Title	Position vectors, homologous chromosomes and gamma rays: Promoting disciplinary literacy through Secondary Phrase Lists
Author(s)	Clarence Green and James Lambert
Source	*English for Specific Purposes*, *53*, 1-12
Published by	Elsevier

Copyright © 2019 Elsevier

This is the author's accepted manuscript (post-print) of a work that was accepted for publication in the following source:

Green, C., & Lambert, J. (2019). Position vectors, homologous chromosomes and gamma rays: Promoting disciplinary literacy through Secondary Phrase Lists. *English for Specific Purposes*, *53*, 1-12. https://dx.doi.org/10.1016/j.esp.2018.08.004

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source.

[Draft- may differ from the copy of record]

Position vectors, homologous chromosomes and gamma rays: Promoting disciplinary literacy through Secondary Phrase Lists

Published after review as:

Green, C., & Lambert, J. (2019). Position vectors, homologous chromosomes and gamma rays: Promoting disciplinary literacy through Secondary Phrase Lists. *English for Specific Purposes*, *53*, 1-12. DOI: 10.1016/j.esp.2018.08.004

1. Introduction

Disciplinary literacy is a pedagogy and research area that has been rapidly spreading across educational systems worldwide. As a pedagogical orientation, it has been adopted in multiple contexts of education from the primary to the tertiary level (Airey et al., 2017; Wilson, Madjar, & McNaughton, 2016). As a research area, disciplinary literacy overlaps significantly with English for Specific Purposes. Both have generated an evidence base of research suggesting effective instruction needs to take into account the different language practices of the disciplines, rather than minimize these differences under an increasingly challenged construct of a general academic literacy (Hyland, 2017; Shanahan & Shanahan, 2017). Disciplinary language varies at many levels, from the discourse moves instantiated by formulaic language (Biber et al., 2004), to the conceptual knowledge accessed via the lexico-grammar (Martin, 2013). An emphasis on developing conceptual knowledge though language is a key tenet of disciplinary literacy (Britt, Richter, & Rouet, 2016; Martin, 2013). Content words, for example, and their patterns of use provide access to the conceptual knowledge needed for educational success in many subjects (Slobin, 2003; Balota, 2016). One resource that has been proven particularly useful in education has been corpus-derived wordlists for instructed vocabulary acquisition (Schmitt, 2008; Nagy & Townsend 2012). These wordlists, particularly when discipline-specific, can be used pedagogically to facilitate conceptual access to the discipline by providing a distillation of a subject's core lexis (Todd, 2017). In other words, a high quality wordlist is both a vocabulary learning instrument and a pedagogical resource that offers a lexical profile of a discipline (Nation, 2016). Wordlists that can be used by teachers to promote disciplinary literacy at the secondary school level are, however, currently lacking. In order to provide such a resource we have produced discipline-specific lists of bigrams (two-word collocates) for Biology, Chemistry, Economics, English, Geography, History, Mathematics, and Physics, for the following lexcio-grammatical combinations: noun+noun, adjective+noun, noun+verb, verb+noun, and verb+adverb. The first two of these capture discipline-specific noun phrases, the second two capture the most common discipline-specific verb/subject and verb/object pairs, and the last captures the most common discipline-specific phrasal verbs.

Two recent trends have occurred in the long tradition of building academic wordlists. The first is that the usefulness of general academic wordlists has been questioned in the context of debates over whether there is a core academic vocabulary (Stein, 2016). The field has therefore seen

a movement toward discipline-specific pedagogical wordlists drawing on the current state of the art of corpus linguistics (Lei & Lui, 2016). The second trend has been to move beyond single-word resources and to develop pedagogical phrase lists. The functions and importance of frequent, multi-word phraseology have been long studied (Biber et al., 2004; Hyland & Tse, 2007; Sinclair, 1991), yet only recently has there been a specific focus on developing pedagogical resources that capture them. Two recent examples that represent the state of the art include the PHRASE List of Martinez and Schmitt (2012) and the Academic Formulas List of Simpson-Vlach and Ellis (2010). The former is a general phrase list, while the latter is restricted to academic language. To date, however, no discipline-specific pedagogical phrase lists have been developed. Furthermore, phrase lists tend to be a mixture of lexical content words and function words, highlighting discourse moves rather than concept-heavy lexical words as one finds in discipline-specific wordlists (Hyland, 2008; Simpson-Vlach & Ellis, 2010). Durrant (2009), for example, developed a list of 1000 academic two-word bigrams, of which 763 were lexical words combined with function words. He notes that teachers might find this disappointing if they are looking for the more concept-carrying phrases, e.g. *economic development* or *gravitational fields*. With respect to this, the bigrams generated by the present research will be less disappointing.

The current research fills several gaps in the theory and practice of disciplinary literacy. Firstly, this paper reports on the development of the eight discipline-specific pedagogical phrase lists, based on a 16.5 million word corpus of secondary-level textbooks for the following core subjects: Biology, Chemistry, Economics, English, Geography, History, Mathematics, and Physics. These pedagogical resources consist of frequent discipline-specific two word phrases vetted for usefulness according to frequency, dispersion, ratio, and mutual information (MI) criteria. Unlike previous phrase lists, these contain part of speech information and are controlled to lexical word combinations. This is a significant resource that fills an important gap between discipline-specific wordlists of single words that capture the prominent concepts of disciplines, and phrase lists which tend to capture discourse moves of the disciplines. The current list presents a unique range of both concept-bearing language and discourse patterns relevant each discipline. Secondly, while the technical advances in wordlist development and the insights gained about discipline-specific language in the context of ESP has previously focused on the tertiary educational context, the present research realises the need to bring these methods and resources into new contexts and has changed the focus to secondary-school disciplinary literacy. Finally, the paper contributes to theory as well as practice. The debate over a general academic language has been heated (Hyland, 2008; Brezina, & Gablasova, 2017), with growing evidence suggesting words and even more so phrases vary across disciplines so much that a discipline-specific approach to education should be adopted (Hyland, 2008). This paper shows that when one considered phrases controlled for the four major parts of speech, the disciplinary variation in academic language is even more marked than previously studies have reported. Thus, this study not only produces educational resources for disciplinary literacy, but contributes to the case for the need for disciplinary literacy.

2. Discipline-specific language and the educational turn to disciplinary literacy

Disciplinary literacy has been defined as "learning how to read, think about, write, communicate, and use information like each discipline's experts" (Zygouris-Coe, 2012: 36). In the classroom,

disciplinary literacy entails seeing each subject area as a community of practice (Lave & Wegener, 1991), in which the role of the teacher is to facilitate entry for the students by helping them understand the discourse of the discipline and how to use language to access its conceptual knowledge. To gain expertise in a discipline, a student must acquire the concept-bearing technical words of that discipline, yet they also must learn how words that are formally similar across subjects may create different meanings when used in different disciplines. For example, Hyland and Tse (2007) note that while the word 'process' is frequent in both humanities and science subjects, but in the humanities it if more often a verb and in the sciences a noun. From a cognitive linguistics standpoint, this indicates that the word has distinct cognitive profiles (Langacker, 2008).

Shanahan and Shanahan (2008, 2017) argue that literacy development involves three developmental trends, with disciplinary literacy being the desired final outcome. Initially, learners learn high frequency words that are used across all disciplines, such as the Dolch (1936) and Fry (1957) lists, or the top 2000 words of the General Service List (West, 1953). In some educational contexts, these may not need to be the focus of explicit instruction but can be learned implicitly (Schmitt, 2008). Secondly, learners build fluency by focussing on the vocabulary, grammar and registers that are a common in academic schooling. Amongst other things, Shanahan and Shanahan (2008) would perhaps include here the words on the Academic Wordlist of Coxhead (2000) and perhaps even the Academic Formulas List (Simpson-Vlach & Ellis, 2010). Finally, students acquire the specialized language patterns of the subject matter and it is this final step in which students become disciplinary literate. It is only the attainment of disciplinary literacy that allows a student to succeed in any give discipline (Gillis, 2014). The disciplinary literacy approach has recently been adopted as a model in the United States, Norway, New Zealand, Singapore, Australia and elsewhere (Airey et al., 2017; Shanahan & Shanahan, 2017), demonstrating its growing impact on international education. Disciplinary literacy is a broad educational and research paradigm (Gillis, 2016; Koomen et al., 2016), but a core component has been the emphasis on the language of the disciplines (Goldman et al., 2016). This is where the technical work of linguists has an important role to play, as the advances in the field of English for Specific Purposes and corpus linguistics can be leveraged for disciplinary literacy.

Schmitt (2008) states resources that target vocabulary provide the foundation for explicit vocabulary teaching and this "almost always leads to greater and faster gains, with a better chance of retention and of reaching productive levels of mastery". This raises the vital question of what vocabulary needs to be taught. Hyland and Tse (2007: 113) questioned the received wisdom of a general academic literacy: "[I]t is by no means certain that there is a single literacy which students need to acquire to participate in academic environments". They examined the coverage of the Academic Wordlist (Coxhead, 2000) in a 3.3 million word corpus of academic texts with subcorpora of the following disciplines: Biology, Physics, Computer Science, Engineering, Sociology, Business and Applied Linguistics. Hyland and Tse (2007) found that although the Academic Wordlist covered slightly above 10% of overall words, once the corpus was examined by discipline, this coverage varied significantly, and that 27% of the terms in the Academic Wordlist have an extremely low chance of being encountered in at least one academic field. When collocations are considered the profiles of the disciplines are thrown into even sharper relief. The word 'strategy', for instance, is frequent across all disciplines, but in Sociology it is preceded by 'coping' 31% of the time, in Applied Linguistics by 'learning' 9% of the time, and in Business by 'marketing' 11% of the time. These collocational differences represent different conceptual units. A *coping strategy*, *learning strategy*

and *marketing strategy* are discipline-specific conceptual phrases made up of lexical word combinations. Hyland and Tse (2007: 251) conclude "the best way to prepare students for their studies is not to search for overarching, universally appropriate teaching items, but to provide them with an understanding of the features of the discourse they will encounter in their particular courses".

Similar insights on the extent of discipline-specific language variation have been found via other studies of corpus-derived pedagogical wordlists. Such lists provide lexical profiles of disciplinary areas which can then be compared to other language use domains. For example, Lei and Lui (2016) developed a pedagogical wordlist of medical terms using a 2.7 million word corpus of journal articles. A range of metrics adapted from Gardner and Davies (2013) were used to determine which words in the corpus profiled the discipline and were therefore of high pedagogical value. These metrics included a minimum frequency of 28.57 per million words, a minimum range and dispersion across the corpus, as well as criteria such as whether the word could be found in a specialised medical dictionary. The final medical wordlist, when matched in their source corpus, was found to cover 20.18% of words in the data, indicating it contained about one in every five words a medical student would encounter during professional reading. However, the wordlist covered only 6.64% of other academic areas in the British National Corpus (BNC), and only 3.69% of the BNC more generally. These figures demonstrate quite clearly the large differences amongst discipline-specific language, general academic language, and general language use.

Over the past decade or so, therefore, English for Specific Purposes, using the tools of corpus linguistics, has advanced the methodologies for investigating discipline-specific language, and yet there has been little cross-fertilization with the field of disciplinary literacy. Perhaps this is because of a certain academic fragmentation in which disciplinary literacy has focused on primary and secondary school native speaker educational contexts and the research has been by academics in education, while ESP has focussed on tertiary, non-native speaker education by academics in (applied) linguistics. The current researchers, however, argue that one can see disciplinary literacy as the wider paradigm, in which all educational levels and speaker identities can be included, whether native or non-native speaker. As Wingate (2012: 2) states, there has been somewhat of a "failure to recognise that both native- and non-native speakers of English are equally novices to academic writing" and that "every new student needs to learn the specific conventions and discourses of their discipline is inappropriate."

3. From word-list to phrase list

The importance and utility of corpus-derived wordlists is long established in educational and applied linguistics (Nation, 2016). The basic insight is that frequency reflects usefulness. Words that occur frequently are worth learning first as they cover more text than words that occur less frequently and thus provide greater text comprehension (Coxhead, 2016). A commonly estimated number, noted in Schmitt's review of instructed vocabulary acquisition (2008: 332), is that the most frequent 8000 word families in English should enable a reader to comprehend about 98% of words in general texts. This figure is, however, an average for a general corpus and the extent of coverage is reduced with discipline-specific corpora. Here, the high frequency words of a discipline are required, underpinning

the recent trend to move from general wordlists to discipline-specific ones (Todd, 2017; Lei & Liu, 2016 Martínez, Beck, & Panza, 2009).

Frequency progression as a language learning pedagogy is an effective learning mechanism, supported by theory, experiment and classroom intervention studies (Nation & Waring, 1997; Coxhead, 2016). There is a clear relationship between language proficiency and knowledge of words at different frequency bands (e.g. the first most frequent 1000 words, the second most frequent 1000 words, etc.). The widely used Vocabulary Levels Test (Laufer & Nation, 1999) show that learners of lower proficiency score worse on accuracy and recognition of words taken from lower frequency bands (Jiang, 2010). Frequency progression is a useful in planning curricula that approximate comprehensible input based on the assumption that once higher frequency items are learned, the learner is developmentally ready for lower frequency words (Saito, Webb, Trofimovich, & Isaacs, 2016). As Nation and Waring (1997: 10) note "there is a very large number of studies showing the effectiveness of such learning in terms of amount and speed of learning".

Nation (2016) points out that pedagogical frequency lists are invaluable reference resources for a teacher's professional knowledge. From this perspective, there is no need to limit the length of a pedagogical word or phrase list to make it manageable for students (Durrant, 2009; Simpson-Vlach & Ellis, 2010), any more than a thesaurus or dictionary on a teacher's desk need be limited to only a handful of items. Nation (2016: 146) notes that even if never used in a class activity, a good frequency list "can help in studies of English for Academic purposes courses and can support teachers' awareness of words that may need attention". In regards to this, it has been demonstrated that effective teaching involves, among other things, repetition of key terms and their phonological stress, and emphasis on key concept-bearing words and phrases (Fraser, 2015; Smart & Marshall, 2013). While many good teachers may intuitively know some words and phrases which need to be emphasised in an effective class discussion, corpus-derived reference lists highlighting the most frequent terms in a discipline offer a handy resource for lesson planning and instruction. For instance, our lists note that the lexical phrase 'homologous chromosomes' will be frequently encountered by students in biology, and that it has a strong MI indicating its importance as a disciplinary concept (Simpson-Vlach & Ellis, 2010).

A number of discipline-specific wordlists for adults have recently been released, including wordlists for engineering (Todd 2017), medicine (Lei & Lui, 2016), business (Konstantakis 2007), of agriculture (Martinez et al. 2009), nursing (Yang 2015), and others (Coxhead & Hirsh 2007). Such wordlists represent the discipline-specific vocabulary students require to access conceptual knowledge in the field. Martin (2013: 23) notes that each discipline has its own 'power words' which carry the key concepts of its discipline, and these words vary in technicality and abstraction across subjects. In a comparative study of power words in secondary school books, Martin (2013: 29) argues:

> History also makes use of a number of technical terms – terms like *trade*, *economy*, *society* and culture. These terms to not refer to concrete entities and so cannot be learned ostensively; they have to be construed through language… [in Biology] depending on the unit of work, textbook or exam to hand, *society* seems to cover some or all of economy, culture, social structure, politics, religion and possibly other things as well.

Phrases also carry this discipline-specific concept bearing role, as pointed out in a follow-up intervention study that aimed to raise awareness in teachers of the power words in the discipline of Biology. Macnaught, Maton, Martin, & Matruglio (2013: 7) argue that the role of biology teachers is to not only teach what *cell division* or *cell replication* means, but to overcome "the challenge of teaching students to construct a wide range of specialised biological meanings *through* the language of Biology". In a case study of a Year 11 biology teacher they report that the teacher was able to construct this type of knowledge using strategies such as nominalisation; for example, having teachers give their students a process such as dividing [the] cell, and having them nominalize this into the concept *cell division*.

Pedagogical phrase lists are now beginning to be developed to supplement wordlists, though as yet these have been discipline specific nor focus on nominal phrases of the CELL DIVISION kind. The importance of phraseology in language, otherwise known as lexical bundles or formulaic language has long been recognized as a property of language as central as single vocabulary items (Sinclair, 1991; Wray, 2013). Two major recent phrase lists have been produced by Martinez and Schmitt (2012) and Simpson-Vlach and Ellis (2010). The former is a general list of pedagogical phrases and the latter an academic one. These two phrase lists might be seen as phraseological equivalents of the GSL (Brezina & Gabslova, 2017) and the AWL (Coxhead, 2000). As such, despite the fact they have advanced the field, they nevertheless come in for the same criticisms that have been levelled against the GSL and AWL by those who favour a discipline-specific approach to vocabulary instruction. In fact, with phraseology, there seems to be even greater evidence for disciplinary variation than single vocabulary words. Hyland (2008) computed the disciplinary variation of all 3, 4 and 5 word phrases across a 3.5 million word corpus of biology, engineering, business studies and applied linguistics. He found that while some patterns are highly frequent across disciplines, e.g. *on the other hand*, fewer than 50% of the 50 most frequent phrases were shared amongst disciplines. Hyland (2008: 8) concludes "we need to be sure we are assisting learners towards an appropriate disciplinary-sensitive repertoire of bundles", and this is a goal of the current study.

4. This study

The present study fills important gaps both in theory and practice. The above review has demonstrated the need for resources to develop disciplinary literacy in students, and that the tools and methods of English for Specific Purposes in wordlist building and lexical profiling can be leveraged to do so in the secondary educational context. Recent years have also seen a movement from extending wordlists to phrase lists, as well as toward discipline-specific wordlists in the context of the debate over the utility of a general academic language. With few exceptions (Durrand, 2009), the movement toward phrase lists has not yet become discipline-specific, and those that have been developed thus far are largely lists of discourse functions made up of a mixture of lexical and grammatical words. Between the discipline-specific wordlists that provide core concept-bearing lexical items and phrase lists that provide the core discourse functions, exist lexical word phrases that carry core conceptual knowledge in a discipline. These include the very important language that captures *cell division*, *position vectors*, *homologous chromosomes* and *gamma rays*. In order to access this language, the researchers have developed a set of discipline-specific lexical bigram

phrase lists for no less than eight of the core academic disciplines at secondary school level. Furthermore, the research also examines the variation across disciplines, making a contribution to the theoretical discussion around whether or not there is a core general academic vocabulary.

The research sought to answer two questions, one pedagogical and the other theoretical:

1. What are the pedagogically useful two-word phrases students need to access to conceptual knowledge in eight core subjects: Biology, Chemistry, Economics, English, Geography, History, Mathematics, and Physics?

2. Using these pedagogical lists as discipline-specific lexical profiles, how much variation is there across disciplines? Do the results show significant overlap and support the notion of a general academic vocabulary, or is there significant variation suggesting the need for the disciplinary literacy approach?

5. Methods

5.1. Corpus profile

The corpus used for this study was compiled from an approximately 16.25 million word corpus of secondary school textbooks. Since the goal was to advance disciplinary literacy at the secondary school level, we followed the advice of Nation (2016) and designed a corpus to represent reading in the intended educational context (i.e. secondary as opposed to tertiary textbooks). As far as we are aware, this is the largest specialised corpus built from secondary school material. The corpus consisted of 206 textbooks representing eight subjects. Over 80% of the texts were published within the past five years and are on recommended reading lists for upper secondary, O-levels or A-levels. The eight disciplines covered are Biology, Chemistry, Economics, English, Geography, History, Mathematics, and Physics. More details on the technical aspects of the creation of this corpus can be found in (Authors, under review). Table 1 is an estimate of word count per discipline, measured by the corpus tool Wordsmith (Scott, 2016)

Table 1

Word counts per discipline in corpus

Biology	*Chemistry*	*Physics*	*Geography*
2,011,083	1,908,228	1,911,574	2,221,239
English	*Mathematics*	*Economics*	*History*
2,110,857	1,404,280	2,297,055	2,389,034
Total:	16,253,350		

5.2. Procedures

All texts were scanned with OCR technology and the corpus tagged using CLAWS version 7. Despite extensive data cleaning, the graphic design of secondary schoolbooks, which includes many text boxes, captions, labelled diagrams, thought bubbles, etc., introduced a certain level of noise into the

data. Thus, while CLAWS estimates an error rate of around 3%, we computed our own error rate on a random sample of 26,000 tags and calculated an overall error rate of 6.4%. Though it is not possible to correct all tag errors in a 16 million word corpus, it was crucial to ensure that the final pedagogical lists were accurate. As the lists are designed for pedagogical purposes, we implemented a set of selection criteria in order to remove bigrams that were errors or otherwise constituted unwanted material. The selection criteria were parsimonious with the goal of keeping as much as possible. The bigrams removed fell into the following categories:

1. Cases where the bigram was partly missing due being a hyphenated compound: *angled triangles* (for 'right-angled triangles')
2. Parts of longer multiword phrases (not discipline specific): *other hand* (for 'on the other hand')
3. Parts of proper names with more than two elements: *Nobel Peace* (for 'Nobel Peace Prize')
4. Bigrams that crossed clause boundaries and commas: *October Kennedy* ('in October, Kennedy signed'); *adenine cytosine ('*adenine, cytosine, etc.)
5. Bigrams that are purely textbook discourse (not discipline specific): *exam hint*; *following diagram*; *key terms*.
6. Bigrams that are final word of a name followed by abbreviation: *organisation Nato* (for North Atlantic Treaty Organisation NATO)

Each list was checked independently the two researchers, who then vetted each other's work and discussed problematic cases. CLAWS tags were also checked by hand for the final lists and vetted form the final lists to ensure accuracy for pedagogical use. In some cases, when a systematic tagging error occurred, e.g. CLAWS tagged *light speed* as an adjective noun combination, rather than tagging *light* as a noun, the correction was made.

A series of metrics were used to determine the pedagogical usefulness of phrases and whether they were discipline-specific. First, all combinations of the four major word classes were extracted: noun, verb, adjective, adverb. Following Simpson-Vlach and Ellis (2010) and Biber et al. (1999), a phrase had to occur at a minimum of 10 times per million words to be on the final lists; however, while they applied their metric across disciplines, we set the minimum frequency criteria within each discipline. A Mutual Information (MI) threshold was also set. Simpson-Vlach and Ellis (2012) in their phrase list argue that not only is this useful for teachers who may prefer to teach by MI ranking rather than frequency, but the MI score is a better indicator of the holistic nature of the phrase. Following Hunston (2002), any phrase in which the two words had a minimum MI below three was excluded from the list (see also Durrant, 2009). Following Lei and Liu (2016) and Gardner and Davies (2013), it was further required that to be pedagogically useful a phrase had to be of generally utility in the discipline, rather than occurring only in few texts and contexts. For this, two threshold metrics were applied to the phrases. The first was the Oakes dispersion test, which computes the distribution of an item across a corpus on a scale of 0-1. An item that occurs frequently and with a high MI may nevertheless only occur in a cluster of passages in a corpus. All phrases therefore had to meet a minimum dispersion metric of 0.5, following Lei and Liu (2016). Secondly, a range criterion to ensure the phrase occurred in a minimum number of texts for the target discipline. The range chosen was that the item had to occur in at least 20% of all corpus files at the minimum frequency of 10 occurrences per million words.

The order of the procedures was as follows. Firstly, all bigrams from each discipline were computed, using Wordsmith 7 (Scott, 2016). From these lists, all bigrams that were combinations of the four major part of speech tags were extracted and each bigram was checked for dispersion and range. Then, using the tool Collocate (Barlow, 2004), mutual information scores were computed for all remaining phrases and finally, the accuracy of all part of speech tags in the final lists were checked by hand independently by the researchers.

6. Results and discussion

The full pedagogical resources are available in the supplementary materials of this journal. The lists, for ease of use, are organised by the type of part of speech combination and frequency, rather than presented in one long list This allows a teacher to quickly see, for example, all noun-noun combinations without having to isolate these part of speech combinations themselves. Not all possible two-word combinations of the four lexical parts of speech were included in the final lists as some combinations, e.g. adverb+adverb, had very few terms left after all statistical criteria were applied and those that remained did not carry disciplinary content. The combinations on the final lists are: noun+noun, adjective+noun, noun+verb, and verb+noun. It was decided to also include lists of verb+adverb and verb+preposition combinations in order to present discipline-specific phrasal verbs. Even these phrases move somewhat away from the focus on concept bearing phraseology, it was felt that could be a useful pedagogical reference. Each phrase in the lists also has its associated MI score reported, so the lists can also be sorted by MI, following the recommendation of Simpson-Vlach and Ellis (2010) that this might be preferred for some classroom contexts. Further, as the lists are in Excel format, they can be easily sorted alphabetically in order to bring together all the discipline-specific words that go with a certain noun, verb, or adjective.

Table 2 presents the ten most frequent noun+noun combinations in each of the disciplines, with the STEM subjects on the top row and the humanities subjects on the lower.

Table 2

Most frequent noun + noun phrases

	Biology	**Mathematics**	**Chemistry**	**Physics**
1	carbon dioxide	surface area	carbon dioxide	field strength
2	amino acids	position vector	sodium hydroxide	field lines
3	water potential	Pythagoras' theorem	sodium chloride	heat capacity
4	blood cells	chain rule	carbon atoms	air resistance
5	surface area	arithmetic progression	room temperature	sound wave
6	blood glucose	scale factor	enthalpy change	sound waves
7	amino acid	line segment	carbon monoxide	flux density
8	cell wall	Venn diagram	oxidation state	power supply
9	cell surface	position vectors	sulfur dioxide	bar magnet
10	blood group	vector equation	carbon atom	gamma rays
	Economics	**Geography**	**History**	**English**
1	demand curve	climate change	Security Council	topic sentence
2	interest rate	population growth	East Germany	summary question
3	supply curve	sea level	Warsaw Pact	sentence structures
4	price level	land use	Weimar government	subject matter

5	interest rates	carbon dioxide	Marshall Plan	climate change
6	money supply	food production	US President	body language
7	balance sheet	life expectancy	Pearl Harbor	punctuation marks
8	exchange rate	birth rate	air force	flow chart
9	health care	fossil fuels	arms race	family members
10	trade receivables	case study	Berlin Wall	essay questions

It is clear from Table 2 that the phrases which have emerged from the procedures are clearly key lexis of the target disciplines. These are not just frequent phrasal patterns but language units carrying core concepts of the disciplines. A list not controlled to lexical part of speech phrases would not highlight the centrality of concepts such as *carbon dioxide* in Biology, *position vectors* in Mathematics, *sodium hydroxide* in Chemistry, and *gamma rays* in Physics.

Many of these terms will be found in dictionaries since they represent compound nouns. Dictionaries are useful resources both for teachers and students, however, the wordlists of the present research capture a lot lexis than is to be found in dictionaries. Dictionaries only record compounds that (a) act as lexemes, that is, have referents that are conceived of as a single unit, idea, physical object, etc., and that (b) are not easily understood through simple analysis of the two parts of the compound (e.g. *cupboard door* which is easily analysed as 'a door of a cupboard'). Numerous compound nouns of category (b) are perfectly valid lexemes, but are not considered dictionary material since their definition (e.g. 'a door of a cupboard') is no more elucidatory than the original compound itself ('cupboard door'). Lexicographers often refer to these as 'transparent compounds' and omit them because the number "of possible compounds is practically infinite" and requirements of space and time mean that "no dictionary can include more than a limited number of entries" (Lutstorf 1960: 29).

At the same time, dictionaries also define transparent compounds especially when they have some saliency as a cultural item or object. For example, most dictionaries have a definition for *apple pie*, but not *rhubarb pie* or *pumpkin pie*. In our lists, bigrams of the noun+noun (and many adjective+noun) form capture compound nouns of both types. To assess lexicographical treatment, a check of the terms in Table 2 was made in four comprehensive up-to-date dictionaries (the *Macquarie Dictionary*, *Oxford* online, *Webster's Third*, and *Wiktionary*). Each dictionary recorded a different selection of the terms, but individual dictionary coverage ranged between approximately 60% to 76% of the phrases listed in Table 2. Some of the terms were only in one dictionary (e.g. *land use* and *population growth* were only defined in the *Macquarie*), while a number of terms were missing from all four dictionaries, for example, *cell surface*, *enthalpy change*, *food production*, *scale factor*, *sentence structure*, *vector equation*, and *US president*. Illustrative of the utility of the lists is that all four dictionaries defined *cell wall* but not *cell surface*. Compositionally there is no great distinction between *cell wall* and *cell surface*, both are relatively transparent compounds, but both parts of the cell are of high significance in Biology due to their central role in cell functioning. Thus, although omitted from dictionaries, *cell surface* is a concept-bearing phrase that acts as a unit within the discipline and is therefore a term that students need to learn in order to improve their disciplinary literacy. From a psycholinguistic standpoint, in a model of lexical access such phrases may be lexicalized as two words but at the conceptual stratum they are represented is a single semantic complex (Levelt et al. 1999).

There are some interesting similarities and differences amongst the disciplines. For example, in the STEM subjects of Biology and Chemistry, *carbon dioxide* is a fundamental concept to both, being at the top of the list, yet also the phrase is important in the discourse of Geography. In the context of Geography, *carbon dioxide* will be discussed less in terms of its biological or chemical importance and instead discussed with regards to its importance to *climate change*, the term which tops the Geography list. Of note is that *climate change* is also an important topic in English in secondary school discourse. Table 2 indicates that in the STEM subjects, the phrases are closer to technical language (Hyland, 2008, 2017), while this is less true of subjects such as History and English. In History, the noun-noun combination are less about technical concepts and rather provides a list of key actors, events and places in modern history, as well a number of military related terms.

The most frequent adjective and noun combinations across the eight disciplines are reported in Table 3.

Table 3

Most frequent adjective + noun phrases

	Biology	**Mathematics**	**Chemistry**	**Physics**
1	active transport	straight line	hydrochloric acid	magnetic field
2	nervous system	standard deviation	sulfuric acid	kinetic energy
3	active site	significant figures	periodic Table	potential difference
4	small intestine	shaded region	boiling point	electric field
5	spinal cord	decimal places	ethanoic acid	potential energy
6	homologous chromosomes	quadratic equation	boiling points	thermal energy
7	fatty acids	stationary points	molecular formula	resultant force
8	aerobic respiration	stationary point	molecular mass	gravitational field
9	natural selection	minimum point	nitric acid	time graph
10	living organisms	simultaneous equations	aqueous solution	electrical energy
	Economics	**Geography**	**History**	**English**
1	United States	urban areas	Soviet Union	English language
2	aggregate demand	rural areas	United States	young people
3	economic growth	economic development	Cold War	other people
4	total cost	tropical storms	eastern Europe	different types
5	real gdp	local people	communist party	key words
6	aggregate supply	developed countries	foreign policy	human beings
7	total revenue	other countries	Korean War	short story
8	marginal cost	economic growth	civil war	standard English
9	monetary policy	tropical rainforest	Southeast Asia	main idea
10	short run	different types	Prime Minister	different ways

The adjective and noun combinations in Table 3, as with the noun and noun combinations of Table 2, also provide lexis that clearly profiles the disciplines as well as core concepts to the fields. For example, the methods have highlighted key concept bearing phrases such as *natural selection* and the *homologous chromosomes* in Biology, the *standard deviation* and *quadratic equation* in Mathematics, the *periodic Table* of Chemistry and *kinetic energy* in Physics. Economics also presents some discipline-specific terminology, such as *aggregate demand* and *marginal cost*. Economics and Geography appear to share an interest in *economic growth*, again suggesting overlap in both

language and concepts of the disciplines. How much overlap and whether this constitutes a general academic literacy is an issue that is returned to shortly.

There is little doubt that to be disciplinarily literate in these subjects, a student needs to master the phrases discussed above, the bulk of which are not available in previous pedagogical word and phrase lists. Moving on from noun phrases, Table 4 contains the most frequent noun+verb combinations. These lexical patterns capture many common subject verb combinations in the disciplines, but also help profile some of the typical processes associated with nouns.

Table 4

Most frequent noun + verb combinations

	Biology	**Mathematics**	**Chemistry**	**Physics**
1	water moves	diagram shows	gas produced	diagram shows
2	blood flows	time taken	reaction occurs	time taken
3	Table shows	distance travelled	solution containing	force acting
4	muscles relax	point lies	time taken	distance travelled
5	water enters	region bounded	energy released	forces acting
6	cells contain	line joining	reaction takes	force exerted
7	time taken	line passing	diagram shows	graph shows
8	process called	line intersects	product formed	energy required
9	blood clotting	area bounded	energy required	pressure exerted
10	cells become	problems involving	acid reacts	distance moved
	Economics	**Geography**	**History**	**English**
1	quantity supplied	people living	war ended	writer uses
2	curve shifts	people live	cartoon published	language used
3	services produced	factors affecting	Stalin wanted	words given
4	economists believe	data collected	people lived	writer says
5	income earned	map showing	USA wanted	writer describes
6	prices rise	plates move	Hitler wanted	writers use
7	quantity sold	factors affect	people died	information given
8	curve shows	people died	source suggest	people living
9	government spending	graph showing	speech made	words explain
10	people want	damage caused	people thought	techniques used

Table 4 indicates some interesting discipline-specific collocations. For example, in Biology *blood flows* but *water moves.* In Mathematics, *regions* and *areas* are *bounded*. In Chemistry *energy* is often discussed in terms of processes such as *energy released* during a chemical reaction whereas in Physics the discussion is more frequently of the *energy required* for an action. A different lexical profile emerges for the humanities. In History, the nature of knowledge in this discipline is often to try to account for historical events based on the motivations of important historical players and this is reflected in the phrases in Table 4, e.g. *Stalin wanted, the USA wanted*. Reversing the pattern, Table 5 lists the ten most frequent verb+noun patterns.

Table 5

Most frequent verb + noun phrases

	Biology	Mathematics	Chemistry	Physics
1	takes place	shows part	takes place	takes place
2	take place	comparing coefficients	conduct electricity	take place
3	taking place	solve problems	take place	transfer energy
4	shows part	using Pythagoras	form ions	make use
5	give rise	using substitution	lose electrons	shows part
6	provide energy	takes hours	conducts electricity	gives rise
7	release energy	solving equations	taking place	taking moments
8	carries blood	involving modulus	gain electrons	makes use
9	produce antibodies	solving problems	write equations	generate electricity
10	carry blood	make use	produce hydrogen	give rise
	Economics	**Geography**	**History**	**English**
1	buy goods	takes place	took place	make use
2	take place	taking place	declared war	take note
3	took place	took place	take place	make sense
4	increase output	taken place	invade Cuba	use language
5	take advantage	using examples	took control	pay attention
6	buy shares	generate electricity	use force	using language
7	make decisions	increase food	seized power	give reasons
8	allocate resources	collect data	pay reparations	take place
9	producing goods	make way	seize power	taking place
10	takes place	allow water	became president	paying attention

Table 5 indicates that several high frequency verb+noun phrases are used in multiple subjects, unlike the concept-heavy noun+noun and adjective+noun combinations previously discussed. For example, the phrase *take place* and its inflectional variants is amongst the ten most frequent verb+noun combinations in Biology, Chemistry, Physics, Economics, Geography, History and English. This raises the question as to whether these discipline-specific lists might contain a core academic vocabulary of general utility. Table 5 nevertheless indicates disciplinary variation. For example, discipline-specific processes are highlighted such as *provide energy* in Biology, *comparing coefficients* in Mathematics, *conduct electricity* in Chemistry, and *transfer energy* in Physics. The phrases of English suggest a different focus in the discipline related to the knowledge construction of the field (Martin, 2013), such as *give reasons*, *use language* and *pay attention*, language related to argumentation and analysis.

Finally, as noted, all verb+adverb combinations were computed, the most frequent of which are reported in Table 6.

Table 6

Most frequent verb + adverb phrases

	Biology	Mathematics	Chemistry	Physics
1	made up	write down	explain why	set up
2	explain why	carried out	carried out	shows how
3	explain how	selected at	made up	switched on
4	carry out	chosen at	explain how	occurs when
5	carried out	made up	write down	write down
6	broken down	multiply both	carry out	happens when
7	describe how	divide both	find out	made up
8	make up	know how	given out	find out
9	break down	occurs when	held together	slow down
10	crossing over	drawn at	used up	slows down
	Economics	**Geography**	**History**	**English**

1	explain how	explain how	set up	explain why
2	set up	explain why	explain why	find out
3	occurs when	find out	explain how	explain how
4	find out	describe how	carried out	write down
5	become more	occurs when	made up	show how
6	see how	set up	became more	see how
7	give up	carried out	broke out	know how
8	shows how	become more	set out	become more
9	end up	carry out	carry out	go back
10	show how	suggest why	went on	set up

Table 6 captures a range of patterns, including phrasal verbs constituted by verb + adverbial particle combinations, as well as more free combinations. Note that these reflect discipline-specific language patterns, e.g. matter is *broken down* in Biology, compounds are *held together* in Chemistry. Phrasal verbs also vary by discipline, things *break out* in History, and experiments are *set up* in Physics. While recent corpus-based research has developed pedagogical phrasal verb lists (Garnier & Schmitt, 2015), these have not included disciplinary variation. Nevertheless, Table 6 also indicates a significant amount of shared language amongst the most frequent phrases, which brings us to the important question of a general academic vocabulary.

6.1. Is there a general academic vocabulary across the phrase lists?

We turn now to second of our research questions. The above discussion has indicated some overlap between the high frequency part of speech combinations on several of the lists, so the question can be asked as to whether there is a core academic vocabulary or whether disciplinary variation is so great as to support the disciplinary literacy approach that focus on discipline-specific phrases. While Simpson-Vlach and Ellis (2010) were able to derive a list of phrases they suggested constituted a core academic language, other researchers such as Hyland (2008) found disciplinary variation to the extent that there was less than 50% overlap amongst phrases in different disciplines. Table 7 reports overlap in the number of phrases between the disciplines in terms of the number of shared items in a pairwise matrix.

Table 7

A pairwise matrix of number of shared phrases across the eight disciplines

		Biology	*Math*	*Chemistry*	*Physics*	*English*	*History*	*Geography*
1	*Biology*	-						
2	*Mathematics*	34	-					
3	*Chemistry*	242	40	-				
4	*Physics*	150	87	207	-			
5	*English*	62	22	48	38	-		
6	*History*	53	16	31	25	92	-	
7	*Geography*	170	36	109	106	109	165	-
8	*Economics*	82	33	59	43	91	105	217

Table 7 shows that in general, the STEM subjects had more shared phrases with each other, and less in common with humanities subjects, while the opposite was true for the humanities subjects. This reflects not just language similarity but reflects that the conceptual domains in these disciplines are closer to each other. Geography, interestingly, shares many concepts with both the STEM subjects and Humanities, which indicates it is a discipline that is both a hard science (e.g. geology) and a social science (e.g. social geography). Overall, however, while Table 7 indicates 2472 occurrences of phrases occurring in more than one list, this is in fact a very low proportion of overlap across all eight subjects. In total, there were 7682 phrases in the pedagogical lists across all disciplines, and in this entire resource, only 8 phrases occurred in all 8 disciplines. There were a further 8 that occurred in 7 disciplines, 17 that occurred in 6 disciplines, 35 that occurred in 5 and 62 that occurred in 4 of the 8 disciplines. In other words, phrases that occurred in more than 50% of the disciplines amounted to only 130 of 7682 phrases; i.e. approximately 2% of the lexical word combinations can be said to be of general utility across disciplines, despite meeting stringent criteria for discipline-specificity. Further, 6780 phrases, approximately 88%, occur in the lexical profile of only one discipline. The results provide further evidence in support of a discipline-specific approach to language education and the importance of a decline-specific phraseology (Hyland, 2008; Martin, 2013).

7. Limitations

A few limitations and caveats need to be outlined. First, the scanning and OCR of texts, in combination with the graphic design of high school text book, introduced a certain extent of noise into the corpus which in turn led to a higher tagging error rate than usual. This was controlled for, however, by vetting the final phrase lists. Second, the pedagogical lists do not represent all possible part of speech combinations, but only those deemed to have the greatest educational utility. In fact, the coverage in each disciplinary corpus for the target part of speech combinations is quite good, and comes close to that of recent discipline specific lemma list (e.g. Lei & Liu, 2016): approximately 28% of all bigram tokens in Chemistry, 25% in Physics, 16% in Biology, 15% in Geography, 14% in History and Math, yet only 5% in English- a reflection of the diverse vocabulary of his discipline. Further, while previous research informed all technical methodological decisions, it should be noted that there is nothing particularly special about, for example, a dispersion metric of 0.5 or a minimum frequency of 10 occurrences per million words. Finally, the lists are extensive and unlike previous research, such as Durrant (2009) who set a cut-off of the top 1000 bigrams, we made a conscious methodological decision to provide teachers with as much information as possible about discipline-specific lexis for the eight secondary school subjects. The lists are thus intended to be a resource that can be plumbed in order to better inform materials design and curriculum development, which as Nation (2016) points out, is an essential function of wordlists.

8. Conclusion

This paper offers several major contributions. It combines the work done in ESP and corpus linguistics with that being done in disciplinary literacy, drawing on the technical advances of the former to produce a resource that can enhance the teaching of disciplinary literacy in the secondary school context. The resulting lists fill an important gap between discipline-specific lemma lists that

represent the language of disciplinary content, and phrase lists which tend to be combinations of lexical and grammatical words that represent discourse moves and functions. The research indicates that important concepts that are fundamental to disciplinary literacy are carried by two-word lexical combinations also, which have hitherto been given little attention in research and materials development. Finally, the paper makes a theoretical contribution by adding to the growing amount of evidence challenging the notion of a core academic vocabulary. The results of this study indicate large variation in the language and concepts secondary school students need to learn in the different disciplines, and therefore suggests that the most appropriate pedagogical approach is one of disciplinary literacy rather than a decontextualized general literacy approach.

References

Airey, J., Lauridsen, K. M., Räsänen, A., Salö, L., & Schwach, V. (2017). The expansion of English-medium instruction in the Nordic countries: Can top-down university language policies encourage bottom-up disciplinary literacy goals? *Higher Education*, *73*(4), 561-576.

Balota, D. A. (2016). Speed reading: you can't always get what you want, but can you sometimes get what you need? *Psychological Science in the Public Interest*, *17*(1), 1-3.

Barlow, M., (2004) Collocate [Computer Software]. Houston: Athelstan Publications

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied linguistics*, *25*(3), 371-405.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.

Brezina, V., & Gablasova, D. (2017). How to produce vocabulary lists? Issues of definition, selection and pedagogical aims. A response to Gabriele Stein. *Applied Linguistics*, *38*(5), 764-767.

Britt, M. A., Richter, T., & Rouet, J. F. (2014). Scientific literacy: The role of goal-directed reading and evaluation in understanding scientific information. *Educational Psychologist*, *49*(2), 104-122.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238.

Coxhead, A. (2016). Reflecting on Coxhead (2000),"A New Academic Word List". *TESOL Quarterly*, *50*(1), 181-185.

Coxhead, A., & Hirsch, D. (2007). A pilot science-specific word list. *Revue Française de Linguistique Appliquée*, *12*(2), 65-78.

Dolch, E. (1936). A basic sight vocabulary. *The Elementary School Journal*, *36*(6): 456-460.

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, *28*(3), 157-169.

Fraser, B. (2015). Classroom learning environments. In R. Gunstone (Ed.), *Encyclopedia of Science Education* (pp. 154-157). Netherlands: Springer

Fry, E. (1957). Developing a word list for remedial reading. *Elementary English, 34*(7): 456-458.

Gardner, D., & Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics*, *35*(3), 305-327.

Garnier, M., & Schmitt, N. (2015). The PHaVE List: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research*, *19*(6), 645-666.

Gillis, V. (2014). Disciplinary literacy. *Journal of Adolescent and Adult Literacy*, *57*(8), 614-623.

Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M. A., Greenleaf, C. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist*, *51*(2), 219-246.

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

Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing*. Ann Arbour, MI: University of Michigan Press.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*(1), 4-21.

Hyland, K. (2017). English in the disciplines: Arguments for specificity. *ESP Today-Journal of English for Specific Purposes at Tertiary Level*, *5*(1), 5-23.

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL quarterly*, *41*(2), 235-253.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, *57*(1), 1-33.

Konstantakis, N. (2007). Creating a business word list for teaching business English. *Elia: Estudios de Lingüística Inglesa Aplicada*, 7, 79-102.

Koomen, M. H., Weaver, S., Blair, R. B., & Oberhauser, K. S. (2016). Disciplinary literacy in the science classroom: Using adaptive primary literature. *Journal of Research in Science Teaching*, *53*(6), 847-894.

Langacker, R. W. (2008). *Cognitive grammar: A basic introduction*. Oxford: Oxford University Press.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 33-51.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.

Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, *22*, 42-53.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1-38.

Lutstorf, H. T. (1960). *The stressing of compounds in modern English*: *A study in experimental phonetics*. Unpublished doctoral thesis: University of Zurich.

Macnaught, L., Maton, K., Martin, J. R., & Matruglio, E. (2013). Jointly constructing semantic waves: Implications for teacher training. *Linguistics and Education*, *24*(1), 50-63.

Macquarie Dictionary. (2017). 7th edition. http://www.macquariedictionary.com.au

Martin, J. R. (2013). Embedded literacy: Knowledge as meaning. *Linguistics and Education*, *24*(1), 23-37.

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied linguistics*, *33*(3), 299-320.

Martínez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes*, *28*(3), 183-198.

Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. Amsterdam: John Benjamins.

Nagy, W. & Townsend, D. (2012). Words as tools: Leaning academic vocabulary as language acquisition. *Reading Research Quarterly*, 47(2): 91–108.

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. *Vocabulary: Description, acquisition and pedagogy*, *14*, 6-19.

Oxford Dictionary. (2017). http://www.oxforddictionaries.com/

Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2016). Lexical profiles of comprehensible second language speech: The role of appropriateness, fluency, variation, sophistication, abstractness, and sense relations. *Studies in Second Language Acquisition*, *38*(4), 677-701.

Schmitt, D., Schmitt, N., & Mann, D. (2011). *Mastering the academic word list*. London: Longman.

Schmitt, N. (2008). Instructed second language vocabulary *learning. Language Teaching Research*, 12(3), 329-363.

Scott, M. (2016). *WordSmith Tools* [Computer software]. Liverpool.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, *78*(1), 40-59.

Shanahan, T., & Shanahan, C. (2017). Disciplinary Literacy: Just the FAQs. *Educational Leadership*, *74*(5), 18-22.

Simpson, R. C., Briggs, S. L., Ovens, J., & Swales, J. M. (2002). The Michigan corpus of academic spoken English. *Ann Arbor, MI: The Regents of the University of Michigan*.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, *31*(4), 487-512.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Slobin, D. I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. *Language in Mind: Advances in the Study of Language and Thought*, *157*(1), 178.

Smart, J. B., & Marshall, J. C. (2013). Interactions between classroom discourse, teacher questioning, and student cognitive engagement in middle school science. *Journal of Science Teacher Education*, *24*(2), 249-267.

Stein, G. (2017). Some Thoughts on the Issue of Core Vocabularies: A Response to Vaclav Brezina and Dana Gablasova:'Is There a Core General Vocabulary? Introducing the New General Service List'. *Applied Linguistics*, 38(5), 759-763).

Todd, R. W. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes*, *45*, 31-39.

Webster's Third New International Dictionary. (1993). Chicago: Merriam-Webster.

West, M. (1953). *A general service list of English words.* Longmans Green and Company.

Wiktionary. (2017). http://en.wiktionary.org/

Wilson, A., Madjar, I., & McNaughton, S. (2016). Opportunity to learn about disciplinary literacy in senior secondary English classrooms in New Zealand. *The Curriculum Journal*, *27*(2), 204-228.

Wingate, U. (2012). Using Academic Literacies and genre-based models for academic writing instruction: A 'literacy' journey. *Journal of English for Academic Purposes*, *11*(1), 26-37.

Wray, A. (2013). Formulaic language. *Language Teaching*, *46*(3), 316-334.

Yang, M. N. (2015). A nursing academic word list. *English for Specific Purposes*, *37*, 27-38.