wordandphrase.info/frequencyList.asp>.
5. See also British Academic Written English Corpus (BAWE): <<http://www.coventry.ac.uk/bawe>>.
6. Available at Sketch Engine Open Corpora: <<https://the.sketchengine.co.uk/open/>>.
7. In the description of phrases and patterns we will use *italics* to denote wordforms and SMALL CAPS to denote lemmas.
8. The concordancer is now located here <<https://lxtutor.ca/conc/eng/>> and currently gives users the option of searching a wider range of corpora than was available to Gaskell and Cobb in 2004.
9. The number of lines that may be considered “manageable” is a subjective matter and teachers may want to take into account factors such as student level and previous exposure to concordances. In many cases, samples of around 20–30 lines have been created. Since the hyperlink facility is available to all, any user of the link would be able to filter this down further and create a new link on Sketch Engine if this was thought to be too many.
10. A collocate analysis of lemmas associated with IMPACT on in the BAWE corpus (-5L to -1L, minimum co-occurrence 3, ranking by logDice values) indicates that HAVE co-occurs in more than half of all instances; no other verb (except *will*) is found in the top 50 collocates. As pointed out by one of our reviewers, this is not the only potentially unidiomatic feature of this instance, but, as probably the most salient one to a proficient writer of English, it is the one we are focusing on here.

11. CQL query: [lemma="impact" & tag="N.*"] "on".
 12. This does not account for the separation of **IMPACT** and *on* in instances such as *the **impact** it had **on** her writing* (from BAWE document 0063d; emphasis added).
 13. This facility is indicated by a “link” icon in Sketch Engine.
 14. The comment in this case was “*bring*” isn’t the usual verb. See [here](#) for examples—the hyperlink was added to the underlined word.
 15. Our thanks to a reviewer who pointed this out.
 16. We thank one of our reviewers for pointing this out.
 17. The quick links of the BAWE web address is <bawequicklinks.coventry.domains>.
-

ABSTRACTS

This paper outlines a new initiative aimed at integrating concordances and other corpus outputs into written feedback for learners of English for Academic Purposes (EAP). Although data-driven learning has by now a 30-year history, it has yet to have a great impact on mainstream pedagogy despite various claims regarding its efficacy and its benefits in terms of promoting learner autonomy. This situation may be due to technical barriers to the use of corpora in teaching, in particular the apparent complexity of most corpus interfaces to the uninitiated. We seek to circumvent these barriers by creating concordances to help students address their written errors. These concordances, found in the free-to-use British Academic Written English corpus, are made available as hyperlinks pasted into student work to help them address their errors by giving them direct access to instances of proficient academic writing. This paper will outline the methods used to create the concordances, the types of writing issues that are most amenable to this treatment and some initial student feedback on the first phase of an ongoing project to identify and categorise typical phraseological problems in student writing.

Cet article a pour but de présenter une nouvelle initiative visant à intégrer des concordances et d’autres éléments de corpus dans les réponses fournies aux apprenant.e.s afin qu’ils/elles puissent remédier aux erreurs de rédaction d’un écrit en anglais sur objectif spécifique. Bien que l’approche *data-driven learning* (DDL) existe depuis maintenant 30 ans, elle n’a pas encore eu de réel impact sur la pédagogie courante, malgré les bénéfices déclarés quant à l’autonomie de l’apprenant.e. Cette situation peut être due à des obstacles techniques relatifs à l’utilisation de corpus dans l’enseignement. Les non-initié.e.s peuvent être particulièrement rebuté.e.s par l’apparente complexité de la plupart des interfaces de corpus. Nous cherchons à remédier à ces difficultés en créant des concordances permettant aux étudiant.e.s de corriger leurs erreurs. Ces concordances, extraites du British Academic Written English Corpus et accessibles via des liens hypertexte collés dans le texte de l’étudiant.e, l’aident à rectifier ses erreurs en lui donnant un accès direct à des exemples d’anglais universitaire de qualité. Cet article décrit les méthodes utilisées pour créer ces concordances et les types de problèmes de rédaction que l’on peut traiter de cette façon. Il inclut aussi les réactions initiales des étudiant.e.s dans le cadre de la première phase de ce projet actuellement en cours qui vise à identifier et à catégoriser les problèmes phraséologiques typiques rencontrés par les étudiant.e.s dans leurs travaux écrits.

INDEX

Mots-clés: data-driven learning (DDL), linguistique du corpus, rétroaction écrite, phraséologie, hyperliens

Keywords: data-driven learning (DDL), corpus linguistics, written feedback, phraseology, hyperlinks

AUTHORS

BENET VINCENT

Coventry University

HILARY NESI

Coventry University