Spark: UAL Creative Teaching and Learning Journal

Creating and using corpora: A principled approach to identifying key language within art & design

David C. King, Acting Head of Insessional Programmes and Helen Hickey, Acting Head of Presessional Programmes, Language Centre, University of the Arts London

Abstract

A considerable body of research points to the importance of lexical knowledge for students studying, working and communicating in a second language (Carver, 1994; Hu and Nation, 2000; Schmitt and Schmitt, 2014), but decisions regarding content to prioritise can be difficult. Although there are many books aimed at teaching English for Academic Purposes (EAP), the language of and for art and design is conspicuous by its absence. Tutors face challenges in identifying relevant input texts and then creating appropriate language materials for students. This article shows how 'corpus' informed approaches can aid in the identification and selection of lexis, with relevant art-related words and vocabulary through which they can communicate their ideas and better understand the subject the learn.

Keywords

EAP; English; corpus; vocabulary; pedagogy; language

Introduction

A corpus is a systematic and principled compilation of written texts, which are analysed using various software programmes (Bennett, 2010). Using corpora offers a relatively quick and efficient way of examining the assumptions made regarding linguistic patterns, the specific phrasings of a specific topic, and aids with identifying key linguistic features of both written and spoken discourses, including:

- the most frequent words in a text (word lists)
- words that are unusually (in)frequent in a text when compared to how often they appear in other texts (keywords)
- frequent phraseological patterns (n-grams and nominalisation[1])
- the vocabulary profile of a text (the lexical spread).

Analyses of corpora provide empirical confirmation of what might have been suspected or intuitively known. At times, however, results have been known to challenge preconceptions and there have been instances where analysis has yielded genuinely surprising results. This hints at some of the complexities involved when trying to intuitively identify the features of language that are relevant for students at University of the Arts London. What we think we know and what we think our students need to know are not often self-evident and can even be incorrect.

To help mitigate such challenges, we have compiled several corpora relating to art and design courses; for example, English for Graphic Design Communication (EGDC)[2]. A separate corpus exists

---

[1] An n-gram is a phrase of, typically, three or more words that occurs frequently within a text. A 3-gram is a three-word phrase, a four-gram is a four word phrase, and so on. Nominalisation refers to changing verbs or adjectives to noun forms; e.g., identify to identification.

for learner writing from the Presessional Academic English Programme (PAEPLC)[3], another for assessment briefs (CAB)[4] and finally a gender based sub-corpus (GSC) has been collated from a larger reading list, comprising texts used in a first year taught module in Cultural and Historical Studies at London College of Fashion[5]. These have provided the foundation for a number of investigations into word lists and keywords, semantic domain, phraseology (n-grams), multi-word lexical 'chunks', nominalisation (phrases using noun forms of verbs or adjectives), pronoun use and lexical spread (the range of vocabulary used within a text).

Word lists and keywords

A key concern for EAP tutors is that the texts that are used as the basis for language input are as 'authentic' as possible. Authentic here describes the types of texts students are likely to encounter or produce during their studies. This means that in contrast to most texts used in English as a Foreign Language (EFL) classroom teaching, the texts have not been graded nor modified. The texts encountered in EAP classrooms have often been written with an English L1 readership in mind, which means that making principled decisions regarding what language to focus on in a class of English L2 speakers can be highly problematic. Software programmes can analyse texts to produce lists of words that have high surrender values (i.e. those words which are the most frequently occurring), which may at first seem appealing when deciding which lexical items to focus on. However, even a cursory examination reveals this is often not the case, as can be seen in Figure 1 (showing 'High Surrender Value'). It shows the most frequently occurring words in three corpora. The British Academic Written English Corpus (BAWE Nesi *et al*, 2007)[6] is shown alongside the two corpora the Language Centre has compiled: the Graphic Design related EGDC; and gender-based sub-corpus, GSC. It is obvious that functional lexis (i.e. prepositions, articles – so-called 'grammar words') dominate the word frequency lists. Where content words do appear, for example 'design' within the EGDC and 'women' in GSC, these are closely connected to the domain, or subject matter, of the corpora. As can be seen, relying on lexical frequency is not necessarily a useful guide when deciding upon language for classroom input and materials design.

Wordlist – high frequency					
Rank	BAWE		EGDC		GSC
1	the		the		the
2	of		of		of
3	and		and		and
4	to		to		in
5	in		a		to
6	a		in		a
7	is		is		as
8	that		that		is
9	as		design		that
10	be		as		women

Figure 1: High Surrender Value (King and Hickey, 2017).

---

[2] Although our EGDC is not huge (170,395 words from 24 texts), it is the only such corpus we know to exist.
[3] Compiled over 3 years, this corpus of presessional student essay writing contains 1,219,956 words.
[4] This corpus comprises 89 assessment briefs from all UAL colleges, across four levels from foundation through to taught masters, and totals 115,045 words.
[5] This sub-corpus consists of 50,261 words.
[6] Created between 2004 and 2007 (Nesi *et al*) at Coventry University, the BAWE comprises 2,761 assessed, written academic texts (student assignments which range from 500 words to approximately 5,000 words). The corpus totals 6,506,995 words, covering four broad disciplinary areas (Arts and Humanities, Social Sciences, Life Sciences and Physical Sciences). Thirty-five subject areas are represented from four levels; undergraduate through to taught masters.

A keyword analysis is a computation of words that are unusually (in)frequent within a text or a body of texts when compared to a larger, more generally representative body of texts and can reveal language of potentially much greater pedagogic value. The table which follows (Figure 2) lists the salient keywords for our EGDC and GSC when compared to the BAWE.

Keyword list comparison (to BAWE)			
Rank	EGDC		GSC
1	design (including designer/s)		women
2	graphic		fashion
3	visual		gaze
4	typography		media
5	art		beauty
6	color		salon
7	communication		magazines
8	elements		dress
9	printing		body
10	book		gender

Figure 2: Keyword list comparison (King and Hickey, 2017).

The results may, on the surface, seem unsurprising. One might reasonably expect 'typography' to be an especially salient word in a body of texts on graphic design. Similarly, one might be unsurprised to see 'women' featuring so prominently in a body of texts on gender. What such lists do highlight is the relative ranking of lexical items, and it may be surprising that 'typography' does not appear either higher or lower in the list. The appearance of 'book' within the top ten might raise questions as to where 'poster', 'map' or 'website' are, and whether this says anything about graphic design itself and its main concerns, or whether this reflects the age and provenance of the texts included in the corpus[7]. Similarly, one might reasonably ask where 'men' are in a corpus on gender. The absence of lexis related to the semantic domain of transgender may also call into question the topicality and relevance of the texts included in the Cultural and Historical Studies Reader.

Within the UAL Language Centre such keyword analyses have been instrumental in devising specific Academic English Skills courses, as well as in informing the materials design of both our Presessional Academic English Programme (PAEP) and Insessional language provision. We believe that the value of such analyses can be extended beyond the Language Centre; that these could for example result in the production of course-specific glossaries of key terminology. If compiled in collaboration with subject experts (e.g. lecturers, course leaders, subject-specialist librarians, etc.) we could provide the requisite linguistic expertise to ensure that such a resource would be relevant to the students' needs.

Semantic domains

Attempts have been made to identify key academic vocabulary, most notably Coxhead's (2016) Academic Word List (AWL). The AWL identifies '570 word families that account for approximately 10.0% of the total words (tokens) in academic texts but only 1.4% of the total words in a fiction collection of the same size' (Coxhead, 2000, p.213). The corpus the AWL is derived from contains 3,500,000 words from academic journals, textbooks, course workbooks, lab manuals and course notes. As such, taking vocabulary items from the AWL to inform classroom input and materials design

---

[7] The texts included in our EGDC came from suggestions from BA and MA discipline tutors at CSM and Chelsea, as well as from recommendations from subject specialist librarians.

can be useful, but this is not always as straightforward as it first seems. This is because the words on the AWL frequently occur in a range of academic texts, meaning they tend to be very general in nature, and are often not directly connected with any particular subject. The following very small, random sample of words (Figure 3), exemplifies properties and characteristics that need to be considered when teaching vocabulary.

AWL	Primary semantic field	
	BAWE	UAL
found	(vb) discovered; as in a scientific discovery	(vb) physically located, as in a space or within a work of art (PAEPLC)
image	(n) a look; as in an illusory appearance, or an attempt to conform to an expected appearance	(n) a material representation, as in a photograph or a picture (PAEPLC)
pose	(vb) to set forth or to come to attention; as in questions, risks, threats, challenges	(n) a sustained posture, as in one assumed for artistic effect (GSC)
vision	(n) sight, as in physiology; a sensory experience	(n) seeing, as in a feature of gaze and power relations (GSC)

Figure 3: AWL and Semantic Domain (King and Hickey, 2017).

Even these simple examples suggest the importance of not assuming that a student's knowledge of a word extend into other semantic domains, and thus it is essential for a teacher to consider a word's co-textual and contextual features and relations. Corpus-informed approaches to text analysis can be invaluable in this regard[8]. This undoubtedly extends to those lexical items which exhibit more complex relationships within various art and design discourses, and these almost certainly demand explicit instruction.

Phraseology: n-grams and nominalisation

Formulaic language (e.g. fixed phrases, collocations, situationally-bound expressions etc.) is almost universally found in EFL course-books and often taught in terms of exam preparation. Considering the attention paid to the teaching of such phrases within EFL and exam preparation training, one might suspect that students for whom English is not their first language would have developed a discernible ability in deploying an array of these phrases in their writing. Certainly, within the BAWE, formulaic language can be frequently seen and is a noticeable feature of these texts. Surprisingly, this is not evidenced to nearly the same extent within our students' writing (PAEPLC). Analysis revealed that a significant proportion of the most common lexical 'chunks' produced by our students came directly from the rubric of a PAEP assessment brief, which asked students to respond to the statement – 'A picture tells a thousand words' when analysing visual imagery (their responses are shown in bold italics in Figure 4 below). There is much less evidence of appropriate use of other formulaic language. As can be seen in Figure 4, many of the phrases prevalent within the BAWE serve vital rhetorical purposes (e.g. cause and effect, compare and contrast, problem and solution, exemplification, etc.) and their appropriate use can enhance both coherence and cohesion within students' writing.

n-gram	BAWE		PAEPLC
3-gram	in order to		*a thousand words*
	as well as		*of the image*

---

[8] It is worthwhile noting that the use of corpora for such purposes is not limited to tutors. Students can be instructed and encouraged to construct their own corpora, for example of their own writing or texts they are reading, which they can then analyse for themselves. A key outcome of this is that it can aid student-centred learning and further student autonomy (Guan, 2013, p.111).

	due to the		one of the
	one of the		there is a
	the use of		*tells a thousand*
4-gram	as a result of		*tells a thousand words*
	the end of the		On the other hand
	On the other hand		*picture tells a thousand*
	as well as the		is one of the
	in the form of		one of the most
5-gram	at the end of the		*a picture tells a thousand*
	due to the fact that		No other 5-grams present
	it can be seen that		No other 5-grams present

Figure 4: N-gram Analysis (King and Hickey, 2017).

It is also worth noting that of the remaining five n-grams not directly lifted from the rubric, students over-relied on 'one of the', with a variant of this phrase appearing three times among the list of most common n-grams ('one of the'; 'is one of the'; 'one of the most'). In effect, students' use of important rhetorical and cohesive devices appears to be severely limited. These results serve as a timely reminder for anyone involved in language instruction of the need to explicitly teach, or at least review, these seemingly basic phrases and the focus should not always be on subject specific, technical, or academic language.

As many have noted, (Cooper, 2012; Hewings, McCarthy and Thaine, 2012), nominalisation is generally considered an important feature of academic writing and it frequently appears in English for Academic Purposes course-books and reference books. Typically, nominalisation results from changing verbs or adjectives into nouns and employs the relatively fixed syntax of 'a/an/the *noun* of a/an/the *noun*'. There has, however, been extremely limited analysis of this within language for art and design. The only attempt of which we are aware is 'International Art English' by Rule and Levine (2012), but this limits itself to art-world press releases. Analysis of our own students' writing in the PAEPLC reveals an ability to replicate explicitly taught nominalisations (following the aforementioned pattern of 'a/an/the *noun* of a/an/the *noun*' (e.g. 'the *denotation* of the *image*' or 'the *connotation* of the *colour*'). However, our analysis also highlighted that although the students could repeat what had been explicitly taught they were unable to manipulate the grammar structure and create an extended range of nominalised phrasing. As a result, we have since developed materials for the PAEP which explicitly introduce students to this feature of academic writing within the discourses of art and design.

Pronoun use: 1st and 2nd person usage
Many EAP course-books and EAP reference materials advise against the use of first and second personal pronouns (e.g. 'I', 'me', 'you', etc.), which may be appropriate advice for many disciplines, particularly those more deeply rooted within positivist paradigms. However, our analysis of the EGDC shows that 'you' is an important keyword within texts on graphic design. Less salient, but still important keywords, were both 'I' and 'my'. Supplementary analysis of written materials from an Exceed workshop on Graphic Branding and Identity revealed both 'you' and 'your' to be within the top 25 keywords. Given that much of the published literature advises against using first and second person pronouns in academic writing, this was surprising, but when asked, discipline tutors confirmed that a reasonable use of first and second person personal pronouns and possessive adjectives (e.g. 'my', 'your', 'our') was considered an appropriate feature of speaking and writing within their

disciplines. At UAL students' practices as artists and designers are often seen as an extension of self, or at the very least, an expression of self, so personal pronouns such as 'I', 'me', and 'my' are not often viewed as something to be avoided. Interestingly, anecdotal evidence gleaned from discussions with UAL postgraduate European students indicates that most had been previously instructed by lecturers on their undergraduate degrees in their home countries to avoid the use of 'I' in academic writing and many expressed a deep sense of unease and discomfort with writing in the first person. However, avoiding these pronouns and possessive adjectives often serves to detract from the expression of the personal and this can be problematic because, as mentioned earlier, such expression tends to be highly valued within art and design discourses. This is, of course, by no means universal. We are aware of a number of disciplines within the university in which the use of personal pronouns is actively discouraged. This does raise a number of related questions though, such as whether this is personal stylistic preference on the part of degree tutors or whether it is a specific feature of the genres reproduced within a particular discourse community. It may also result from general (mis?)perceptions of what constitutes acceptable academic writing. Tutors and students alike may be labouring under more traditional assumptions of what constitutes suitably 'academic' writing in varying circumstances across the many genres demanded of students.

Lexical spread
The lexical spread of a text generally indexes the level of vocabulary knowledge required for a reader to comprehend a text. The New General Service List (NGSL) is a wordlist derived from the two billion words contained on the Cambridge English Corpus. Created by Browne, Culligan and Phillips, the NGSL contains approximately 2,800 'core high frequency vocabulary words for students of English as a second language' (2013). Although typically seen as a list of 'general' vocabulary, analysis indicates that the first three-thousand most common words in English (K1 – K3) includes approximately 64% of the words in the Academic Word List (Cobb, 2010), and so the NGSL can be seen as potentially relevant for English for Academic Purposes.

In terms of text comprehension, it had been suggested that a reader needs to have passive knowledge of at least 95% of the language in a text (Laufer, 1989), but more recent research has revised this figure upwards and it is now thought that readers typically require vocabulary recognition of 98–99% of a text (Hu and Nation, 2000). In other words, between one in fifty and one in a hundred words can be unknown before comprehension is impaired, making understanding texts in a foreign language notably difficult (Carver, 1994).

The ability of many students who do not have English as a first language to effectively engage with written materials is further compounded by entry level requirements. Most BA programmes at the university ask students who do not have English as a first language to provide proof of attainment of IELTS 6. This corresponds with level B2 in the Common European Framework of Reference (CEFR), yet according to our own analysis, 5.1% of the items in the NGSL occur at higher levels (3% at C1 and 2.1% at C2). In addition, many of the key terms situated within a particular field of discourse exist 'off-list' or beyond level C2, thereby creating further challenges.

This clearly implies that much more needs to be done to ensure that students whose first language is not English are explicitly exposed to and taught vocabulary that will bring them to this 98–99% threshold. Unfortunately, research has highlighted 'the lack of a principled approach to teaching mid-frequency vocabulary' (Schmitt and Schmitt, 2014, p.498). Mid-frequency vocabulary can be operationalised as those lexical items occurring within the spread of 4,000–8,000 word level (K4 – K8) from the New General Service List. At this point learners are moving out of the most common vocabulary and into the range of vocabulary Schmitt and Schmitt (2014) are suggesting should be taught to bring them to the required threshold to understand an academic text. Our own corpus based

research shows the potential gains to be had when this mid-frequency range of vocabulary is explicitly introduced in the classroom. Initial analysis of the PAEPLC reveals that slightly over 14% of students' writing contained vocabulary from the mid-frequency range and that, encouragingly, a significant number of these items had been repeatedly encountered by our students via our Presessional Core Materials. These encounters were established according to four core principles:

- Target language should be presented within texts that are as 'authentic'[9] as possible.
- The texts should also be as relevant as possible to the students' immediate and future academic needs, following a key principle of 'situated cognition' (Brown, Collins and Duguid, 1989).
- Students also require repeated exposure to key lexical items (Nation, 1990; Barcroft, 2004).
- Students need to be 'scaffolded' towards actively using the target vocabulary in order for them to move beyond recognition (passive knowledge) to controlled application (active knowledge).

The mid-frequency lexical items uncovered by our research are embedded within contexts relevant to art and design students, and are often repeated. Nation (1990) indicates the value of such repeated exposure. The table which follows (Figure 5) provides a sample of mid-frequency words and how often they explicitly occurred within our PAEP materials.

K4	metaphor (metaphorical)	10
	celebrity	17
K5	signifier	18
	signified	17
K6	denotation (denote)	5
	mythology (myth)	12
K7	juxtaposition	5

Figure 5: Explicit Lexical Exposure (King and Hickey, 2017).

Many researchers have tried – with varied results – to determine the ideal number of times that a learner needs to encounter a word to actually learn it (Horst, Cobb, and Meara, 1998; Hulstijn, Hollander and Greidanus, 1996; Pigada and Schmitt, 2006; Rott, 1999; Waring and Takaki, 2003; Webb, 2007). Research into reading indicates that new words need to be seen around 8 to 10 times in order to be learned (Schmitt, 2008; Teng, 2016). The figures in Figure 5 reflect exposure via reading texts and do not include how often students would have heard the words used in listening exercises, lectures and in the classroom, or how often they would have used them in their own discussions. It is reasonable to assume then that students' 'true' exposure would have far exceeded this already substantial exposure.

We believe that we have implemented a principled approach to expanding our students' exposure to mid-frequency vocabulary, and that this is showing indications of success. The implications for non-language classrooms is that repeated exposure to words that are otherwise mid-low frequency is

---

[9] What constitutes 'authentic' can be contentious, and limitations on this paper prevent a thorough exploration of this, but for us, an authentic academic text is one which has been written with, 'a native speaker in mind' (Nation and Bonesteel, 2010, p.1) and which 'needs to be situated to some extent in its intended academic context' (de Chazal, 2014).

essential for students for whom English is their L2, particularly if the appropriate use of such language is seen as marking one out as a member of the discourse community in question[10].

Conclusion

Corpus-based approaches to analysing texts can provide a wealth of linguistic information. Wordlists of the most frequently occurring vocabulary can be created quickly and used to identify which words are particularly salient to a text (keywords). Analyses can also reveal the different ways in which words inhabit semantic domains and highlight the use (or non-use) of phraseological patterns, for example rhetorical phrasing or nominalisation. Analysis of a corpus can also provide empirical evidence of specific features of discourse, placing the practitioner at the heart of writing.

The implications of our research are numerous, but of primary importance is the fact that as educators we cannot assume that students who have attained the requisite English language entry requirements have the necessary breadth and depth of language knowledge to successfully navigate the vast array of texts they will encounter, and be required to produce, at university. For language teachers – who are not experts in the range of art and design disciplines within the university – it is imperative that we find willing partners in subject tutors, course leaders and subject specialist librarians in order to provide students with targeted, relevant support that will expand their language portfolios. For subject tutors, we believe a heightened awareness of the implications of their language choices (in writing assessment briefs, in providing feedback, in delivering lectures and so forth) is required, but that this need not occur in isolation. A wealth of linguistic expertise resides within the Language Centre and drawing upon this, by working in ever closer collaboration with language tutors, will ensure that the language challenges faced by students are better addressed. Finally, we believe that it could be of immense benefit if students were taught and encouraged to construct and analyse corpora of their own writing, or of texts they encounter, in order to further their own language development as autonomous learners. We believe this warrants further research to determine whether this would in fact, lead to improvements in student writing as indicated by this study.

References

Barcroft, J. (2004) 'Second language vocabulary acquisition: a lexical input processing approach', *Foreign Language Annals*, 37(2), pp.200-208. https://dx.doi.org/10.1111/j.1944-9720.2004.tb02193.x.

Bennett, G. (2010) *Using corpora in the language learning classroom*. Ann Arbor: Michigan University Press.

Brown, J.S., Collins, A. and Duguid, P. (1989) 'Situated cognition and the culture of learning?', *Educational Researcher*, 18(1), pp.32-42.

Browne, C., Culligan, B. and Phillips, J. (2013) *The new general service list*. Available at: http://www.newgeneralservicelist.org (Accessed: 6 June 2017).

Carver, R.P. (1994) 'Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: implications for instruction', *Journal of Literacy Research*, 26(4), pp.413-437. https://dx.doi.org/10.1080/10862969409547861.

Cobb, T. (2010) 'Learning about language and learners from computer programs', *Reading in a Foreign Language* 22(1), pp.181-200.

---

[10] The Language Centre can provide training in corpus-informed approaches to discourse analysis.

Cooper, J. (2012) 'Nominalization', *Academic English online*, Queen Mary University London. Available at: http://aeo.sllf.qmul.ac.uk/Files/Nominalization/Nom%20LOC.html (Accessed: 6 June 2017).

Coxhead, A. (2000) 'A new academic word list', *TESOL Quarterly*, 34(2), pp.213-238. https://dx.doi.org/10.2307/3587951.

Coxhead, A. (2016) *The Academic Word List*. Available at: http://www.victoria.ac.nz/lals/resources/academicwordlist/information (Accessed: 6 June 2017).

de Chazal, E. (2014) 'Using authentic texts in the EAP classroom', *Oxford University Press: English Language Teaching global blog*. Available at: https://oupeltglobalblog.com/tag/authentic-texts/ (Accessed: 6 June 2017).

Guan, X. (2013) 'A study on the application of data-driven learning in vocabulary teaching and learning in China's EFL class', *Journal of Language Teaching and Research*, 4(1), pp.105-112.

Hewings, M., McCarthy, M. and Thaine, C. (2012) *Cambridge academic English student's book [C1 advanced]: an integrated skills course for EAP*. Cambridge: Cambridge University Press.

Horst, M., Cobb, T. and Meara, P. (1998) 'Beyond a clockwork orange: acquiring second language vocabulary through reading', *Reading in a Foreign Language*, 11(2), pp.207-223.

Hu, M. and Nation, I.S.P. (2000) 'Vocabulary density and reading comprehension', *Reading in a Foreign Language*, 23(1), pp.403-430.

Hulstijn, J.H., Hollander, M. and Greidanus, T. (1996) 'Incidental vocabulary learning by advanced foreign language students: the influence of marginal glosses, dictionary use, and reoccurrence of unknown words', *The Modern Language Journal,* 80(3), pp.327-339. https://dx.doi.org/10.2307/329439.

Laufer, B. (1989) 'What percentage of text-lexis is essential for comprehension?' in Laurén, C. and Nordman, M. (eds.) *Special language: from humans to thinking machines*. Clevedon: Multilingual Matters, pp.316-323.

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

Nation, I.S.P. and Bonesteel, L. (2010) *The authentic reading experience: Building reading comprehension and fluency*. Available at: http://www.pearsonlongman.com/tertiaryplace/pdf/authentic_reading_experience.pdf (Accessed: 6 June 2017).

Nesi, H., Gardner, S., Thompson, P. and Wickens, P. (2007) *British Academic Written English Corpus*. University of Oxford. Available at: http://ota.ahds.ac.uk/headers/2539.xml (Accessed: 6 June 2017).

Pigada, M. and Schmitt, N. (2006) 'Vocabulary acquisition from extensive reading: a case study', *Reading in a Foreign Language*, 18(1), pp.1-28.

Robbins P. and Aydede, M. (eds.) (2012) *The Cambridge handbook of situated cognition*. Cambridge: Cambridge University Press. https://dx.doi.org/10.1017/cbo9780511816826.025.

Rott, S. (1999) 'The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition through reading', *Studies in Second Language Acquisition*, 21(4), pp.589-619. https://dx.doi.org/10.1017/s0272263199004039.

Rule, A. and Levine, D. (2012) 'International art English: on the rise, and the space, of the art world press release', *Triple Canopy*, 16, pp.7-30. Available at: http://canopycanopycanopy.com/16/international_art_english (Accessed: 6 June 2016).

Schmitt, N. (2008) 'Instructed second language vocabulary learning', *Language Teaching Research*, 12(3), pp.329-363. https://doi.org/10.1177/1362168808089921.
Schmitt, N. and Schmitt, D. (2014) 'A reassessment of frequency and vocabulary size in L2 vocabulary teaching', *Language Teaching*, 47(4), pp.484-503. https://doi.org/10.1017/S0261444812000018.

Teng, F. (2016) 'The effects of context and word exposure frequency on incidental vocabulary acquisition and retention through reading', *The Language Learning Journal*, 44(4), pp.1-14. https://dx.doi.org/10.1080/09571736.2016.1244217.

Waring, R. and Takaki, M. (2003) 'At what rate do learners learn and retain new vocabulary from reading a graded reader?', *Reading in a Foreign Language*, 15(2), pp.130-163.

Webb, S. (2007) 'The effects of repetition on vocabulary knowledge', *Applied Linguistics*, 28(1), pp.46-65. https://doi.org/10.1093/applin/aml048.

Biographies
David C. King is Acting Head of Insessional programmes, based in UAL's Language Centre. David has an MA in Applied Linguistics from King's College London. His research interests include genres of academic writing within art and design, corpus linguistics, and materials design.

Helen Hickey is Acting Head of Presessional Programmes in the Language Centre at UAL. Helen has an MA in Applied Linguistics and English Language Teaching from King's College London. She has supported a range of degree courses from Foundation through to postgraduate in a variety of disciplines (Product Design, Fashion, Fine Art, Photography). Her current research interest is the compilation of corpora and their application to materials design and the classroom.