COMMUNITY DIRECTIONS

Utilising Data-driven Learning in Chemistry Teaching: a Shortcut to Improving Chemical Language Comprehension

Simon Rees,[1] Megan Bruce[1] & Steven Bradley[2]

[1]The Foundation Centre, Durham University, UK
[2]Department of Computer Sciences, Durham University, UK

Abstract

This article describes the development of the FOCUS project and specific pedagogical strategies to improve understanding of the language of chemistry. The importance of language comprehension skills for success in learning chemistry has recently been highlighted by Pyburn *et al*. (2013). The FOCUS project has involved the construction of a database of student writings (from foundation to Ph.D. level) to create a corpus that can then be analysed for the occurrence of key words in context. Using the principles of concordance and data-driven learning (where student becomes language researcher) a number of teaching activities have been developed to enhance the understanding of subject-specific language for both home and international students studying on a foundation level Chemistry course.

Keywords: data driven learning, chemical language comprehension, corpora

Introduction

Mathematical ability is often considered to be an important indicator for student success in chemistry (Rixse & Pickering 1985) but more recently Pyburn *et al*. (2013) have highlighted the significant contribution of language comprehension skills to student achievement. They demonstrated that comprehension skill correlated with general chemistry performance and also partially

Corresponding author:
Simon Rees, The Foundation Centre, Pelaw House, Durham University, Durham DH1 1TA, UK
Email: simon.rees@durham.ac.uk, Phone: 0191 3348282

compensates for deficits in prior knowledge. They state that “efforts to prepare students for success in general chemistry should include both content and the development of language comprehension skill”.

Specific studies documenting the language issues in chemistry are relatively limited. Initial work by Cassels & Johnstone (1985), however, highlighted confusion caused for school-age pupils by non-technical words in a scientific context and Jasien (2010, 2011) and Jasien & Oberem (2008) similarly described confusion for undergraduates with terms such as dense, energy, neutral and strong. These issues have also been demonstrated to apply for students for whom English is their second language (Johnstone & Selepeng 2001, Childs & O’Farrell 2003). Cassels & Johnstone (1985) suggested that language is a contributor to information overload and Gabel (1999) states that the difficulties students have with chemistry might be more to do with how chemistry knowledge is linguistically expressed than with the subject matter itself. Herron (1996) provides an excellent and comprehensive discussion of the role of language in teaching chemistry in which he summarises many of the key aspects of chemical language and offers useful practical advice for any educator in the subject to reflect upon.

This body of research reinforces our own personal teaching experience where students are often unable to frame their question about the content of a module because they lack the vocabulary to do so. As a consequence, we have undertaken work to develop pedagogical strategies that will assist students to develop their understanding of the language of chemistry. Initial work to assist students with vocabulary issues involved developing a suite of activities including a student generated online glossary (Rees *et al.* 2013). In this article we describe the development of the FOCUS project (www.dur.ac.uk/foundation.focus) and the use of data-driven learning (DDL) as a tool to enhance language understanding and demonstrate how it can be used to successfully support and enhance student learning in chemistry.

The learning context

Durham University Foundation Centre has an annual intake of approximately 200 non-traditional students. The department aims to widen participation in and access to higher education and therefore accepts students who lack some of the formal qualifications. The majority of students are either mature learners returning to education or international students who are unable to study to a sufficient level in their own country for direct entry to UK degree programmes. The Foundation Centre offers progression to all departments in Durham University and a significant number of students study chemistry in order to progress to a number of different subjects such as Biomedical Science, Chemistry, Earth Sciences and Medicine.

Constructing FOCUS

To assist student understanding of the language of chemistry we decided to create a corpus of student texts which our students could then search using a concordancer. The project is entitled “FOCUS” (**Fo**undation **C**orp**us**) and is a collection of academic writings produced by Durham University students (undergraduate and postgraduate) in Chemistry and an increasing number of other subjects. Tribble (1997) cautions against using apprentice performances as corpus data. However, we would argue that since the function of the corpus is to teach Foundation students to write like conventional university students, in this instance student writings are rather more expert than apprentice performances.

Criteria for text inclusion

Acquiring texts to include in a project of this nature is always difficult as it requires the co-operation of a range of different people (Alsop & Nesi 2009, pp76–81). Some of our texts are introductions in Ph.D. and Masters theses which are freely accessible within the university. However, the majority of our texts have been sent to us by students for inclusion in the corpus. In collaboration with the Chemistry Department we obtained a list of students who had scored 60% or more in that particular assignment and contacted them to ask them to send us a copy of their assignment. The contact email outlined the aims and ethical procedures of the project. To incentivise students to send us their writing, we entered all names of contributors into a draw for a £100 Amazon voucher.

Some departments expressed concern that our corpus could turn into an essay bank, which students could plagiarise. In fact, this is not possible as a search for a particular word reveals the keyword plus around 40 characters of text to the left and to the right of the keyword totalling 100 characters per line. Clicking on a single instance of the keyword gives a slightly larger text fragment, but still only around 200 characters. No coding is made available to the user which would allow them to identify and piece together fragments to form a complete assignment.

What is a Concordancer?

The corpus is the collection of texts whilst the concordancer is the actual program that enables analysis of the texts. Johns (1991) defines a concordancer as:

> ***"able to recover from the text all the contexts for a particular item (morpheme, word or phrase) and to print them out in a way which facilitates rapid scanning and comparison. The most usual format is the keyword-in-context (KWIC) concordance in which the keywords are arranged one below the other down the centre of the page, with a fixed number of characters of context to the left and to the right. A useful refinement, particularly where one is concerned with regularities and patterns in large numbers of citations, is the ability to sort alphabetically the contexts to the left or right of the keyword so that similar contexts are grouped together"***
>
> ***(Johns 1991, p2)***

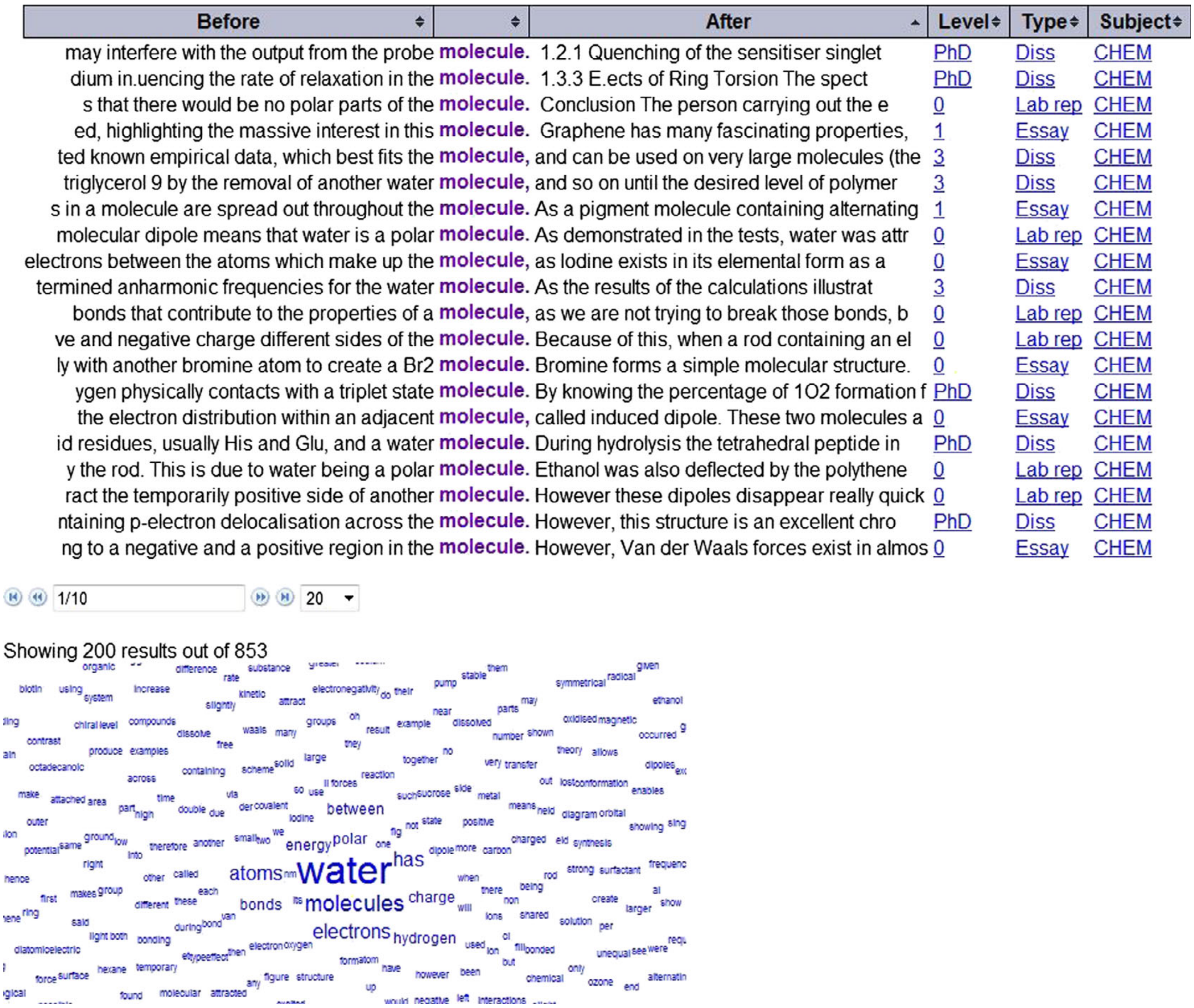
Alternatively, a more concise definition from Tribble (1997, p11) states:

> ***"What the concordance does is make the invisible visible."***

Functionality of FOCUS

The concordancer can be accessed via http://www.community.dur.ac.uk/foundation.focus/. It does require a Durham University login but a video walkthrough is available (guest logins can be provided for anyone who would like to try the tool). A user can perform a simple search by just entering a keyword. This can be a word, morpheme (using the wildcard symbol %) or phrase. A more advanced search can also be performed which limits the findings by level (e.g. UG/PG), type (e.g. essay/lab report) or subject (e.g. Chemistry/Earth Sciences).

The screen can be set up to show 20, 40 or 100 concordance lines and it chooses a random 200 lines from the total held in the corpus. Users



Before		After	Level	Type	Subject
may interfere with the output from the probe	molecule.	1.2.1 Quenching of the sensitiser singlet	PhD	Diss	CHEM
dium in.uencing the rate of relaxation in the	molecule.	1.3.3 E.ects of Ring Torsion The spect	PhD	Diss	CHEM
s that there would be no polar parts of the	molecule.	Conclusion The person carrying out the e	0	Lab rep	CHEM
ed, highlighting the massive interest in this	molecule.	Graphene has many fascinating properties,	1	Essay	CHEM
ted known empirical data, which best fits the	molecule,	and can be used on very large molecules (the	3	Diss	CHEM
triglycerol 9 by the removal of another water	molecule,	and so on until the desired level of polymer	3	Diss	CHEM
s in a molecule are spread out throughout the	molecule.	As a pigment molecule containing alternating	1	Essay	CHEM
molecular dipole means that water is a polar	molecule.	As demonstrated in the tests, water was attr	0	Lab rep	CHEM
electrons between the atoms which make up the	molecule,	as Iodine exists in its elemental form as a	0	Essay	CHEM
termined anharmonic frequencies for the water	molecule.	As the results of the calculations illustrat	3	Diss	CHEM
bonds that contribute to the properties of a	molecule,	as we are not trying to break those bonds, b	0	Lab rep	CHEM
ve and negative charge different sides of the	molecule.	Because of this, when a rod containing an el	0	Lab rep	CHEM
ly with another bromine atom to create a Br2	molecule.	Bromine forms a simple molecular structure.	0	Essay	CHEM
ygen physically contacts with a triplet state	molecule.	By knowing the percentage of 1O2 formation f	PhD	Diss	CHEM
the electron distribution within an adjacent	molecule,	called induced dipole. These two molecules a	0	Essay	CHEM
id residues, usually His and Glu, and a water	molecule.	During hydrolysis the tetrahedral peptide in	PhD	Diss	CHEM
y the rod. This is due to water being a polar	molecule.	Ethanol was also deflected by the polythene	0	Lab rep	CHEM
ract the temporarily positive side of another	molecule.	However these dipoles disappear really quick	0	Lab rep	CHEM
ntaining p-electron delocalisation across the	molecule.	However, this structure is an excellent chro	PhD	Diss	CHEM
ng to a negative and a positive region in the	molecule.	However, Van der Waals forces exist in almos	0	Essay	CHEM

Figure 1 Screenshot of FOCUS showing search returns for the word "molecule".

can sort the data alphabetically using the "Before" or "After" tabs to identify common collocations. The tool also includes a word cloud feature to identify common collocations and guide users into useful explorations about their chosen keyword (see Figure 1).

Another feature of the tool is its wild card function using the % symbol. This allows users to search for all forms of a word family, so searching for combin% would give: combination, combinations, combinatorial, combine, combines, combined, and combining. The wild card symbol also allows users to explore particular affixes, such as %icity. Clicking on the central column would alphabetise the keywords with the –icity suffix and allow users to see which words occur most frequently. Students can also use this type of search to deduce the meaning of various affixes, which in turn improves their ability to deduce the meaning of unknown words in the future.

Teaching with FOCUS

Data-Driven Learning (DDL)

Johns (1991) coined the term "data driven learning" to describe a learning situation in which "the language learner is also, essentially, a research worker whose learning needs to be driven by access to linguistic data – hence the term "data-driven learning" (DDL) to describe the approach" (Johns 1991, p2). In DDL the learner uses data to uncover the rules behind the language while the teacher "provides a context in which the learner can develop strategies for discovery" (ibid).

DDL using corpora and concordancing programs has been commonly used in language classrooms for the past 20 years. Its use is wide-spread but it has been focused on the learning of second languages. Some of our Foundation students are learning in a second language but the majority are native speakers who are trying to learn a subject-specific vocabulary in their own language. Hence, from our perspective, chemistry is their second language. The use of concordancing with native speakers in a subject-specific context represents a novel application of this pedagogical approach.

Sample activities

This section briefly summarises some teaching activities that have been developed utilising FOCUS.

What does "pressure" mean?

In response to this question, a class may be observed reaching for their smartphones and other devices, typing the word into a search engine and producing a response along the lines of something to do with an amount of force over a given area. This is a definition of the word in isolation but only provides a very limited understanding of the meaning and usage of the word in a chemistry context.

The students are then asked to enter the word into FOCUS and search chemistry texts. By clicking on the "Before" tab, the search returns 343 results which are then sorted alphabetically by the word immediately preceding "pressure" (see Figure 2).

The trained user of this tool quickly recognises common collocations with the word "pressure". In this particular sample we can see "exerts a

Before		After	Level	Type	Subject
e conditions, of approximately 1000-3000 atm.	**pressure**	and temperatures of around 50-300C [9] These	1	Essay	CHEM
even reaching 5MPa (about 50 atmospheres)[9].	**Pressure**	is kept constant by feeding monomer into the	1	Essay	CHEM
illary of length L to be (with an atmospheric	**pressure**	difference at each end):	PhD	Diss	CHEM
is really quite necessary, as at atmospheric	**pressure,**	simply 1 gram of hydrogen will volumetricall	1	Essay	CHEM
ymerization may be carried out at atmospheric	**pressure**	under an inert atmosphere of dry nitrogen or	PhD	Diss	CHEM
t evaporates off at 65 C under atmospheric	**pressure.**	11 Before the transesterification reaction ca	3	Diss	CHEM
y temperatures when subjected to super atomic	**pressure**	or in the presence of a catalyst or solvent s	1	Essay	CHEM
, ranging from 65-100C and between 10-37 bar	**pressure.**	The ethylene monomer is fed into the reactor	1	Essay	CHEM
l effects are: increased heart rate and blood	**pressure,**	a sense of alertness, decreased perception o	1	Essay	CHEM
adrenaline which increases heart rate, blood	**pressure**	and respiration and leads to higher blood-glu	1	Essay	CHEM
Pressure methods, such as the Maximum Bubble	**Pressure**	technique, measure the pressure difference a	PhD	Diss	CHEM
il wet capillary. The difference in capillary	**pressure**	across the meniscus drives the imbibition alo	PhD	Diss	CHEM
and therefore it can influence the capillary	**pressure**	on both sides of the meniscus. 31 Hammond and	PhD	Diss	CHEM
ons between the particles and the walls cause	**pressure.**	The smaller the volume of a container, the m	0	Essay	CHEM
pend on the product of the stagnation chamber	**pressure**	and the diameter of the nozzle aperture, Pd (	PhD	Diss	CHEM
etransition,inducedby a changeintemperature,	**pressure**	or concentration, the structure of the state	PhD	Diss	CHEM
alls of the container increases; consequently	**pressure**	increases (Taylor, 2000). Water on the other	0	Essay	CHEM
ecific heat capacities of the gas at constant	**pressure,**	CP , and volume, CV, respectively. As an exa	PhD	Diss	CHEM
used in MD simulation for generating constant	**pressure**	and temperature ensembles. There are also o	PhD	Diss	CHEM
which can be attributed to the higher contact	**pressure**	(710 MPa) for the same applied load. From the	PhD	Diss	CHEM

Figure 2 Sample of search returns for the term "pressure" sorted alphabetically by the preceding word.

pressure" and "temperature and pressure" are common collocations and several more, such as "high" and "low", can be identified further down the results. This evidence is supported by the word cloud where gas, temperature, surface and container are the largest and hence the most frequent words.

The students may now be asked to construct a concept map style diagram to show the connections between the words and this can then lead to a discussion as to why they are often associated with the key term and what the meaning of those words is, e.g. to exert. The search also identifies a number of different types of pressure, for instance partial, transmembrane, radiation and osmotic. The meaning of these specific types of pressure can then be explored. Furthermore, a search within a different subject area can reveal results with similarities and differences, e.g. a search in Earth Sciences texts reveals the common occurrence of hydrostatic, lithostatic and pore pressure. This can then promote a valuable discussion amongst a mixed discipline group of students with regard to the subject-specific usage of a word.

Consequently, in a short amount of time, the learner has undertaken some language research that has quickly exposed them to a rich and diverse sample of sentences. The usage of the term is demonstrated in a wide variety of different contexts so improving their understanding of their subject-specific language and deepening the student's understanding of what the word "pressure" and associated terms mean. The value of this teaching activity is enhanced by the use of authentic chemistry texts and the fact that it promotes learner discovery of the connections between words and their meaning.

Reinforcement of new and problematic terms

Students can also use FOCUS to carry out activities to reinforce their understanding of new and problematic terms as they arise on the course. The students are asked to find out the definition of a word and then to select the correct definition for the meaning of word in context using an example from FOCUS. This is particularly useful when a word has multiple meanings in different context. Table 1, for example, shows the dictionary entry that a student obtained for the word "contract" and also an example sentence from FOCUS. The task is to select the correct definition for the context. Clearly, the first thing to realise is that 'contract' is a verb and not a noun in this instance (note the potentially confusing reference to "bridge" in the noun

Table 1 Results of a student search for the definition of the word "contract" and a sample sentence from FOCUS.

Word	Definition	Example sentence
Contract	n.	The graphene capillaries will expand or contract depending on localised humidity to allow water and nothing else through
	1. a binding agreement between two or more persons that is enforceable by law	
	2. (contract bridge) the highest bid becomes the contract setting the number of tricks that the bidder must make	
	3. a variety of bridge in which the bidder receives points toward game only for the number of tricks he bid	
	v.	
	1. enter into a contractual arrangement	
	2. engage by written agreement	
	3. squeeze or press together	
	4. become smaller or draw together	
	5. be stricken by an illness, fall victim to an illness	
	6. make smaller	
	7. compress or concentrate	
	8. make or become more narrow or restricted	
	9. reduce in scope while retaining essential features	

definitions). There are now a further nine definitions, some of which are more distinct than others but equally contain further words which the student needs to understand. Some students may opt for “make smaller” and, whilst this may be considered correct, there is a better answer based on the context of the example sentence. The sentence refers to “capillaries” and “water flow” which could lead one to the conclusion that the better answer is “make more narrow”. Hence, there is a subtlety to the language understanding that is determined by the context and revealed in the example sentence from FOCUS (also of interest is the common collocation of “expand and contract”).

This activity has elicited some very interesting responses from students. For example, when asked to find out the definition of “complex ion”, one student arrived with the definition “the colour and appearance of someone’s face”. It became apparent that the search tool on his smartphone had combined the two words to provide a definition for “complexion”, something which the student had failed to appreciate. On another occasion, the students were asked to define the prefix “inter” as in “intermolecular” and one student returned with the definition “to place in a grave!” These two extreme examples serve to illustrate the language comprehension challenges that some students face and the importance of developing strategies to improve this situation.

Self-study activities

Class time constraints in respect of carrying out these sorts of activities and the desire to target language support for individual students have led to the development of a range of concordance-based self-study activities. Simply to present the FOCUS tool to students to use is not successful since guidance is required to understand the best way to use it and interpret the results. Hence, the self-study activities are designed to guide the student through the task and to address issues that students have identified as of concern. These activities are summarised in table 2. The design cycle of activity identification, development and review with the Foundation students was effective in developing meaningful and engaging activities over the course of the year.

Table 2 Outline of activities designed to accompany the corpus.

Activity	Summary
Writing laboratory reports	Students undertake a series of activities utilising FOCUS to gain a better understanding of writing methods, conclusions and the meaning of the term “control variable”
Affixes in science	Students follow a Camtasia-produced video activity to improve their understanding of common affixes in science
Words with multiple meanings	Students improve their understanding of words such as “weak” and “strong” in different contexts
Developing reading	Students develop their reading and summarising skills and identify key terms to explore the meaning
Types of scientific language	Students undertake an assessment of their understanding of different types of scientific language and demonstrate how to produce a personal glossary
Using connectives	Students explore the appropriate use of connectives in academic writing

Central to the success of the project has been the development of activities that are interactive and a variety of software applications were utilised to do

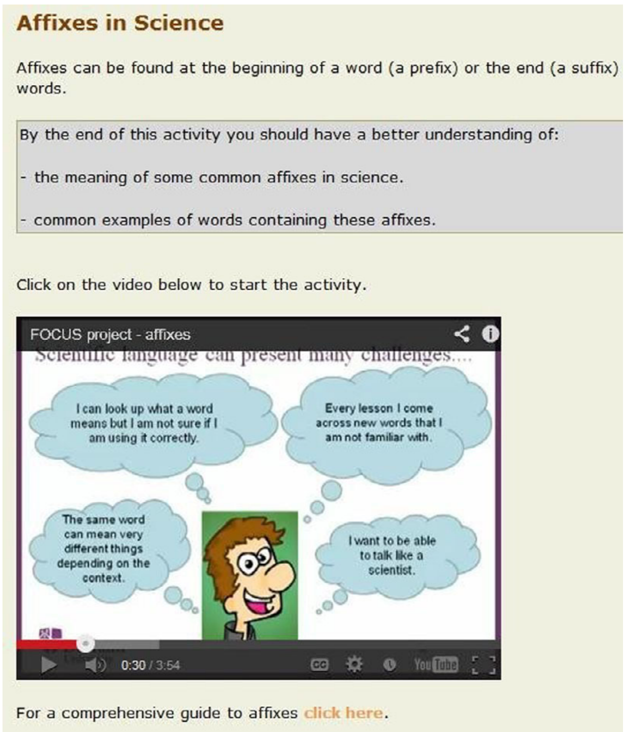


Figure 3 Screen capture of one of the activities with embedded video tutorial.

this. The basis of the activity design uses WIMBA create (an add-in to Microsoft Word) which provides a straightforward way to write interactive content using a variety of question styles with feedback. Google docs was used to write a subject-specific language questionnaire that enables the author to record and analyse the student responses. Camtasia was used to produce tutorial screencasts to guide the student through the activity. The content was further enriched through the embedding of material such as links to the FOCUS concordancing tool, relevant articles and extension material (see Figure 3).

In-class questionnaires were completed and furnished very positive feedback from the students (see Figure 4). This was supported by focus group discussions which responded that the design of the activities was clear and well organised with an appealing blend of different forms of media. The respondents felt that the activities would be useful for the next student cohort. Initial trials with FOCUS and the current student cohort are also providing encouraging results. At the end of the first term the current student cohort were asked their opinion of a variety of resources and to respond to the statement:

> ***"On a scale from 1 to 10 (1 = not useful, 10 = extremely useful), please rate the FOCUS tool used during the course in relation to improving your understanding of scientific terms and language."***

The FOCUS resource was given an average score of 6.7/10.

Conclusions

To understand the meaning of a word requires exposure to its authentic usage in a variety of different contexts. Nation (1990) concluded that "it takes from five to sixteen or more [encounters] for a word to be learned". He states that a learner needs to notice a word in context a number of times in order to understand and learn all its different characteristics. It can take a very long time to notice a word in context on up to 16 different occasions, so the development of the FOCUS concordancing tool and associated activities provides a shortcut for students to experience key words in multiple contexts. Its presentation, however, as a stand-alone tool is not sufficient to engage with students. It can be difficult for a non-specialist to appreciate the value of the tool and how it can impact on learning. The development of the activities outlined in this article, however, demonstrates how structured learning activities can be designed which are enriched by the corpus content. These activities can be used to provide meaningful subject-specific language for students.

To gain a better understanding of the significance of these language issues and to assess the impact of these strategies on student learning we are currently developing a chemical language diagnostic assessment. This assessment explores

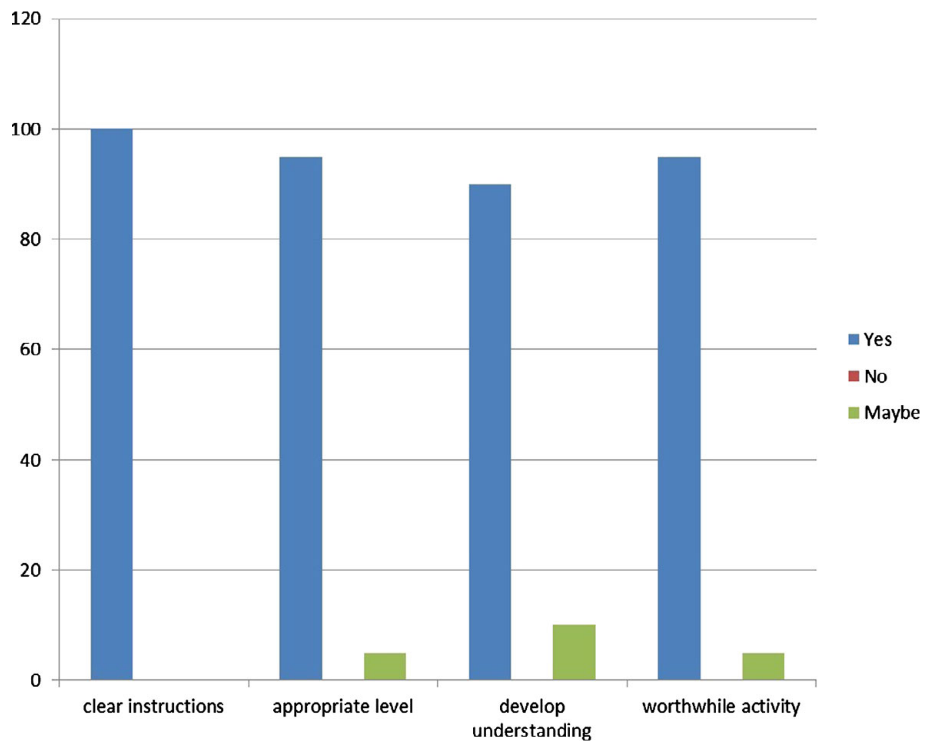


Figure 4 Results of an in-class questionnaire about the FOCUS activities.

student understanding and awareness of a range of different types of chemical language, e.g. affixes, non-technical words in a scientific context, symbolic language, and so on. It is hoped that the results of this assessment will enable us to provide even better tailored language support for our students.

Acknowledgements

The authors gratefully acknowledge the financial support received from HEA/UKCISA (initial scoping and chemistry corpus development), Durham University (bespoke FOCUS tool development) and the HEA (FOCUS activities development). The authors would also like to acknowledge the contribution of Ben Douglas for inputting student texts and Rachel Dunn for assistance in designing the activities.

References

Alsop, S. and Nesi, H. (2009) 'Issues in the development of the British Academic Written English (BAWE) Corpus'. *Corpora* **4** (1), 71–83.

Cassels, J.R.T. and Johnstone, A.H. (1985) *Words that matter in science*. London: Royal Society of Chemistry.

Childs, P.E. and O'Farrell, J.O. (2003) Learning science through English: an investigation of the vocabulary skills of native and non-native English speakers in international schools. *Chemistry Education Research and Practice* **4**, 233–247.

Gabel, D. (1999) Improving teaching and learning through chemistry education research: a look to the future. *Journal of Chemical Education* **76**, 548–554.

Herron, J.D. (1996) *The Chemistry Classroom*. Washington D.C.: American Chemical Society.

Jasien, P.G. (2010) You said "neutral", but what do you mean? *Journal of Chemical Education* **87**, 33–34.

Jasien, P.G. (2011) What do you mean that "strong" doesn't mean "powerful"? *Journal of Chemical Education* **88**, 1247–1249.

Jasien, P.G. and Oberem, G.E. (2008) Student contextual understanding of the terms sense and energy: a qualitative study. *The Chemical Educator* **13** (2), 46–53.

Johns, T.F. (1991) 'Should you be persuaded: Two examples of data-driven learning'. In *Classroom Concordancing* (eds. T.F. Johns and P. King), pp1–13. Birmingham: ELR.

Johnstone, A.H. and Selepeng, D. (2001) A language problem revisited. *Chemistry Education Research and Practice* **2**, 19–29.

Nation, I.S.P. (1990) *Teaching and learning vocabulary*. New York, NY: Heinle and Heinle.

Pyburn, D.T., Pazicni, S., Benassi, V.A. and Tappin, E.E. (2013) Assessing the relation between language comprehension and performance in general chemistry. *Chemistry Education Research and Practice* **14**, 524–541.

Rixse, J.S. and Pickering, M. (1985) Freshman chemistry as a predictor of future academic success. *Journal of Chemical Education* **62**, 313–315.

Rees, S.W., Bruce, M. and Nolan, S. (2013) Can I have a word please - strategies to improve understanding of subject specific language in non-traditional students studying foundation chemistry. *HEA: New Directions* **9** (1).

Tribble, C. (1997) Corpora, Concordances and ELT. In *New ways of using computers in language teaching* (ed. T. Boswood). Alexandria, VA: TESOL.