

# Exploring polysemy in the Academic Vocabulary List:

## A lexicographic approach

Sophia Skoufaki\* and Bojana Petrić

*Department of Language and Linguistics, University of Essex, Colchester, UK*

*Department of Languages, Cultures and Applied Linguistics, Birkbeck, University of London, London, UK*

Sophia Skoufaki

Department of Language and Linguistics

University of Essex

Wivenhoe Park

Colchester

CO4 3SQ

UK

Email: [sskouf@essex.ac.uk](mailto:sskouf@essex.ac.uk)

ORCID ID: 0000-0002-9992-3977

Bojana Petrić

Department of Languages, Cultures and Applied Linguistics

Birkbeck, University of London

26 Russell Square

London

WC1B 5DQ

UK

Email: [b.petric@bbk.ac.uk](mailto:b.petric@bbk.ac.uk)

ORCID ID: 0000-0001-6855-3785

Declarations of interests: none

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

# Exploring polysemy in the Academic Vocabulary List:

## A lexicographic approach

### Abstract

The Academic Vocabulary List (AVL) (Gardner & Davies, 2014) is a valuable resource for EAP teachers and students as it identifies potential lexical learning/teaching targets. This study enhances the AVL's pedagogical usefulness by identifying polysemous lemmas in it. Polysemous AVL lemmas are operationalised as those with more than one definition in two lexicographic resources, the *Collins COBUILD Advanced Learners' Dictionary* and WordNet. This study also examines a theoretical issue, the relationship between the number of meaning senses of AVL lemmas and their frequency in an academic-English corpus. To this end, correlations were calculated between the numbers of AVL lemmas' meaning definitions listed in both lexicographic resources and their frequency in the COCA-Academic corpus. 34.38% of the 2,673 AVL lemmas included in both lexicographic resources, excluding homonyms, are polysemous. Most (66.05%) come from the most frequent 1,000 AVL lemmas. The number of meaning definitions of AVL lemmas and their frequency are positively correlated. This correlation is non-linear, i.e., low-frequency words tend to be monosemous but beyond a frequency threshold, word definitions increase as word frequency increases. Implications for future research and teaching are discussed.

**Keywords:** polysemy, vocabulary, Zipf, meaning sense, word frequency, EAP

## 1. Introduction

Lists of English general academic words, i.e., words occurring frequently in academic discourse across disciplines (e.g., Nation, 2013), such as the Academic Word List (AWL) (Coxhead, 2000) and the Academic Vocabulary List (AVL) (Gardner & Davies, 2014), are of great value to EAP course designers, materials developers and teachers since they facilitate the selection of words to be included in course materials or prioritised for direct teaching. However, the utility of wordlists is limited because, among other things, they lack supplementary information about words in them (e.g., Thompson & Alzeer, 2019). Recent research has attempted to make academic wordlists more user-friendly by providing, for example, grammatical (Green, 2019) and collocational (Lei & Liu, 2018) information about the words listed.

Another way of supplementing academic wordlists would be to indicate which words are polysemous. Information on polysemy in an academic wordlist would be useful to teachers and learners because language learners are unlikely to infer unknown meaning senses of polysemous words from context because they tend to assume that words are monosemous (e.g., Bensoussan & Laufer, 1984; Frantzen, 2003) whereas at least some English general academic words are polysemous (e.g., Granger & Paquot, 2009; Rhees, 2018). Examples include academic words with different meanings when used in general academic English and discipline-specific contexts, such as *solution*, denoting ‘solution to a problem’ in general academic English but also used as a discipline-specific term in chemistry (e.g., ‘solution of ammonia in water’) (Mudraya, 2006), and academic words whose meanings vary across discipline groups, such as *analyse*, which refers to ‘methods of determining the composition of a substance’ in science as opposed to the meaning of ‘considering something carefully’ in the social sciences (Hyland & Tse, 2007, p.244). Although, as shown above, previous studies

indicate that some academic words are polysemous, while academic wordlists have not been examined for polysemy. Information on how many and which English academic words are polysemous can have implications for academic vocabulary teaching and thus enhance the pedagogical utility of academic wordlists for EAP teaching. With this aim, this study will examine polysemy in the AVL by identifying which AVL lemmas are polysemous and how many meaning senses they have.

This study also addresses a theoretical research gap. Zipf's (1945, 1949) claim that the more frequent a word is, the more meaning senses it tends to have has received support in relation to general vocabulary both in Zipf's aforementioned publications and others (e.g., Casas, Hernandez-Fernandez, Catala, Ferrer-i-Cancho, & Baixeries, 2019; Kuiper, Fromont, & Gerhard, 2017). However, research on the relationship between academic word frequency and meaning-sense number has not been conducted. Establishing the relationship between academic word frequency and meaning-sense number is of both theoretical and practical interest. Theoretically, it is of interest to determine whether academic words behave similarly to general words in terms of their polysemy-frequency relationship. Practically, the findings can be of benefit to EAP pedagogy by providing useful information about the profile of frequent academic words.

To summarise, this study aims to identify polysemous words in general English academic vocabulary operationalised as the AVL and examine the relationship between the frequency of academic words in an academic corpus and the number of their meaning senses. Our method relies on lexicographic and corpus data. In section 2 we discuss the notion of polysemy and dictionary treatment of polysemy, which is followed by an overview of previous research on the relationship between polysemy and word frequency, and on polysemous academic words. The study's research questions are presented in section 3.

Section 4 describes the method employed in the study. Section 5 presents results. Section 6 discusses results and their implications for research and pedagogy.

## **2. Literature review**

### *2.1 Defining polysemy*

Polysemy is commonly defined as a type of meaning variation where a lexical item has two or more distinct meaning senses (e.g., Murphy, 2010). Unlike homonymy, where words sharing the same phonological or written form have unrelated meanings (e.g., *bank*<sub>1</sub>, which means ‘a financial institution’ and *bank*<sub>2</sub>, which means ‘the land alongside a river’), in polysemy the different senses of a word are related to each other: for example, *bank*<sub>1</sub> also has the meaning sense ‘a stock of something available for use when required, such as a blood bank’ (*Oxford Dictionary of English*).

In the traditional conceptualisation of polysemy, the meaning senses of a polysemous word are considered discrete and well-defined linguistic units, which extend from its core meaning. However, research in corpus semantics and psycholinguistics in the last three decades has revealed that the boundaries between meaning senses of polysemous words are fuzzy rather than strict (Storjohann, 2016), leading to a shift to a more dynamic conceptualisation of polysemy based on the idea of words having ‘meaning potentials’ (Hanks, 2013; Geeraerts, 2016). To understand the meaning of a word in a particular context, it is necessary to investigate the phraseology of which it is part (Hanks, 2013).

Finally, it is not always easy to decide whether a word is polysemous or two or more homonyms (e.g., Cruse, 2004; Murphy, 2010), how many meaning senses a lexical item has

(e.g., Pragglejazz Group, 2007) and which senses need to be taught to foreign language learners (e.g., Garnier & Schmitt, 2015). These practical problems have led researchers to operationalise polysemous words as those that have more than one definition in a learner dictionary (e.g., Dobrić, 2015; Pragglejazz Group 2007), which is the approach we follow in this study. The next section discusses polysemy treatment in dictionaries.

## *2.2 Polysemy in dictionaries*

To infer word meaning, lexicographers examine word usage in large corpora. This process begins with studying many concordance lines with the key word in context (KWIC) and identifying usage patterns. Lexicographers also obtain summaries of a word's grammatical and collocational behaviour from corpus query tools (see, e.g., Kilgarriff, 2013). Word occurrences are then grouped according to meaning similarity and/or collocation and other usage patterns. Once distinct meaning senses are established, their definitions are developed and examples written or selected from the corpus (Alexander, 2015).

Dictionaries vary in the extent to which they differentiate among the senses of polysemous words. A distinction is commonly made between 'splitters', i.e., lexicographers tending to make very fine-grained word sense distinctions, and 'lumpers', i.e., those leaning towards grouping broadly similar senses together (Walter, 2010; Lew, 2013). A dictionary's approach to sense distinction also depends on its intended users. A learner dictionary aimed at beginners, for instance, will tend to adopt a 'lumping' approach, while a dictionary for linguists, such as the *Oxford English Dictionary*, is likely to fall on the 'splitter' side of the continuum.

Dictionaries typically represent polysemy as a numbered list of meaning senses within an entry for a polysemous word (however, see also, e.g., Nesi & Tan, 2011, for the use of signposts and menus to help users navigate longer entries easily). The senses of a polysemous word are listed in the order of frequency; however, this is mostly indicative only, because identifying frequencies of the different senses of a polysemous word would require an enormous amount of laborious sense coding since sense identification cannot be done automatically. Frequency counts of individual senses are affected by whether the dictionary adopts a ‘splitter’ or a ‘lumper’ approach to sense identification (Walter, 2010). Dictionaries also differ in where they draw the line between the polysemy (where meaning sense definitions are listed in a single dictionary entry) and homonymy (where meaning sense definitions are presented in separate entries) of some words, as comparative analyses of commonly used English learners’ dictionaries have shown (e.g., Moerdijk, 2003).

Relevant to this discussion is the emergence of online lexicographic resources and databases based on advances in computational lexicography, such as Wordnik<sup>1</sup> and WordNet (Fellbaum, 1998). Wordnik is a dictionary and a language resource which incorporates existing dictionaries and automatically sources examples illustrating a word’s meaning senses from the internet, offering users large numbers of examples of target words in context without any editorial interventions. WordNet is a large database of words structured in networks based on their semantic relations, used for research in computational linguistics and natural language processing (see Method for more information).

Notwithstanding the challenges in word sense distinction described above, dictionaries are an excellent tool for examining polysemy from both a theoretical and an applied perspective. Corpus-based dictionaries in particular, such as most English learner dictionaries today (Yamada, 2013) (e.g., *MacMillan English Dictionary for Advanced Learners*, *Collins*

---

<sup>1</sup> Wordnik is freely available at <https://www.wordnik.com/>

*COBUILD Advanced Learners' Dictionary*), are considered good substitutes for the laborious process of intuitive sense tagging of corpora by linguists (Deignan, 2015). Moreover, consulting a corpus-based dictionary helps to limit the subjectivity of a researcher's sense tagging (Dorst & Reijnierse, 2015).

### *2.3 Word frequency and polysemy*

A positive relationship between a word's frequency and its meaning senses was first formally posited by Zipf (1945, 1949). In addition to his own research, this relationship has been examined through various operationalisations of 'word frequency' and 'meaning senses' of English words and words in other languages (e.g., Casas et al., 2019; Kuiper et al., 2017). Although the exact statistical findings differ among studies, they all support Zipf's claim. In applied linguistics, the validity of Zipf's exact formula about the relationship between word frequency and word meaning senses has not been examined per se but Reynolds, Wu, Kuo, and Yeh (2015) found a strong positive relationship between word frequency in the British National Corpus and the number of word definitions in WordNet ( $r = 0.596, p < .0001$ ).

This issue has implications for vocabulary learning because being aware that high-frequency words tend to have more meaning senses than low-frequency words can inform language teaching. For example, according to Kuiper et al. (2017), a positive relationship between word frequency and the number of word's meaning senses means that language teaching materials for beginner learners should not only include high-frequency English words but also present them in contexts illustrating their different meaning senses.



The studies mentioned above have been conducted on general vocabulary. Examining the relationship between frequency and polysemy in English academic vocabulary can inform the learning and teaching of English academic vocabulary.

#### 2.4 *English academic vocabulary*

Although *general academic vocabulary* is broadly defined as the vocabulary used in academic writing and speech across disciplines (e.g., Nation, 2013), which words one considers academic depends on which academic wordlist one uses. This section will explain why English academic vocabulary was operationalised as the AVL (Gardner & Davies, 2014) in the present study.

The AVL consists of 3,014 word lemmas<sup>2</sup> (i.e., root word forms each with a specific part of speech [POS] and its inflected forms) which occur at least 50% more frequently in the Academic section of the Corpus of Contemporary American English (COCA) than would normally be expected. A lemma wordlist was preferred over a word-family wordlist<sup>3</sup>, such as the AWL (Coxhead, 2000), because the lack of POS tagging in word-family wordlists means that words with the same spelling are counted together although their frequency and meaning may differ depending on POS. Another reason why we preferred the AVL over the AWL is

---

<sup>2</sup> The Excel file with the AVL provided as supplementary material in Gardner and Davies (2014) consists of 3,105 lemmas but the entry for the word *disproportionately* appears twice (Durrant, 2016, p. 53). Therefore, the real number of lemmas in the AVL is 3,014. In our study, we worked with 3,013 AVL lemmas. We excluded the two *disproportionately* AVL entries from data analysis because these two entries provide different information about the frequency of *disproportionately* in COCA-Academic and, consequently, its rank in the AVL list.

<sup>3</sup> In a word-family wordlist each headword stands for itself and all its inflected and derived forms up to a certain level of Bauer and Nation's (1993) English affixes list, a list of affixes ordered in levels according to their frequency in written English.

because in the former academic words are evenly distributed across disciplinary sections of the Academic section of COCA whereas the AWL has been found to be more representative of the social sciences than of the hard sciences (e.g., Durrant, 2014).

Finally, we operationalise English academic vocabulary as the AVL because it includes both high-frequency and low-frequency words. In this way, it differs from the AWL, which assumes that academic vocabulary is different from high-frequency vocabulary. This assumption has been criticised because it means that many high-frequency words used both in everyday life and academic settings can be labelled ‘general’ or ‘academic’ depending on which words are considered high-frequency (e.g., Cobb, 2010).

### *2.5 Polysemy in English academic vocabulary*

Research has not examined how many English academic words are polysemous. Nevertheless, various studies indicate that at least some English academic words are polysemous.

Studies indicate considerable word overlap between the AWL and wordlists other than the General Service List (West, 1953), such as wordlists derived from the British National Corpus (Cobb, 2010; Masrai & Milton, 2018) and the BNC/COCA word family lists (Masrai & Milton, 2018). For example, Cobb (2010, p.192) found that a “total of 280 AWL items falls within the first two 1,000 levels of the BNC”; this overlap accounts for nearly half (49.12%) of the 570 AWL word family headwords. Although some of these AWL and BNC overlapping words may be homographs, this lexical overlap also indicates that some high-frequency English academic words are polysemous.

Corpus studies on various disciplines indicate that specific academic words have more than one meaning (e.g., Granger & Paquot, 2009; Martínez, Beck, & Panza, 2009; Mudraya, 2006; Partington, 1998). Within an academic wordlist, meaning variation has only been investigated with reference to the AWL; however, researchers tended to focus on homonymy rather than polysemy. For instance, Wang and Nation (2004) examined how many AWL word families contain homographs by classifying dictionary definitions of AWL items on a semantic relatedness scale, thus using intuition-based judgements to distinguish between polysemy and homography. They found that 60 out of the 570 AWL items contained homographs. Hyland and Tse (2007) investigated the cross-disciplinary variation in meaning and use of AWL headwords in a corpus of selected hard science, engineering and social science disciplines. Among others, they provide further evidence of cross-disciplinary semantic variation of AWL items due to the existence of homographs, based on the analysis of concordance lines of selected AWL headwords with potential homographs. Although they point to cross-disciplinary differences in the meaning senses of several individual words, such as 'analysis' (as mentioned in the Introduction), which indicate polysemy, polysemy in the AWL is not explored in their study.

Rhees (2018) is another study of English academic vocabulary use across disciplines. In particular, this study conducted Corpus Pattern Analysis (CPA) on the concordance lines which included 30 AWL verb headwords in a corpus made out of research articles in History, Microbiology and Management. CPA is a corpus annotation approach whereby the collocates of verbs are annotated in terms of the broad semantic set they belong to (e.g., if the verb's subject is a person, the tag 'HUMAN' is used to annotate it). These annotations map onto different verb meaning senses and the aim of such annotations is to map meaning on text

(Hanks, 2004) because different collocation patterns map onto different meaning senses<sup>4</sup>. By analysing only AWL headwords which are unambiguously verbs, this study examined, in effect, verb lemmas and, consequently, is an examination of polysemy as it is commonly defined. Another methodological strength of Rhees (2018) is that in addition to comparing patterns and meaning senses across the three disciplinary corpora, it examines pattern and meaning sense variation inside each subcorpus. Results indicated that with the exception of *accomplish*, all verbs appeared with more than one pattern, an indication that nearly all verbs are polysemous. Moreover, some of these patterns appeared significantly more in one discipline than in another and others were shared, an indication that some verb meanings are discipline-specific and others are not. Some of these patterns also appear in *The pattern dictionary of English verbs* (Hanks, 2001), an indication that they express meaning senses which are shared between General English and Academic English.

By combining quantitative and qualitative analysis with CPA, this study offers an in-depth view into the polysemy of academic vocabulary. However, this study is limited to a sample of AWL verbs and, therefore, like the other studies reviewed in this section, it does not give us an indication of polysemy in a whole academic wordlist.

### **3. The present study**

The present study is motivated by the research gaps discussed in sections 2.3 and 2.5, namely, the lack of research on the relationship between English academic word frequency

---

<sup>4</sup> The patterns and associated meanings of English verbs in general English appear in *The pattern dictionary of English verbs* (Hanks, 2001), available at <https://www.pdev.org.uk/>.

and polysemy and the lack of a large-scale identification of polysemous English academic words, respectively.

Regarding the first research gap, by examining whether English academic words tend to have more meanings the more frequent they are, this study addresses a theoretical issue that has implications for English academic vocabulary teaching and learning (see section 2.3). In relation to the second research gap, the present study identifies polysemous academic vocabulary in the AVL. This research aim is pedagogically worthwhile because it will provide EAP and ELT materials designers and practitioners information on which AVL words have more than one meaning sense, thus helping them to identify AVL words with meaning senses their students need to learn. In this way, this study contributes to recent research (e.g., Green, 2019; Lei & Liu, 2018) which supplements English academic wordlists with information teachers and language learners may find useful.

This study addresses these research questions:

1. What is the relationship between AVL lemma frequency in a corpus of academic English and AVL lemma number of meaning senses?
2. Which AVL lemmas are polysemous?

## **4. Method**

### *4.1 Selection of lexicographic resources*

Since dictionaries differ in their treatment of polysemy (see section 2.2), to avoid the shortcomings of identifying polysemous AVL lemmas based on a single dictionary, this study uses two different lexicographic resources, the *Collins COBUILD Advanced English Dictionary (COBUILD)* and WordNet (Fellbaum, 1998).

The *Collins COBUILD Advanced English dictionary* is a dictionary with a long history in corpus-based research; its current version is based on the Collins Corpus, which contains 4.5 million words. Aimed at advanced learners of English, such as those typically enrolling in pre-sessional EAP courses, it provides explicit explanations of sense distinctions in a language easily accessible to learners, using only the 3,000 most frequent words for definitions and examples. Conversely, WordNet is a large online lexical database created within a research project at Princeton University. Focusing on words' lexical and semantic relationships, it is primarily aimed at researchers, and has been extensively used as a resource in studies of polysemy (e.g., Casas et al., 2019, Kuiper et al., 2017).

Although both are monolingual, they differ in terms of their intended users and purposes: *COBUILD* is an English learner dictionary, while WordNet is a lexicographic database aimed at researchers in computational linguistics and natural language processing. Based on this distinction, *COBUILD* is likely to follow a 'lumping' approach, focusing on sense distinctions of relevance to language learners, while WordNet is likely to be a 'splitter' when it comes to sense distinctions of polysemous words (see section 2.2). Consequently, *COBUILD* can be expected to underestimate polysemy, while WordNet is likely to overestimate it. Using these two lexicographic resources in the study is therefore likely to promote the validity of our AVL polysemous word identification thanks to triangulation (Cohen, Manion, & Morrison, 2018).

We use general English lexicographic resources rather than those limited to academic usage, such as the *Oxford Learner's Dictionary of Academic English*, because students on EAP courses will have encountered AVL lemmas in academic and non-academic discourse in their previous instruction. Research has shown that students tend to assume that words are monosemous and that the meaning they have first encountered applies to other usages of the

word (Bensoussan & Laufer, 1984; Frantzen, 2003). We therefore believe it is useful for practitioners to take into account the full spectrum of meanings of AVL lemmas.

#### *4.2 Polysemous AVL lemma identification*

Polysemous lemmas in the AVL were operationalised as those with more than one definition in both *COBUILD* and WordNet.

The online *COBUILD* was searched via its Application Programming Interface (API) for the 3,013 AVL lemmas (i.e., all AVL lemmas except the two *disproportionately* entries, see footnote 2) using a script written in the programming language Python (van Rossum, 1995). The results of this search were output in an Excel file which contained the lemma and POS columns of the AVL Excel file and an extra column entry for each definition found for every AVL lemma identified in the online dictionary.

202 of the 3,013 AVL lemmas were not identified in the *COBUILD* API. A subsequent manual search of *COBUILD* for these lemmas did not result in the identification of any of them.

It should be pointed out that the AVL contains separate lemma entries for words which are spelled differently in American and in British English (e.g., British English *fulfil* and American English *fulfill*); due to the existence of these separate lemma entries in the AVL, we did not consider lemmas with different spellings as interchangeable and, consequently, API-undetected lemmas of this sort (namely, *fulfil*, *judgement*, *underly*) were not lumped together with their American English equivalents.

The definitions extracted from the *COBUILD* API for 100 randomly selected words were compared against those in the online Collins *COBUILD* dictionary (<https://www.collinsdictionary.com/>) to test the consistency of the API and online dictionary

definitions. This comparison showed that the definitions extracted from the *COBUILD* API were consistent with those in the online *COBUILD* dictionary for all but the adjective *low*. For *low*, the API search yielded 15 whereas the online *COBUILD* consultation yielded 17 definitions; the definitions ‘If you drive or ride a bicycle in a low gear, you use a gear, usually first or second, which gives you the most control over your car or bicycle when travelling slowly.’ and ‘If you describe someone such as a student or a worker as a low achiever, you mean that they are not very good at their work, and do not achieve or produce as much as others.’ appeared in the online *COBUILD* dictionary but not in the API. This discrepancy is probably due to changes made in online *COBUILD* entries between the time when the *COBUILD* API was searched (2017) and the time when we checked the aforementioned 100 AVL lemmas in online *COBUILD* (2021). Such discrepancies are not unexpected since online dictionaries are continuously updated. As we considered the single discrepancy identified in the sample acceptable, we did not change the number of definitions for *low* and did not extend the test to further words.

The aforementioned comparison also alerted us to the fact that the *COBUILD* entry ‘*ethic*’ included definitions for both *ethic* and *ethics* whereas the AVL includes two separate lemmas, *ethic* and *ethics*. Consequently, the number of *COBUILD* definitions of *ethic* was manually changed from 3 to 1 and that of *ethics* from 4 to 3.

AVL words were manually looked up in the online version of WordNet<sup>5</sup>. The AVL lemmas found in WordNet and the number of WordNet definitions for each of these AVL lemmas were logged in an Excel file. 176 of the 3,013 AVL lemmas were not found in WordNet.

---

<sup>5</sup> The online version of WordNet is version 3.1. It is freely available at

<http://wordnetweb.princeton.edu/perl/webwn>.



Some of the entries in WordNet include definitions of proper names which are irrelevant to the head entries. For example, the entry for the noun male includes the definition of Male, the capital of the Maldives. Such definitions were not included in the definition counts of AVL words.

The next step in the procedure involved the exclusion of homonyms to ensure that only polysemous AVL lemmas are included in the analysis. As mentioned in section 2.2, dictionaries differ in their approach to distinguishing between polysemy and homonymy. WordNet does not include separate entries for homonyms while analyses of *COBUILD* have shown inconsistent findings. For example, it treats *bank*<sup>1</sup>, ‘financial institution’, and *bank*<sup>2</sup>, ‘the bank of a river’, as homonyms (Oliviera, Miranda, & Siqueira, 2012) whereas *school*<sup>1</sup>, ‘place for the education of children’, and *school*<sup>2</sup>, ‘school of fish’, appear within a single entry, as different senses of a polysemous word (Moerdijk, 2003).

Given that the distinction between polysemous lemmas and homographs is treated differently in these lexicographic resources, to identify AVL lemmas which were truly polysemous according to *COBUILD* and WordNet we needed to identify and exclude from further analysis any AVL items which appear in *COBUILD* and WordNet but are homographs. The AVL items in *COBUILD* and WordNet were looked up in a dictionary which extensively distinguishes between homographs, the English-US dictionary at the Oxford Dictionaries Premium online resource<sup>6</sup>. 26 AVL items which appeared in both *COBUILD* and WordNet were homographs and 2 AVL items which appeared only in WordNet were homographs. This total of 28 AVL items were excluded from further analysis. They appear in the Supplementary materials.

---

<sup>6</sup> This resource is subscription-based and available at <https://premium.oxforddictionaries.com>. In this resource homographs are numbered (e.g., *bank*<sub>1</sub> and *bank*<sub>2</sub>).

## 5. Results

This section first presents an overview of the AVL lemmas in each lexicographic resource and those shared between them (5.1). It then reports on correlations between the number of definitions for the shared AVL lemmas in each lexicographic resource and COCA-Academic frequency (5.2); these correlations address research question 1. This section then gives an overview of the AVL lemmas with more than one definition in each lexicographic resource and in both resources (5.3) as a preamble to our answer to research question 2. Research question 2 is addressed via a discussion of the shared AVL lemmas which have more than one definition in both lexicographic resources (5.4).

### 5.1 AVL lemmas in *COBUILD* and *WordNet*

Table 1 provides a breakdown of the AVL lemmas in *COBUILD* and *WordNet* and those shared between them per POS. The per POS percentage of AVL lemmas found in each resource appears within parentheses.<sup>7</sup> For example, in the first row of Table 1, the 1054 nouns in *COBUILD* account for 93.61% of the 1126 nouns in the AVL following the exclusion of those which correspond to homographs (see section 4.2).

Table 1 about here

---

<sup>7</sup> All numbers are rounded up to the second decimal place.

Table 1 shows that both *COBUILD* and WordNet provide excellent coverage of the AVL. The lemmas shared between *COBUILD* and WordNet are fewer than those found in either lexicographic resource due to a lack of complete lemma overlap. Nevertheless, a high percentage of AVL lemmas (89.55%) are shared between them.

### *5.2 Correlations between COCA-Academic frequency and number of definitions per AVL lemma*

Correlation analyses were conducted between COCA-Academic frequency and the number of AVL lemma definitions in the 2673 AVL lemmas shared between *COBUILD* and WordNet. COCA-Academic frequency is the frequency each AVL lemma has in the 120 million words that the Academic section of COCA had when the AVL list was compiled<sup>8</sup>. AVL lemmas are ordered in the AVL list according to their COCA-Academic frequency.

More precisely, the logarithm with base 10 of the COCA-Academic frequency of each word was calculated and correlations were conducted with this variable,  $[\log_{10}(\text{COCA-Academic frequency})]$ <sup>9</sup>. Table 2 provides information on the variables involved in these correlation analyses. These variables are the  $\log_{10}(\text{COCA-Academic frequency})$  of AVL lemmas shared between *COBUILD* and WordNet and the number of definitions for these

---

<sup>8</sup> COCA-Academic frequency for each AVL lemma can be found in the Excel file which also contains the AVL at <https://www.academicwords.info/>.

<sup>9</sup> Word counts from large corpora can be very disparate (e.g., the most frequent AVL lemma occurred 137,208 times in COCA-Academic at the time of AVL construction whereas the least frequent lemma occurred 111 times). Extreme word counts are likely to become outliers in statistical analyses. To decrease the possible effect of outliers on statistical results and increase symmetry around the mean score, researchers can replace raw word counts with their  $\log_{10}$  (e.g., Allen & Conklin, 2013; Author).

lemmas in each lexicographic resource. Because correlations were conducted both between the whole variables and on a POS basis, the aforementioned variables also appear per POS. Because the number of times a lemma occurred in COCA-Academic is more easily interpretable than its log10 transformation, descriptive statistics for both measures are presented.

Table 2 about here

Kolmogorov-Smirnov tests indicate that none of the variables in Table 2 were normally distributed ( $p < 0.001$  for each variable). Therefore, for all variables the median and interquartile range are more valid measures of centrality and dispersion, respectively, than the mean and SD (Field, 2013). A comparison of the median and interquartile range scores as well as the maximum number of definitions indicates that WordNet tends to include more definitions per AVL lemma in general.

In addition to not meeting the normality assumption, the variables in Table 2 also do not meet the linearity assumption of Pearson correlation, as the Loess lines on the scatterplots in Figures A, B and C in the Supplementary materials indicate. Consequently, Spearman correlations and Kendall correlations – which are appropriate for data which do not meet the assumptions of Pearson correlation (Field, 2013) – were conducted instead of Pearson correlations. We conducted Spearman correlations because they are the most commonly used correlations for non-parametric data (Field, 2013). We conducted Kendall correlations to triangulate the findings of the Spearman correlations. Table 3 reports on these correlations.

Table 3 about here

Table 3 shows that Spearman and Kendall correlations are both significant and positive. Thus, irrespective of how non-parametric correlation is calculated, findings indicate that the more frequent an AVL lemma is, the more definitions it is likely to have both in *COBUILD* and in WordNet<sup>10</sup>. Another important observation is that in both *COBUILD* and WordNet correlations for verbs have the highest strength, followed by correlations for nouns, then adjectives and finally adverbs.

### *5.3 AVL lemmas with more than one definition in COBUILD, in WordNet and in both*

Table 4 provides a breakdown per POS of the AVL lemmas which have more than one definition in each of *COBUILD* and WordNet and those shared between them. The per POS percentage of AVL lemmas found in each resource appears within parentheses.

Table 4 shows that across POS a higher percentage of AVL lemmas have more than one definition in WordNet than in *COBUILD*. The consistently higher proportion of polysemous lemmas according to WordNet than according to *COBUILD* agrees with other researchers' observation that WordNet tends to offer more definitions for words than other lexicographic resources (e.g., Navigli, Litkowski, & Hargraves, 2007).

Table 4 about here

---

<sup>10</sup> Kendall correlation coefficients are lower than Spearman correlation coefficients, as is often the case (Conover, 1999).

In terms of which parts of speech tend to be polysemous, in all columns (*COBUILD*, WordNet and Shared) verbs have the highest proportion of polysemous lemmas, followed by nouns, then adjectives and finally adverbs.

#### 5.4 Polysemous AVL lemmas

As explained in section 4.2, we consider an AVL lemma polysemous if it has more than one definition in both *COBUILD* and WordNet. Table 5 shows the 50 most frequent of these lemmas.

Table 5 about here

The full list of the resulting 919 polysemous AVL lemmas appears in Appendix 1 in the Supplementary materials, a file with the same structure as Table 5. In both, the second column indicates the ‘AVL ID number’ of each polysemous lemma, that is, the rank this lemma has in the AVL list. For example, in Table 5, *science* has AVL ID number 61, that is, it is lemma number 61 in the AVL list<sup>11</sup>.

The majority of polysemous AVL lemmas come from the first 1,000 AVL lemmas. Figure 1 shows the breakdown between polysemous AVL lemmas with IDs 1-1000 (‘AVL

---

<sup>11</sup> Table 5 does not include AVL IDs 7 (noun *research*), 13 (noun *use*), 15 (noun *data*), 20 (noun *policy*), 21 (noun *university*), 32 (adverb *both*), 50 (noun *need*), 51 (verb *base*), 53 (adjective *international*), 54 (noun *technology*), 55 (noun *individual*). Out of these lemmas, AVL ID 32 (adverb *both*) appears only in *COBUILD*, not in WordNet as well; the rest are shared between *COBUILD* and WordNet and do not have more than one definition in both.

frequency band 1'), IDs 1001-2000 ('AVL frequency band 2'), and IDs 2001-3015 ('AVL frequency band 3')<sup>12</sup>.

Figure 1 about here

Figure 1 shows that most polysemous AVL lemmas are from the first AVL frequency band irrespective of POS. Taken as a whole, polysemous AVL lemmas are also concentrated in the first AVL frequency band, making up 607 (65.05%) of the 919 polysemous AVL lemmas.

These polysemous AVL lemmas also tend to have more meaning definitions than those from the second or third AVL frequency band. Table 6 presents descriptive statistics about the number of definitions of polysemous AVL lemmas across POS and AVL frequency bands as well as overall. The data in all variables are positively skewed. Consequently, Table 6 includes median scores, maximum scores and interquartile range scores, descriptive statistics which are more appropriate to non-normally distributed data than mean and standard deviation scores (Field, 2013).

Table 6 about here

Table 6 shows that polysemous nouns, verbs and adjectives tend to have more *COBUILD* and WordNet definitions when they come from the first AVL frequency band than from the other bands. Adverbs are the only exception to the general decrease in the number of definitions as

---

<sup>12</sup> AVL homographs (see section 4.2) and the duplicate lemma *disproportionately* were excluded from this analysis. The AVL IDs in Table 5, Appendix 1 in the Supplementary materials and Figure 1 are those in the original AVL list so that readers can easily map our findings onto lemmas in the AVL list.

we move down the frequency bands; they have a median number of 2 definitions irrespective of AVL frequency band.

Finally, a comparison between the median number of definitions in the polysemous AVL lemmas which appear in both *COBUILD* and WordNet suggests that WordNet tends to list a higher number of definitions than *COBUILD* across POS and AVL frequency bands. This finding agrees with our finding that across POS a higher percentage of AVL lemmas have more than one definition in WordNet than in *COBUILD* (see Table 4).

## **6. Discussion and conclusion**

This study breaks new ground by examining polysemy in the whole AVL instead of only in a small set of English academic words. In particular, this study examined a) the relationship between AVL lemmas' frequency in COCA-Academic and their number of definitions in each of two lexicographic resources and b) the identification of polysemous AVL lemmas.

The relationship between AVL lemmas' COCA-Academic frequency and their number of meaning definitions was examined through Spearman and Kendall correlations. A statistically significant non-linear positive relationship between lemma frequency and number of meaning definitions was found both for *COBUILD* and WordNet definitions. This relationship is non-linear and positive in that low-frequency words tend to have only one definition but once a frequency threshold is passed, the number of word definitions increase as word frequency increases. As Figure A in the Supplementary materials shows, in the correlation between  $\log_{10}(\text{COCA-Academic frequency})$  and *COBUILD* definitions, the frequency threshold beyond which the correlation becomes positive is 3; a  $\log_{10}(\text{COCA-Academic frequency})$  of 3 maps onto AVL lemmas with COCA-Academic word frequencies



spanning 990 to 1008 counts in the 120 million words that the COCA-Academic corpus consisted of when the AVL was constructed. In the correlation between WordNet definitions and  $\log_{10}(\text{COCA-Academic frequency})$ , the frequency threshold beyond which the correlation becomes positive is  $\log_{10}(\text{COCA-Academic frequency}) 2.7$ , which maps onto AVL lemmas with COCA-Academic word frequencies spanning 496 to 506 counts.

These findings agree with findings of earlier studies on the relationship between the frequency of general English words and their number of meaning definitions in lexicographic resources. Out of these studies, Casas et al. (2019) is the most comparable to ours because it statistically addressed this issue via correlation (not regression). Casas et al. (2019) found a significant and non-linear correlation between WordNet definitions and frequency in a large corpus, both when the latter was operationalised as frequency in the CHILDES database (MacWhinney, 2000) and as frequency in Wikipedia.

In our study, this statistically significant non-linear and positive relationship was also found across POS (i.e., noun, verb, adjective, adverb). Since Casas et al. (2019) did not conduct correlations per POS, this analysis and finding are novel. A related notable finding is that verb definitions (both in *COBUILD* and WordNet) had the strongest correlation with COCA-Academic frequency than the definitions of any other POS, followed by noun, adjective and adverb definitions.

In terms of the second research question, our study found that, after excluding the duplicate AVL item ‘disproportionately’ and AVL homographs from further analysis, 919 (34.38%) out of the 2,673 AVL lemmas appearing in both *COBUILD* and WordNet have more than one definition in both resources. In addition to this high occurrence of polysemous lemmas in the AVL, 66.05% of the 919 polysemous lemmas come from the most frequent 1000 AVL lemmas, 24.7% from the second most frequent 1000 AVL lemmas, and 9.25% from the last 1013 AVL lemmas. This finding agrees with the findings from the correlation

analyses that the more frequent an AVL lemma is, the more likely it is to be polysemous and that below a frequency threshold very few AVL lemmas are polysemous.

These 919 polysemous AVL lemmas seem to be of high utility to university students not only because they are likely to appear often in their reading materials but also because many of them are used in university students' writing. Durrant (2016) identified 427 AVL lemmas – all from the most frequent 1000 AVL lemmas – which appear more than 12 times per million tokens in 28 or more of the 31 disciplines in the British Academic Written English (BAWE) corpus. Because of their high cross-disciplinary occurrence in British university students' writing, these 427 AVL lemmas are likely to be useful for the writing needs of all British university students (Durrant, 2016). A cross-check between them and the 919 AVL lemmas identified as polysemous in the present study shows that 299 (70.02%) of these 427 AVL lemmas are polysemous.

However, it needs to be acknowledged that not all meanings of the 919 polysemous AVL lemmas are equally likely to occur in academic contexts. In addition, some polysemous AVL lemmas may have specific meanings in specific fields. While a qualitative analysis of the meaning definitions of polysemous AVL lemmas is out of the scope of this paper, we discuss the meaning definitions of the most frequent polysemous AVL lemmas, the nouns *study* and *group* as examples, illustrated with excerpts from COCA. Of the ten definitions of *study* in WordNet and four in *COBUILD* (see Appendix 2), some (e.g., definition 3 in WordNet, 'a written document describing the findings of some individual or group', and definition 2 in *COBUILD*, 'a study of a subject is a piece of research on it') are more likely to occur in academic texts than others (e.g., meaning 5 in WordNet, 'a room used for reading and writing and studying' and meaning 4 in *COBUILD*, 'a study is a room in a house which is used for reading, writing, and studying'). Some meanings are likely to appear in specific disciplines, such as meaning 10 in WordNet ('a composition intended to develop one aspect of the

performer's technique; "a study in spiccato bowing"). By contrast, other meanings are likely to appear in both academic and non-academic texts; for instance, meaning 2 in WordNet ('applying the mind to learning and understanding a subject (especially by reading)') and meaning 1 in *COBUILD* ('study is the activity of studying') can occur in both academic and everyday contexts as the following excerpts from different sections of the COCA corpus show: 'some of them choosing Schussler Fiorenza as their object of study' (Academic), 'His rigorous study of the two thinkers, his attempt to understand them' (Fiction), 'these are fascinating creatures, worthy of study' (TV/Movies), and 'it does not require a lot of study to understand what the best situation for a child is' (Blog).

The definitions of noun *group*, the second most frequent polysemous AVL, offer further examples of the difficulty to distinguish between academic and non-academic meanings of some polysemous AVL lemmas: meaning 1 in both WordNet ('any number of entities (members considered as a unit)') and *COBUILD* ('a group of people or things is a number of people or things that are together in one place at one time') are likely to occur in both academic and non-academic discourse. As explained in section 2.2, this difficulty reflects the nature of polysemy and the fact that meanings of polysemous words are not strictly delineated in real-life discourse; rather their meaning is shaped in interaction with the context in which the word occurs. This has important implications for future research and pedagogy, which we discuss in the next sections.

### *6.1. Implications for future research*

This study identified polysemous lemmas in the whole AVL thanks to its lexicographic approach to polysemy. However, this lexicographic approach has also meant that examining

which AVL lemma meaning senses are shared across disciplines and which occur only or mainly in one discipline or a group of disciplines was out of the scope of this study. Identifying the meaning senses of AVL lemmas across disciplines is of high pedagogical value because it would direct EAP teachers and students to meaning senses that are worth learning across disciplines or within disciplinary groups. Such research would require coding many instances of polysemous AVL lemmas in large academic corpora representative of all disciplines.

Which academic-word meaning senses students will need to learn also depends on which meaning senses they already know. This fact calls for the development of academic-word tests which assess not only one's ability to recognise the most frequent meaning sense – as, for example the academic-words sublist in Schmitt, Schmitt and Clapham's (2001) versions of the Vocabulary Levels Test – but also other meaning senses of academic words. Moreover, tests examining students' ability to produce the right academic word in specific sentential contexts are needed. As we saw in section 6, 70.02% of the 427 AVL lemmas which appear more than 12 times per million tokens in 28 or more of the 31 disciplines in the BAWE corpus are polysemous according to our analysis. Finally, tests examining students' ability to recall (not just recognise) academic-word meaning senses will be useful indicators of students' reading comprehension ability because the ability to recall word meaning (meaning recall) is a more reliable predictor of reading comprehension than the ability to recognise word meaning (meaning recognition) (e.g., McLean, Stewart, & Batty, 2020).

## *6.2. Pedagogical implications*

The fact that most polysemous AVL lemmas (66.05%) are among the most frequent 1000 AVL lemmas (see Figure 1) and that, as mentioned above, 70.02% of the AVL lemmas shared across disciplines in BAWE are polysemous indicate that university students are likely to encounter many of the polysemous AVL lemmas and use them in academic writing tasks. Unfortunately, research indicates that the incidental learning of the meaning senses of words is very slow (e.g., Pigada & Schmitt, 2006). This slow learning rate can be due to aspects of students' input, such as the insufficient and ambiguous contextual clues common in natural text (e.g., Frantzen, 2003). It may also be due to learners' behaviour. For example, learners have been found to resist inferring meaning senses from context because of a 'monosemy bias'. Research suggests that learners tend to assume that the meaning sense they already know for a word applies to all contexts (e.g., Bensoussan & Laufer, 1984; Frantzen, 2003). This can cause particular problems if the meaning sense they already know is the one typically used in everyday life rather than in academic discourse. Combined with L2 university students' difficulty with selecting appropriate definitions from the dictionary entries of polysemous words (Nesi & Haill, 2002), this monosemy bias indicates the need to alert students to the polysemous nature of high-frequency academic words.

To raise students' awareness of the polysemous nature of many academic words, EAP teachers and materials developers can design tasks consisting of sets of sentences or passages containing selected polysemous academic words used with different meaning senses; students can then be asked to infer a word's meaning from context and/or select the relevant definition for each word from a dictionary. Because, as mentioned, L2 students may face difficulty with dictionary consultation, teachers may need to raise learners' awareness of various aspects of dictionaries which facilitate lookup (e.g., sense signposts, grammatical/collocational information in a dictionary entry). The list of polysemous AVL items provided in this study (see Appendix 1 in the Supplementary materials) can serve as a resource for word selection

for such tasks, while suitable authentic sentences and passages containing target AVL items can be sourced from freely accessible academic and general corpora (e.g., COCA, including its academic section, and the British National Corpus are available at <https://www.english-corpora.org>).

Dictionary selection is another important issue to consider. As our study has found, dictionaries differ in how they treat polysemy and how many meaning senses of academic words they identify. While our study used general English dictionaries aimed at learners (*COBUILD*) and expert users (WordNet), EAP practitioners may consider using a dictionary of academic vocabulary usage, such as the *Oxford Learner's Dictionary of Academic English*, which uses sentences from the Oxford Corpus of Academic English to illustrate the meaning definitions of words, including those that students may previously have encountered in non-academic contexts.

Tasks applying the principles of data-driven learning (see, e.g., Friginal, 2018) can be designed to promote students' discovery of meaning variation of polysemous AVL items across registers and disciplines. Such tasks would require students to analyse concordance lines containing target polysemous AVL lemmas and classify their occurrences according to similarities in meaning, which can be followed by small group comparison and discussion activities. Existing academic and discipline-specific corpora, such as the disciplinary subsections of COCA-Academic, can be used for such data-driven tasks. Alternatively, students can be encouraged to compile their own small corpora from articles in their discipline (see, e.g., Charles, 2012), which they can explore to gain an understanding of the meaning senses of target polysemous AVL words commonly used in their field.

## References

- Alexander, M. (2015). Words and dictionaries. In: J.R. Taylor (Ed.), *The Oxford handbook of the word* (pp. 37–53). Oxford: Oxford University Press.
- Allen, D.B., & Conklin, K. (2013). Cross-linguistic similarity and task demands in Japanese-English bilingual processing. *PLOS ONE*, 8, e72631.  
<https://doi.org/10.1371/journal.pone.0072631>
- Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, 6, 253–279.
- Bensoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading*, 7, 15–32.
- Casas, B., Hernández-Fernández, Català, N., Ferrer-i-Cancho, R., & Baixeries, J. (2019). Polysemy and brevity versus frequency in language. *Computer Speech and Language*, 58, 19–50.
- Charles, M. (2012). ‘Proper vocabulary and juicy collocations’: EAP students evaluate do-it yourself corpus-building. *English for Specific Purposes*, 31, 93–102.
- Cobb, T. (2010). Learning and language and learners from computer programs. *Reading in a Foreign Language*, 22, 181–200.
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education*. (8<sup>th</sup> edn). New York: Routledge.
- Conover, W.J. (1999). *Practical nonparametric statistics*. (3rd edn). New York: Wiley.
- Coxhead, A. (2000). A new Academic Word List. *TESOL Quarterly*, 34, 213–238.
- Cruse, D.A. (2004). *Meaning in language: An introduction to semantics and pragmatics*. Oxford: Oxford University Press.
- Deignan, A. (2015). MIP, the corpus and dictionaries. What makes for the best metaphor analysis? *Metaphor and the Social World*, 5, 145–154.
- Dobrić, N. (2015). Three-factor prototypicality evaluation and the verb *look*. *Language*

- Sciences*, 50, 1–11.
- Dorst, A.G., & Reijniere, W. G. (2015). A dictionary gives definitions, not decisions. Response 1 to ‘On using a dictionary to identify the basic senses of words’. *Metaphor and the Social World*, 5, 137–144.
- Durrant, P. (2014). Discipline and level specificity in university students’ written vocabulary. *Applied Linguistics*, 35, 328–356.
- Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes*, 43, 49–61.
- Falkum, I. L., & Vicente, A. (2015). Polysemy: Current perspectives and approaches. *Lingua*, 157, 1–16.
- Fellbaum, C. (Ed.). (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.
- Field, A. (2013). *Discovering statistics using IBM SPSS Statistics*. London: Sage.
- Frantzen, D. (2003). Factors affecting how second language Spanish students derive meaning from context. *The Modern Language Journal*, 87, 168–199.
- Friginal, E. (2018). *Corpus linguistics for English teachers: Tools, online resources, and classroom activities*. Abingdon/New York: Routledge.
- Gardner, D., & Davies, M. (2014). A new Academic Vocabulary List. *Applied Linguistics*, 35, 305–327.
- Geeraerts, D. (2016). Sense individuation. In N. Riemer (Ed.), *Routledge handbook of semantics* (pp. 233–247). Abingdon: Routledge.
- Green, C. (2019). Enriching the academic wordlist and secondary vocabulary lists with lexicogrammar: Toward a pattern grammar of academic vocabulary. *System*, 87, Article 102158. <https://doi.org/10.1016/j.system.2019.102158>
- Granger, S., & Paquot, M. (2009). In search of a general academic vocabulary: A corpus-



- driven study. In K. Katsampoxaki-Hodgetts (Ed.), *Options and practices of LSP practitioners* (pp. 94–108). Heraklion: University of Crete Publications. Available at <http://hdl.handle.net/2078.1/75685>
- Hanks, P. (2001). *The pattern dictionary of English verbs*. Masaryk, Czech Republic: Masaryk University, Czech Republic. Available at <http://www.pdev.org.uk>
- Hanks, P. (2004). Corpus pattern analysis. In G. Williams & S. Vessier (Eds.), *Proceedings of the eleventh EURALEX international congress* (pp.87–97). Université de Bretagne-Sud, Lorient: EURALEX. Available at [https://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202004/009\\_2004\\_V1\\_Patrick%20HANKS\\_Corpus%20pattern%20analysis.pdf](https://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202004/009_2004_V1_Patrick%20HANKS_Corpus%20pattern%20analysis.pdf)
- Hanks, P. (2013). *Lexical analysis: Norms and exploitations*. MA: MIT Press.
- Hyland, K., & Tse, P. (2007). Is there an “Academic Vocabulary”? *TESOL Quarterly*, 41, 235–253.
- Hyland, K., & Tse, P (2009). Academic lexis and disciplinary practice: Corpus evidence for specificity. *International Journal of English Studies*, 9, 111–129.
- Kilgarriff, A. (2013). Using corpora as data sources for dictionaries. In H. Jackson (Ed.), *The Bloomsbury companion to lexicography* (pp. 77–96). London: Bloomsbury.
- Kuiper, K., Fromont, R., & Gerhard, D. (2017). Polysemy and word frequency: A replication. *Journal of Research Design and Statistics in Linguistics and Communication Science*, 4, 144–155.
- Lei, L., & Liu, D. (2018). The academic English collocation list: A corpus-driven study. *International Journal of Corpus Linguistics*, 23, 216–243.
- Lew, R. (2013). Identifying, ordering and defining senses. In H. Jackson (Ed.), *The Bloomsbury companion to lexicography* (pp. 284–303). London: Bloomsbury.

- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. (3rd edn). Mahwah, NJ: Lawrence Erlbaum Associates.
- Martínez, I., Beck, S., & Panza, C. (2009). Academic vocabulary in agricultural research articles: A corpus-based study. *English for Specific Purposes*, 28, 183–198.
- Masrai, A., & Milton, J. (2018). Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement. *Journal of English for Academic Purposes*, 31, 44–57.
- McLean, S., Stewart, J., & Batty, A.O. (2020). Predicting L2 reading proficiency with modalities of vocabulary knowledge: A bootstrapping approach. *Language Testing*, 37, 389–411.
- Moerdijk, F. (2003). The codification of semantic information. In P. van Sterkenburg (Ed.), *A practical guide to lexicography* (pp. 273–298). Amsterdam/Philadelphia: John Benjamins.
- Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes*, 25, 235–256.
- Murphy, L. (2010). *Lexical meaning*. Cambridge: Cambridge University Press.
- Nation, P. (2013). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Navigli, R., Litkowski, K. C., & Hargraves, O. (2007). SemEval-2007 task 07: Coarse grained English all-words task. In *Proceedings of the 4th international workshop on semantic evaluations* (pp. 30–35). Stroudsburg, USA: Association for Computational Linguistics.
- Nesi, H., & Haill, R. (2002). A study of dictionary use by international students at a British university. *International Journal of Lexicography*, 15, 277–305.
- Nesi, H., & Tan, K. H. (2011) The effect of menus and signposting on the speed and accuracy

- of sense selection. *International Journal of Lexicography*, 24, 79–96.
- Oliviera, A.F.S.d., Miranda, F.B., & Siqueira, M. (2012). O tratamento da polissemia e da homonímia nos learners' dictionaries: Subsídios da semântica cognitiva para a disposição das acepções. *Alfa Revista de Linguística*, 57, 163–197.
- Partington, A. (1998). *Patterns and meanings. Using corpora for English language research and teaching*. Amsterdam/Philadelphia: John Benjamins.
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, 18, 1–28.
- Pragglejaz Group. (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22, 1–39.
- Reynolds, B. L., Wu, W-H., Liu, H-W., Kuo, S-Y., & Yeh, C-H. (2015). Towards a model of advanced learners' vocabulary acquisition: An investigation of L2 vocabulary acquisition and retention by Taiwanese English majors. *Applied Linguistics Review*, 6, 121–144.
- Rhees, G. P. (2018). *A phraseological multi-discipline approach to vocabulary selection for English for academic purposes*. Unpublished PhD thesis. University Pompeu Fabra, Spain.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing*, 18, 55–88.
- Storjohann, P. (2016). Sense relations. In N. Riemer (Ed.), *The Routledge handbook of semantics* (pp. 248–265). Abingdon: Routledge.
- Thompson, P., & Alzeer, S. (2019). A survey of issues, practices and views related to corpus-based word lists for English language teaching and learning. *International Journal of Applied Linguistics & English Literature*, 8, 43–53.

- van Rossum, G. (1995). Python tutorial, Technical Report CS-R9526. Amsterdam: Centrum voor Wiskunde en Informatica (CWI).
- Walter, E. (2010). Using a corpus to write dictionaries. In A. O’Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 428–443). Abingdon: Routledge.
- Wang, K. M-T., & Nation, P. (2004). Word meaning in academic English: Homography in the academic word list. *Applied Linguistics*, 25, 291–314.
- Ward, J., & Chuenjundaeng, J. (2009). Suffix knowledge: Acquisition and applications. *System*, 37, 461–469.
- West, M. (1953). *A general service list of English words*. London: Longman.
- Yamada, S. (2013). Monolingual learners’ dictionaries: Where now? In H. Jackson (Ed.), *The Bloomsbury companion to lexicography* (pp. 188–212). London: Bloomsbury.
- Zipf, G.K. (1945). The meaning-frequency relationship of words. *Journal of General Psychology*, 33, 251–256.
- Zipf, G. K. (1949). *Human behaviour and the principle of least effort*. Cambridge, MA: Addison-Wesley.

**Table 1**AVL lemma coverage by *COBUILD*, WordNet and shared coverage

POS	AVL items without homographs and 'disproportionately'	<i>COBUILD</i>	WordNet	Shared
Noun	1126	1054 (93.61%)	1087 (96.54%)	1030 (91.47%)
Verb	542	528 (97.42%)	539 (99.45%)	528 (97.42%)
Adjective	1036	963 (92.95%)	915 (88.32%)	884 (85.33%)
Adverb	281	240 (85.41%)	268 (95.37%)	231 (82.21%)
<b>Total</b>	<b>2985</b>	<b>2785 (93.3%)</b>	<b>2809 (94.1%)</b>	<b>2673 (89.55%)</b>

**Table 2**

Descriptive statistics for the variables in the correlation analyses conducted to address Research Question 1

	POS	Mean	Median	Min	Max	SD	Interquartile range	Skewness	Kurtosis
<i>COCA-Academic frequency</i>	Noun	7071.84	1757	111	1757	13310.29	6659.25	3.96	22.49
	Verb	5677.43	1524	120	93212	9926.02	5471.75	3.51	17.92
	Adjective	3409.29	890	111	99744	7433.23	2468.75	5.64	47.55
	Adverb	3465.34	1066	114	90906	7772.74	3071	7.37	73.48
	<i>All</i>	5273.47	1319	111	137208	10669.10	4392	4.60	31.11
<i>Log10(COCA-Academic frequency)</i>	Noun	3.31	3.24	2.05	5.14	0.70	1.09	0.23	-0.84
	Verb	3.27	3.18	2.08	4.97	0.65	1.02	0.32	-0.80
	Adjective	3.05	2.95	2.05	5.00	0.61	0.87	0.54	-0.38
	Adverb	3.09	3.03	2.06	4.96	0.60	0.97	0.45	-0.60
	<i>All</i>	3.20	3.12	2.05	5.14	0.67	1.03	0.39	-0.68
<i>COBUILD definitions</i>	Noun	1.81	1	1	10	1.25	1	2.15	5.54
	Verb	1.95	2	1	11	1.37	1	2.25	6.92
	Adjective	1.52	1	1	15	1.04	1	4.57	39.34
	Adverb	1.19	1	1	4	0.51	0	3.07	10.27
	<i>All</i>	1.69	1	1	15	1.18	1	2.87	13.34
<i>WordNet definitions</i>	Noun	3.1	3	1	16	2.17	2	1.59	3.35
	Verb	3.34	3	1	21	2.52	2	2.63	11.1
	Adjective	2.19	2	1	13	1.57	2	2.37	8.16
	Adverb	1.41	1	1	7	0.75	1	2.9	13.91
	<i>All</i>	2.7	2	1	21	2.08	2	2.31	8.86

**Table 3**

Spearman's rho and Kendall's tau correlations conducted to address Research Question 1

		Log10(COCA-Academic frequency)	
		Spearman's rho	Kendall's tau
<i>COBUILD</i> definitions for	shared AVL nouns	0.54	0.43
	shared AVL verbs	0.59	0.47
	shared AVL adjectives	0.37	0.3
	shared AVL adverbs	0.31	0.26
	<i>all shared AVL lemmas</i>	0.49	0.39
WordNet definitions for	shared AVL nouns	0.5	0.37
	shared AVL verbs	0.52	0.39
	shared AVL adjectives	0.35	0.27
	shared AVL adverbs	0.2	0.16
	<i>all shared AVL lemmas</i>	0.45	0.34

*Note.* All correlations are two-tailed and significant at the 0.001 level

**Table 4**

AVL lemmas with more than one definition in *COBUILD*, WordNet and among those shared between *COBUILD* and WordNet

POS	AVL items without homographs and 'disproportionately'	AVL items with more than one definition		
		<i>COBUILD</i>	WordNet	Shared
Noun	1126	460 (40.85%)	798 (70.87%)	415 (36.86%)
Verb	542	265 (48.89%)	436 (80.44%)	240 (44.28%)
Adjective	1036	300 (28.96%)	532 (51.35%)	244 (23.55%)
Adverb	281	34 (12.1%)	70 (24.91%)	20 (7.12%)
<b>Total</b>	<b>2985</b>	<b>1059 (35.48%)</b>	<b>1836 (61.51%)</b>	<b>919 (30.79%)</b>



**Table 5**50 most frequent polysemous AVL lemmas shared by *COBUILD* and WordNet

Rank	AVL ID number	Lemma	POS	<i>COBUILD</i> definitions	WordNet definitions
1	1	study	Noun	4	10
2	2	group	Noun	5	3
3	3	system	Noun	7	9
4	4	social	Adj.	3	6
5	5	provide	Verb	2	7
6	6	however	Adv.	3	4
7	8	level	Noun	5	8
8	9	result	Noun	4	4
9	10	include	Verb	2	4
10	11	important	Adj.	2	5
11	12	process	Noun	2	6
12	14	development	Noun	6	9
13	16	information	Noun	3	5
14	17	effect	Noun	4	6
15	18	change	Noun	6	10
16	19	table	Noun	2	6
17	22	model	Noun	7	9
18	23	experience	Noun	3	3
19	24	activity	Noun	3	6
20	25	human	Adj.	2	3
21	26	history	Noun	5	5
22	27	develop	Verb	11	21
23	28	suggest	Verb	4	4
24	29	economic	Adj.	2	5
25	30	low	Adj.	15	10
26	31	relationship	Noun	3	4
27	33	value	Noun	5	6
28	34	require	Verb	2	4
29	35	role	Noun	2	4
30	36	difference	Noun	3	5
31	37	analysis	Noun	3	6
32	38	practice	Noun	4	5
33	39	society	Noun	4	4
34	40	thus	Adv.	2	2
35	41	control	Noun	6	11
36	42	form	Noun	6	16
37	43	report	Verb	5	6
38	44	rate	Noun	4	4
39	45	significant	Adj.	2	4
40	46	figure	Noun	10	13
41	47	factor	Noun	3	7
42	48	interest	Noun	7	7
43	49	culture	Noun	4	7
44	52	population	Noun	2	5
45	56	type	Noun	3	6

46	57	describe	Verb	2	4
47	58	indicate	Verb	6	5
48	59	image	Noun	4	9
49	60	subject	Noun	7	8
50	61	science	Noun	3	2

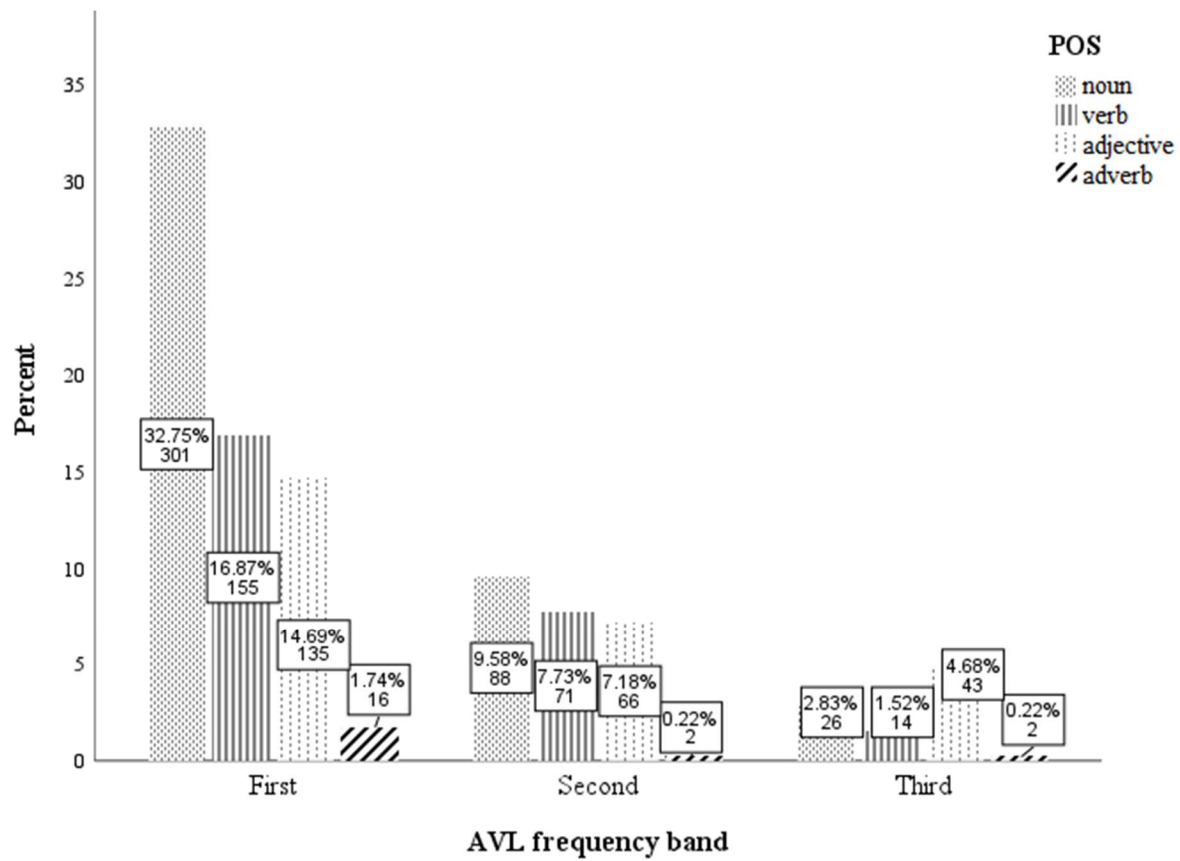
---

**Table 6**Descriptive statistics for the number of definitions of polysemous AVL lemmas shared by *COBUILD* and WordNet

AVL frequency band	POS	<i>COBUILD</i>					WordNet				
		Median	Max	Interquartile range	Skewness	Kurtosis	Median	Max	Interquartile range	Skewness	Kurtosis
1	Noun	3	10	2	1.43	2.24	5	16	3	1.07	1.35
	Verb	3	11	2	1.62	3.67	4	21	3	2.32	7.55
	Adj.	3	15	1	3.62	20.32	4	13	3	1.53	2.76
	Adv.	2	4	1	1.50	1.58	2.5	4	1	0.73	-0.54
	<i>All</i>	3	15	2	2.15	8.39	4	21	3	1.76	5.31
2	Noun	2	4	0	4.26	19.38	3	8	1	1.07	1.11
	Verb	2	6	0	4.36	20.29	3	8	2	0.98	0.85
	Adj.	2	4	0	5.38	30.63	3	9	2	2.25	6.32
	Adv.	2	2	0	N/A	N/A	2	2	0	N/A	N/A
	<i>All</i>	2	6	0	5.51	35.79	3	9	2	1.37	2.27
3	Noun	2	7	0	2.83	8.03	3	6	1	1.93	5.37
	Verb	2	3	1	1.07	-1.03	3.5	9	3	0.98	0.57
	Adj.	2	5	0	2.83	8.00	2	6	1	1.71	2.54
	Adv.	2	2	0	N/A	N/A	2.5	3	0	N/A	N/A
	<i>All</i>	2	7	0	3.38	13.24	3	9	1	2.01	4.76
All AVL	Noun	2	10	1	1.77	3.37	4	16	3	1.31	2.10
	Verb	2	11	2	2.03	5.40	4	21	3	2.55	9.87
	Adj.	2	15	1	4.45	30.88	3	13	2	1.87	4.28
	Adv.	2	4	1	1.84	2.86	2	4	1	0.89	-0.24
	<i>All</i>	2	15	1	2.56	10.81	4	21	3	1.99	6.71

Note 1: Skewness and kurtosis data are not applicable for AVL frequency band 2 and 3 adverbs because in each of these AVL frequency bands only two such adverbs have more than one definition in both *COBUILD* and WordNet.

Note 2: The minimum number of senses for all POS in all AVL frequency bands is 2.



**Figure 1.** Percentage of polysemous AVL lemmas across AVL frequency bands and POS

**Sophia Skoufaki** is Associate Supervisor at the University of Essex. She holds a PhD from the University of Cambridge. She specialises in second language vocabulary learning and teaching. Her research has appeared in journals such as *Applied Linguistics Review*, *English for Specific Purposes*, *Metaphor and Symbol* and *Text & Talk*.

**Bojana Petrić** is Reader at Birkbeck, University of London. She has co-authored *Experiencing Master's supervision: Perspectives of international students and their supervisors* (Routledge, 2017) and has published articles in journals such as *English for Specific Purposes*, *Journal of English for Academic Purposes*, *Journal of Second Language Writing* and *Written Communication*.

**Sophia Skoufaki:** Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Supervision, Visualization, Writing – original draft, Writing – review & editing

**Bojana Petrić:** Conceptualisation, Methodology, Writing – original draft, Writing – review & editing

## Appendix 1

ID	AVL	PoS	COBUILD definitions	WordNet definitions
1	1 study	Noun	4	10
2	2 group	Noun	5	3
3	3 system	Noun	7	9
4	4 social	Adjective	3	6
5	5 provide	Verb	2	7
6	6 however	Adverb	3	4
7	8 level	Noun	5	8
8	9 result	Noun	4	4
9	10 include	Verb	2	4
10	11 important	Adjective	2	5
11	12 process	Noun	2	6
12	14 development	Noun	6	9
13	16 information	Noun	3	5
14	17 effect	Noun	4	6
15	18 change	Noun	6	10
16	19 table	Noun	2	6
17	22 model	Noun	7	9
18	23 experience	Noun	3	3
19	24 activity	Noun	3	6
20	25 human	Adjective	2	3
21	26 history	Noun	5	5
22	27 develop	Verb	11	21
23	28 suggest	Verb	4	4
24	29 economic	Adjective	2	5
25	30 low	Adjective	15	10
26	31 relationship	Noun	3	4
27	33 value	Noun	5	6
28	34 require	Verb	2	4
29	35 role	Noun	2	4
30	36 difference	Noun	3	5
31	37 analysis	Noun	3	6
32	38 practice	Noun	4	5
33	39 society	Noun	4	4
34	40 thus	Adverb	2	2
35	41 control	Noun	6	11
36	42 form	Noun	6	16
37	43 report	Verb	5	6
38	44 rate	Noun	4	4
39	45 significant	Adjective	2	4
40	46 figure	Noun	10	13
41	47 factor	Noun	3	7
42	48 interest	Noun	7	7

43	49	culture	Noun	4	7
44	52	population	Noun	2	5
45	56	type	Noun	3	6
46	57	describe	Verb	2	4
47	58	indicate	Verb	6	5
48	59	image	Noun	4	9
49	60	subject	Noun	7	8
50	61	science	Noun	3	2
51	62	material	Noun	4	5
52	63	produce	Verb	6	7
53	64	condition	Noun	5	8
54	65	identify	Verb	6	6
55	67	support	Noun	3	11
56	68	performance	Noun	3	5
57	69	project	Noun	2	2
58	71	approach	Noun	2	9
59	72	support	Verb	6	11
60	73	period	Noun	6	7
61	74	organization	Noun	3	7
62	76	environmental	Adjective	2	2
63	77	source	Noun	4	9
64	78	nature	Noun	3	5
65	79	cultural	Adjective	2	4
66	80	resource	Noun	2	3
67	81	century	Noun	2	2
68	82	strategy	Noun	2	2
69	83	theory	Noun	3	3
70	84	product	Noun	2	6
71	86	goal	Noun	3	4
72	87	likely	Adjective	2	4
73	88	note	Verb	4	4
74	89	represent	Verb	7	15
75	90	general	Adjective	5	6
76	91	article	Noun	4	4
77	93	environment	Noun	3	2
78	94	language	Noun	6	6
79	95	determine	Verb	4	8
80	96	structure	Noun	3	5
81	97	section	Noun	2	14
82	98	common	Adjective	7	9
83	99	occur	Verb	3	3
84	101	available	Adjective	2	3
85	103	term	Noun	7	7
86	104	reduce	Verb	5	20
87	105	measure	Noun	5	9
88	106	involve	Verb	5	7



89	107	movement	Noun	5	11
90	108	specific	Adjective	3	4
91	109	focus	Verb	3	5
92	110	region	Noun	2	5
93	111	relate	Verb	3	5
94	112	individual	Adjective	2	4
95	113	quality	Noun	4	5
96	114	establish	Verb	4	8
97	115	author	Noun	2	2
98	116	seek	Verb	4	5
99	117	compare	Verb	4	4
100	118	growth	Noun	6	7
101	119	natural	Adjective	5	10
102	120	various	Adjective	2	4
103	121	standard	Noun	3	6
104	122	example	Noun	3	6
105	123	management	Noun	3	2
106	125	argue	Verb	5	3
107	126	degree	Noun	4	7
108	127	design	Noun	4	7
109	128	concern	Noun	5	5
110	131	examine	Verb	4	5
111	132	pattern	Noun	3	8
112	135	traditional	Adjective	2	2
113	136	finding	Noun	2	3
114	137	positive	Adjective	9	11
115	138	central	Adjective	4	2
116	139	act	Noun	5	5
117	140	impact	Noun	2	4
118	141	reflect	Verb	6	7
119	142	recognize	Verb	4	9
120	143	context	Noun	2	2
121	144	relation	Noun	3	6
122	145	maintain	Verb	5	10
123	148	discussion	Noun	2	2
124	149	associate	Verb	3	3
125	150	design	Verb	3	7
126	151	particularly	Adverb	2	3
127	152	purpose	Noun	3	3
128	153	address	Verb	2	10
129	155	particular	Adjective	4	6
130	156	benefit	Noun	5	3
131	157	survey	Noun	3	3
132	158	effective	Adjective	3	6
133	159	apply	Verb	6	10
134	160	contain	Verb	5	6

135	161	understanding	Noun	5	4
136	162	production	Noun	5	8
137	163	form	Verb	7	7
138	164	association	Noun	3	8
139	165	reveal	Verb	2	3
140	166	range	Noun	7	9
141	169	status	Noun	5	2
142	170	necessary	Adjective	2	2
143	171	function	Noun	2	7
144	172	indeed	Adverb	3	2
145	174	global	Adjective	2	2
146	175	conflict	Noun	4	7
147	177	conduct	Verb	4	6
148	178	critical	Adjective	6	7
149	179	perform	Verb	4	4
150	180	discuss	Verb	2	2
151	181	exist	Verb	2	2
152	182	improve	Verb	4	2
153	183	observe	Verb	5	9
154	184	demonstrate	Verb	4	4
155	185	unit	Noun	6	6
156	186	total	Adjective	2	2
157	187	modern	Adjective	4	5
158	188	literature	Noun	3	4
159	189	result	Verb	2	3
160	190	experience	Verb	2	5
161	191	principle	Noun	3	6
162	192	element	Noun	7	7
163	193	challenge	Noun	2	5
164	194	control	Verb	5	9
165	195	historical	Adjective	3	4
166	196	aspect	Noun	2	5
167	199	measure	Verb	3	4
168	201	belief	Noun	3	2
169	202	western	Adjective	2	4
170	207	technique	Noun	2	2
171	210	generally	Adverb	2	3
172	213	importance	Noun	2	2
173	214	application	Noun	4	7
174	215	feature	Noun	5	6
175	216	influence	Noun	3	5
176	217	basis	Noun	4	3
177	219	refer	Verb	8	7
178	220	communication	Noun	2	3
179	221	negative	Adjective	7	9
180	222	primary	Adjective	3	4

181	228	variety	Noun	3	6
182	230	following	Adjective	2	4
183	231	access	Noun	3	6
184	232	contribute	Verb	3	4
185	233	assume	Verb	3	9
186	235	tool	Noun	2	4
187	236	promote	Verb	3	5
188	238	labor	Noun	4	7
189	239	engage	Verb	5	10
190	240	review	Noun	2	10
191	242	highly	Adverb	4	3
192	244	publish	Verb	4	3
193	245	encourage	Verb	4	3
194	247	assess	Verb	2	4
195	248	view	Verb	3	3
196	250	instrument	Noun	4	6
197	252	meaning	Noun	3	2
198	253	limit	Verb	3	3
199	255	directly	Adverb	2	4
200	256	previous	Adjective	2	3
201	257	demand	Noun	3	5
202	258	vision	Noun	5	5
203	259	female	Adjective	2	3
204	260	attempt	Noun	2	2
205	261	influence	Verb	2	3
206	262	independent	Adjective	6	4
207	263	solution	Noun	3	5
208	264	direct	Adjective	5	10
209	265	conclusion	Noun	3	9
210	266	presence	Noun	5	6
211	267	scientific	Adjective	2	2
212	268	ethnic	Adjective	3	2
213	270	active	Adjective	7	13
214	271	male	Noun	2	2
215	272	claim	Noun	3	6
216	274	focus	Noun	4	7
217	275	contrast	Noun	2	5
218	276	failure	Noun	6	7
219	277	internal	Adjective	2	5
220	278	journal	Noun	3	5
221	280	facility	Noun	3	5
222	282	emerge	Verb	5	5
223	283	protection	Noun	3	7
224	284	extent	Noun	2	2
225	285	male	Adjective	2	3
226	286	mental	Adjective	3	5

227	287	explore	Verb	4	4
228	289	generate	Verb	2	4
229	292	requirement	Noun	2	3
230	293	broad	Adjective	5	8
231	294	observation	Noun	4	5
232	296	difficulty	Noun	2	4
233	298	perceive	Verb	2	2
234	301	female	Noun	2	2
235	302	capacity	Noun	6	9
236	309	extend	Verb	9	17
237	310	connection	Noun	4	9
238	311	sector	Noun	3	6
239	312	commitment	Noun	4	5
240	313	interpretation	Noun	2	4
241	315	conclude	Verb	4	5
242	316	notion	Noun	2	4
243	318	domestic	Adjective	5	5
244	319	consist	Verb	2	4
245	320	reference	Noun	6	10
246	322	adopt	Verb	2	7
247	323	comparison	Noun	2	3
248	324	depend	Verb	4	2
249	326	standard	Adjective	2	5
250	328	employ	Verb	3	2
251	329	definition	Noun	2	2
252	330	essential	Adjective	2	5
253	331	contact	Noun	4	9
254	334	actual	Adjective	2	5
255	336	dimension	Noun	3	4
256	337	theme	Noun	4	5
257	338	largely	Adverb	2	2
258	339	link	Verb	4	4
259	340	desire	Noun	2	3
260	343	consistent	Adjective	3	4
261	344	distribution	Noun	2	4
262	345	minority	Noun	2	3
263	346	analyze	Verb	2	4
264	348	psychological	Adjective	2	2
265	349	unique	Adjective	3	4
266	350	experiment	Noun	2	3
267	351	trend	Noun	2	4
268	352	exchange	Noun	3	11
269	355	implication	Noun	2	5
270	356	contribution	Noun	2	5
271	358	organize	Verb	4	6
272	361	emotional	Adjective	3	4

273	362	locate	Verb	3	4
274	366	improvement	Noun	2	3
275	370	rural	Adjective	2	2
276	374	core	Noun	4	9
277	375	volume	Noun	5	6
278	377	limited	Adjective	2	7
279	378	propose	Verb	4	5
280	379	framework	Noun	2	3
281	381	creation	Noun	3	6
282	382	code	Noun	6	3
283	383	emphasis	Noun	2	4
284	384	industrial	Adjective	2	4
285	385	external	Adjective	3	4
286	386	waste	Noun	2	5
287	387	potential	Noun	2	2
288	388	climate	Noun	2	2
289	389	explanation	Noun	3	3
290	390	technical	Adjective	3	7
291	391	mechanism	Noun	3	5
292	392	description	Noun	3	3
293	393	vary	Verb	2	4
294	394	reduction	Noun	2	3
295	395	discipline	Noun	4	5
296	396	construct	Verb	2	6
297	397	equal	Adjective	4	2
298	398	origin	Noun	2	6
299	400	fundamental	Adjective	4	3
300	404	existence	Noun	2	2
301	405	formal	Adjective	5	6
302	406	manner	Noun	3	3
303	407	assistance	Noun	4	3
304	412	planning	Noun	2	3
305	415	cite	Verb	4	7
306	417	judgment	Noun	3	7
307	418	constitute	Verb	2	4
308	420	typical	Adjective	3	3
309	421	selection	Noun	4	5
310	422	incorporate	Verb	2	4
311	423	illustrate	Verb	3	3
312	424	cycle	Noun	3	6
313	425	depression	Noun	4	10
314	426	consideration	Noun	4	6
315	428	arise	Verb	2	7
316	430	separate	Adjective	2	4
317	431	recognition	Noun	4	8
318	432	mode	Noun	3	6

319	434	resistance	Noun	5	11
320	436	diversity	Noun	2	2
321	437	practical	Adjective	4	4
322	439	acquire	Verb	3	7
323	440	characterize	Verb	2	2
324	441	differ	Verb	2	2
325	442	review	Verb	3	5
326	443	interpret	Verb	2	6
327	444	creative	Adjective	3	2
328	445	limitation	Noun	4	5
329	446	resolution	Noun	4	11
330	450	revolution	Noun	2	3
331	451	philosophy	Noun	3	3
332	452	display	Verb	4	2
333	454	publication	Noun	3	4
334	455	variation	Noun	2	11
335	456	derive	Verb	2	5
336	459	permit	Verb	2	3
337	460	alternative	Adjective	4	3
338	462	initiative	Noun	4	2
339	463	employment	Noun	3	4
340	464	regard	Verb	2	3
341	465	estimate	Noun	3	5
342	468	transform	Verb	2	7
343	469	absence	Noun	2	4
344	470	imply	Verb	2	5
345	472	observer	Noun	3	2
346	475	link	Noun	6	9
347	476	evolution	Noun	2	2
348	478	signal	Noun	4	3
349	482	biological	Adjective	5	2
350	483	introduction	Noun	3	7
351	484	boundary	Noun	2	3
352	488	theoretical	Adjective	2	2
353	491	settlement	Noun	4	7
354	492	independence	Noun	2	3
355	493	yield	Verb	7	13
356	494	formation	Noun	4	7
357	495	insight	Noun	2	4
358	496	territory	Noun	5	3
359	497	conventional	Adjective	3	7
360	498	inform	Verb	3	3
361	499	index	Noun	2	5
362	505	distinction	Noun	4	4
363	506	relative	Adjective	3	2
364	507	identification	Noun	5	5

365	509	monitor	Verb	2	2
366	510	domain	Noun	2	5
367	514	strategic	Adjective	3	2
368	515	preference	Noun	2	4
369	516	profession	Noun	2	4
370	517	apparent	Adjective	2	2
371	518	assign	Verb	4	8
372	522	dependent	Adjective	2	6
373	523	presentation	Noun	3	7
374	524	proportion	Noun	4	5
375	525	universal	Adjective	2	3
376	526	norm	Noun	3	2
377	527	tendency	Noun	2	4
378	530	equally	Adverb	3	2
379	531	resolve	Verb	2	7
380	532	competitive	Adjective	3	3
381	533	related	Adjective	3	2
382	534	symbol	Noun	3	2
383	535	consumption	Noun	3	4
384	536	calculate	Verb	2	6
385	538	extensive	Adjective	3	3
386	539	barrier	Noun	5	3
387	540	advanced	Adjective	4	8
388	542	adjustment	Noun	2	5
389	543	shape	Verb	2	3
390	544	integrate	Verb	3	4
391	545	dominate	Verb	3	5
392	546	establishment	Noun	3	7
393	547	entry	Noun	8	6
394	548	visible	Adjective	2	3
395	551	sequence	Noun	2	5
396	552	hence	Adverb	2	3
397	556	dialogue	Noun	2	4
398	557	distinct	Adjective	4	5
399	558	enterprise	Noun	4	3
400	560	scope	Noun	2	4
401	561	assert	Verb	4	4
402	562	capability	Noun	2	3
403	563	reflection	Noun	5	8
404	564	precisely	Adverb	3	3
405	565	electronic	Adjective	2	2
406	567	distinguish	Verb	4	5
407	570	evolve	Verb	2	3
408	572	survival	Noun	2	3
409	573	recommendation	Noun	2	3
410	576	encounter	Verb	2	5

411	577	membership	Noun	2	2
412	579	adapt	Verb	2	2
413	583	rapid	Adjective	2	2
414	584	collective	Adjective	3	3
415	585	reinforce	Verb	4	2
416	586	ethical	Adjective	2	3
417	587	exhibit	Verb	3	4
418	589	function	Verb	2	3
419	590	communicate	Verb	3	7
420	591	valuable	Adjective	2	2
421	594	wealth	Noun	2	4
422	597	initiate	Verb	3	5
423	599	ideal	Adjective	2	3
424	602	indicator	Noun	2	4
425	603	strengthen	Verb	8	3
426	606	accurate	Adjective	4	2
427	607	assembly	Noun	5	6
428	608	acceptance	Noun	4	7
429	609	stress	Verb	2	3
430	610	stable	Adjective	5	5
431	611	guideline	Noun	2	3
432	612	justify	Verb	2	3
433	613	profile	Noun	4	5
434	615	attribute	Verb	3	2
435	618	exclude	Verb	4	5
436	619	advance	Verb	5	12
437	621	sustain	Verb	3	7
438	623	mutual	Adjective	3	2
439	624	radical	Adjective	2	5
440	626	virtue	Noun	3	4
441	627	equation	Noun	2	3
442	628	evident	Adjective	2	2
443	634	highlight	Verb	2	2
444	638	input	Noun	2	4
445	640	functional	Adjective	2	6
446	643	distribute	Verb	3	10
447	645	scheme	Noun	2	5
448	646	ethics	Noun	3	2
449	648	exceed	Verb	2	3
450	651	format	Noun	2	2
451	653	convert	Verb	5	11
452	655	innovation	Noun	2	3
453	656	obligation	Noun	2	5
454	658	inquiry	Noun	3	3
455	659	output	Noun	2	5
456	662	extension	Noun	6	12



457	664	guide	Verb	6	5
458	666	prominent	Adjective	2	2
459	668	logic	Noun	3	5
460	669	maintenance	Noun	3	5
461	673	constraint	Noun	2	3
462	674	dynamic	Adjective	3	3
463	675	circuit	Noun	2	7
464	676	ritual	Noun	2	3
465	679	workshop	Noun	2	2
466	682	discrimination	Noun	3	2
467	685	restriction	Noun	2	3
468	687	compose	Verb	3	6
469	688	autonomy	Noun	2	2
470	690	aim	Noun	2	4
471	691	accuracy	Noun	2	2
472	692	acceptable	Adjective	3	4
473	694	restrict	Verb	4	4
474	695	conception	Noun	2	4
475	699	province	Noun	3	2
476	701	minimize	Verb	3	3
477	702	spectrum	Noun	2	2
478	705	realm	Noun	2	3
479	706	absolute	Adjective	6	5
480	707	array	Noun	2	4
481	712	separation	Noun	3	9
482	713	concern	Verb	4	2
483	714	abstract	Adjective	3	3
484	715	foster	Verb	2	3
485	718	continuous	Adjective	3	2
486	720	colony	Noun	3	6
487	721	undermine	Verb	3	2
488	723	superior	Adjective	5	7
489	724	civilization	Noun	2	4
490	726	transmission	Noun	3	5
491	731	inventory	Noun	2	5
492	733	occupation	Noun	3	5
493	737	underlying	Adjective	2	3
494	739	encounter	Noun	2	4
495	742	readily	Adverb	2	2
496	743	quantity	Noun	3	3
497	751	reliable	Adjective	2	3
498	756	probability	Noun	2	2
499	759	informal	Adjective	4	4
500	760	aggression	Noun	2	5
501	761	acquisition	Noun	3	4
502	764	similarity	Noun	2	2

503	767	govern	Verb	2	4
504	769	integrity	Noun	2	2
505	772	specify	Verb	2	7
506	773	likewise	Adverb	2	3
507	774	correspond	Verb	2	4
508	775	sphere	Noun	2	7
509	777	linear	Adjective	3	5
510	779	model	Verb	6	6
511	781	explicit	Adjective	2	2
512	787	virtual	Adjective	2	2
513	788	necessity	Noun	3	2
514	789	hierarchy	Noun	2	2
515	790	obstacle	Noun	2	2
516	791	import	Noun	2	5
517	793	productive	Adjective	2	4
518	795	progressive	Adjective	2	6
519	796	conversion	Noun	2	8
520	799	label	Verb	2	5
521	800	articulate	Verb	2	5
522	804	profound	Adjective	2	6
523	806	ideal	Noun	2	2
524	807	rational	Adjective	2	4
525	809	value	Verb	2	5
526	813	objective	Adjective	2	4
527	814	sum	Noun	4	6
528	816	consistency	Noun	2	4
529	818	representative	Adjective	2	3
530	819	positively	Adverb	2	2
531	820	survey	Verb	4	6
532	823	comparable	Adjective	2	2
533	824	revolutionary	Adjective	2	4
534	826	isolate	Verb	5	4
535	830	promotion	Noun	2	4
536	831	residential	Adjective	2	2
537	834	adaptation	Noun	2	3
538	836	neutral	Adjective	5	7
539	840	precise	Adjective	2	2
540	842	flexible	Adjective	2	5
541	843	conceive	Verb	4	3
542	845	persist	Verb	2	3
543	846	valid	Adjective	3	2
544	847	embody	Verb	2	3
545	848	stimulate	Verb	3	7
546	853	diminish	Verb	2	2
547	857	comparative	Adjective	3	2
548	858	partial	Adjective	3	3

549	859	fixed	Adjective	3	4
550	860	matrix	Noun	2	6
551	861	innovative	Adjective	2	2
552	863	voluntary	Adjective	3	2
553	864	induce	Verb	2	5
554	871	outline	Verb	2	3
555	872	appreciation	Noun	4	5
556	874	favorable	Adjective	3	4
557	876	morality	Noun	3	2
558	878	philosopher	Noun	2	2
559	880	manipulate	Verb	4	6
560	881	tolerance	Noun	2	5
561	887	accordingly	Adverb	2	2
562	893	logical	Adjective	3	4
563	896	exploit	Verb	4	3
564	897	inadequate	Adjective	2	2
565	898	desirable	Adjective	2	2
566	899	dependence	Noun	3	2
567	903	reside	Verb	2	3
568	906	transport	Noun	3	6
569	908	formulate	Verb	2	4
570	911	aid	Verb	3	2
571	916	frontier	Noun	3	3
572	917	accessible	Adjective	3	4
573	928	exclusion	Noun	2	4
574	930	projection	Noun	2	10
575	933	stance	Noun	2	2
576	935	transmit	Verb	3	4
577	937	governance	Noun	2	2
578	938	notably	Adverb	2	2
579	942	configuration	Noun	2	2
580	944	center	Verb	2	3
581	945	intensive	Adjective	2	3
582	949	regard	Noun	2	7
583	951	attachment	Noun	3	7
584	952	indirect	Adjective	3	5
585	953	interactive	Adjective	2	2
586	954	invoke	Verb	3	3
587	955	differentiate	Verb	2	6
588	958	reproduce	Verb	4	4
589	959	coordinate	Verb	3	4
590	960	revision	Noun	2	3
591	961	passive	Adjective	3	3
592	964	isolated	Adjective	3	6
593	968	modeling	Noun	7	3
594	969	thesis	Noun	2	2

595	971	occurrence	Noun	2	2
596	973	authentic	Adjective	3	2
597	976	marginal	Adjective	3	4
598	978	designate	Verb	3	5
599	979	correspondence	Noun	3	5
600	980	maximize	Verb	2	2
601	981	implicit	Adjective	3	2
602	984	erosion	Noun	3	4
603	985	striking	Adjective	2	2
604	988	invention	Noun	4	3
605	989	fraction	Noun	2	3
606	994	disturbance	Noun	3	7
607	999	large-scale	Adjective	2	2
608	1006	reconstruction	Noun	2	4
609	1009	contrast	Verb	2	2
610	1011	export	Verb	2	3
611	1014	usage	Noun	2	3
612	1015	persistent	Adjective	2	4
613	1019	allocation	Noun	2	3
614	1020	compensate	Verb	3	6
615	1021	coordination	Noun	2	4
616	1023	formulation	Noun	2	3
617	1024	ironically	Adverb	2	2
618	1025	progress	Verb	2	3
619	1028	operational	Adjective	2	4
620	1031	revise	Verb	3	2
621	1034	parallel	Adjective	2	2
622	1040	exert	Verb	2	3
623	1042	absent	Adjective	2	3
624	1044	hostility	Noun	2	4
625	1045	embed	Verb	2	2
626	1046	extract	Verb	6	8
627	1051	suppress	Verb	4	5
628	1054	disagreement	Noun	2	3
629	1057	neglect	Verb	2	4
630	1060	territorial	Adjective	2	3
631	1065	lesser	Adjective	2	2
632	1067	cultivate	Verb	2	4
633	1069	precede	Verb	2	5
634	1074	selective	Adjective	2	2
635	1075	pathway	Noun	2	2
636	1076	autonomous	Adjective	2	3
637	1077	confine	Verb	2	5
638	1080	fulfill	Verb	2	3
639	1083	elicit	Verb	2	3
640	1089	parallel	Noun	2	3

641	1091	synthesis	Noun	2	3
642	1094	underscore	Verb	2	2
643	1105	linkage	Noun	2	4
644	1107	mobilize	Verb	2	4
645	1110	viewpoint	Noun	2	2
646	1111	import	Verb	2	3
647	1119	provincial	Adjective	2	2
648	1126	collaborate	Verb	2	2
649	1129	interference	Noun	2	5
650	1131	verify	Verb	2	4
651	1133	conform	Verb	2	2
652	1136	advancement	Noun	2	3
653	1137	validate	Verb	2	4
654	1143	coherent	Adjective	2	4
655	1144	intermediate	Adjective	2	2
656	1145	empower	Verb	2	2
657	1149	compatible	Adjective	2	3
658	1150	monopoly	Noun	2	3
659	1157	uniform	Adjective	2	4
660	1158	generator	Noun	2	4
661	1160	generic	Adjective	2	3
662	1163	degradation	Noun	2	2
663	1166	maternal	Adjective	2	4
664	1169	primitive	Adjective	2	4
665	1173	coincide	Verb	2	3
666	1177	dependency	Noun	2	3
667	1191	appropriation	Noun	2	3
668	1200	refine	Verb	2	6
669	1202	inconsistent	Adjective	3	3
670	1205	substitute	Noun	2	3
671	1208	apparatus	Noun	2	2
672	1209	thorough	Adjective	2	3
673	1210	robust	Adjective	2	4
674	1212	plausible	Adjective	2	2
675	1214	maturity	Noun	2	3
676	1220	adhere	Verb	2	6
677	1223	elaborate	Verb	2	4
678	1224	focal	Adjective	2	2
679	1226	distortion	Noun	2	6
680	1227	qualification	Noun	4	3
681	1231	alignment	Noun	2	4
682	1233	migrate	Verb	2	2
683	1236	rigorous	Adjective	2	2
684	1239	preclude	Verb	2	2
685	1243	inferior	Adjective	2	6
686	1245	multiply	Verb	2	4

687	1249	enlightenment	Noun	2	3
688	1252	Greek	Noun	2	2
689	1255	reconcile	Verb	3	4
690	1257	straightforward	Adjective	2	4
691	1259	privileged	Adjective	2	3
692	1263	complementary	Adjective	2	2
693	1264	susceptible	Adjective	2	2
694	1266	succession	Noun	2	5
695	1268	decisive	Adjective	2	4
696	1270	spontaneous	Adjective	2	2
697	1277	contention	Noun	2	3
698	1287	accumulation	Noun	2	4
699	1298	displace	Verb	2	4
700	1299	probe	Verb	2	2
701	1300	broaden	Verb	2	4
702	1306	reconstruct	Verb	2	5
703	1309	definitive	Adjective	2	3
704	1313	resonance	Noun	2	5
705	1315	contradict	Verb	2	4
706	1316	unstable	Adjective	2	6
707	1319	appendix	Noun	2	2
708	1320	infer	Verb	2	5
709	1321	distort	Verb	2	5
710	1322	manual	Adjective	2	3
711	1326	confer	Verb	2	2
712	1330	reversal	Noun	2	8
713	1335	imperative	Noun	2	2
714	1338	infinite	Adjective	2	4
715	1342	probable	Adjective	2	2
716	1343	consolidate	Verb	2	5
717	1345	unrelated	Adjective	2	2
718	1347	acknowledgment	Noun	2	3
719	1348	discriminate	Verb	2	3
720	1349	denote	Verb	2	3
721	1350	hinder	Verb	2	3
722	1351	friction	Noun	2	3
723	1354	equilibrium	Noun	2	4
724	1355	mandate	Verb	2	3
725	1370	erode	Verb	2	2
726	1384	fusion	Noun	2	7
727	1390	presently	Adverb	2	2
728	1391	authoritative	Adjective	2	3
729	1399	alteration	Noun	2	3
730	1400	adult	Adjective	2	2
731	1415	collaborator	Noun	2	3
732	1421	contour	Noun	2	3

733	1425	allowance	Noun	3	6
734	1427	citation	Noun	3	6
735	1434	disperse	Verb	2	6
736	1436	unsuccessful	Adjective	2	2
737	1438	transparency	Noun	2	3
738	1439	inheritance	Noun	2	4
739	1441	catalyst	Noun	2	2
740	1442	subordinate	Adjective	2	3
741	1444	conserve	Verb	2	4
742	1450	facet	Noun	2	2
743	1456	preoccupation	Noun	2	3
744	1462	assimilate	Verb	2	5
745	1463	enlarge	Verb	2	4
746	1466	delineate	Verb	2	5
747	1482	estimation	Noun	2	4
748	1485	portrayal	Noun	3	5
749	1497	richness	Noun	2	7
750	1501	bridge	Verb	2	3
751	1502	individualism	Noun	2	3
752	1505	experimentation	Noun	2	2
753	1507	continuation	Noun	2	4
754	1509	disparate	Adjective	2	2
755	1515	degrade	Verb	2	3
756	1516	incompatible	Adjective	2	9
757	1525	nominal	Adjective	2	6
758	1528	substitution	Noun	2	2
759	1529	novelty	Noun	2	4
760	1534	containment	Noun	2	3
761	1540	directory	Noun	2	2
762	1556	restricted	Adjective	2	3
763	1557	synthesize	Verb	2	2
764	1559	customary	Adjective	2	2
765	1571	emanate	Verb	2	2
766	1572	breadth	Noun	2	2
767	1575	unequal	Adjective	2	2
768	1577	multitude	Noun	2	3
769	1578	irregular	Adjective	3	9
770	1580	conformity	Noun	2	5
771	1591	inconsistency	Noun	2	2
772	1612	sharing	Noun	2	4
773	1614	continual	Adjective	2	2
774	1619	capitalist	Noun	2	2
775	1628	amplify	Verb	2	4
776	1629	outweigh	Verb	2	2
777	1634	intellect	Noun	2	3
778	1637	speculative	Adjective	2	3

779	1640	persuasion	Noun	2	2
780	1641	abandonment	Noun	2	3
781	1644	endow	Verb	2	2
782	1648	intrusion	Noun	2	5
783	1665	two-way	Adjective	2	3
784	1666	measurable	Adjective	2	2
785	1688	underline	Verb	2	2
786	1700	refinement	Noun	2	5
787	1701	geometric	Adjective	2	2
788	1703	inadequacy	Noun	2	3
789	1713	lecturer	Noun	2	2
790	1718	completed	Adjective	4	3
791	1722	permeate	Verb	2	3
792	1723	dissolution	Noun	2	5
793	1729	receptive	Adjective	2	4
794	1733	omission	Noun	2	4
795	1739	further	Verb	5	2
796	1741	impart	Verb	2	3
797	1742	impediment	Noun	2	2
798	1750	endemic	Adjective	2	3
799	1761	concentrated	Adjective	2	5
800	1762	suggestive	Adjective	2	3
801	1764	stratum	Noun	2	4
802	1766	bonding	Noun	3	3
803	1787	repetitive	Adjective	2	2
804	1812	unfavorable	Adjective	2	3
805	1814	propagate	Verb	2	8
806	1819	transcription	Noun	2	5
807	1820	diffuse	Verb	3	3
808	1836	neutralize	Verb	2	6
809	1837	refute	Verb	2	2
810	1843	conditional	Adjective	2	2
811	1851	infancy	Noun	2	2
812	1855	approximation	Noun	2	4
813	1865	parochial	Adjective	2	2
814	1880	ordering	Noun	3	2
815	1889	separated	Adjective	2	4
816	1891	expulsion	Noun	2	3
817	1894	immersion	Noun	2	5
818	1908	triad	Noun	2	4
819	1912	pathological	Adjective	2	3
820	1914	negate	Verb	2	4
821	1915	conjunction	Noun	2	6
822	1923	elective	Adjective	2	2
823	1927	retrieval	Noun	2	3
824	1932	predominate	Verb	2	2



825	1940	diverge	Verb	2	4
826	1950	intertwine	Verb	2	3
827	1952	virtuous	Adjective	2	2
828	1955	enlargement	Noun	2	4
829	1963	topography	Noun	2	2
830	1970	impersonal	Adjective	2	2
831	1977	normalize	Verb	2	3
832	1981	outward	Adjective	2	2
833	1993	overestimate	Verb	2	2
834	1995	maturation	Noun	2	3
835	2003	utilitarian	Adjective	2	2
836	2005	colonize	Verb	3	2
837	2029	approximate	Adjective	2	3
838	2031	increment	Noun	2	2
839	2032	rudimentary	Adjective	2	3
840	2034	expressed	Adjective	3	2
841	2035	paternal	Adjective	2	4
842	2039	material	Adjective	2	6
843	2040	potency	Noun	2	4
844	2065	inertia	Noun	2	2
845	2069	compulsion	Noun	2	3
846	2083	index	Verb	2	3
847	2087	repress	Verb	3	5
848	2088	redundant	Adjective	2	2
849	2093	inaccessible	Adjective	3	2
850	2100	predisposition	Noun	2	3
851	2103	directional	Adjective	2	3
852	2106	rupture	Noun	2	3
853	2116	replete	Adjective	2	2
854	2117	devalue	Verb	2	3
855	2119	auxiliary	Adjective	2	2
856	2132	condense	Verb	2	7
857	2133	unskilled	Adjective	2	3
858	2134	monolithic	Adjective	2	2
859	2160	hereditary	Adjective	2	2
860	2166	unaffected	Adjective	2	4
861	2185	doubling	Noun	2	2
862	2187	harmonize	Verb	3	6
863	2189	inward	Adjective	2	2
864	2196	impractical	Adjective	2	2
865	2210	unrestricted	Adjective	2	5
866	2226	hereafter	Adverb	2	3
867	2229	generality	Noun	2	2
868	2231	dynamism	Noun	2	3
869	2235	indefinite	Adjective	2	2
870	2237	prompt	Adjective	2	3

871	2240	collected	Adjective	2	2
872	2251	expert	Adjective	3	2
873	2257	redundancy	Noun	2	4
874	2264	modulate	Verb	2	5
875	2271	recast	Verb	2	3
876	2303	persistently	Adverb	2	2
877	2318	drafting	Noun	4	3
878	2319	credence	Noun	2	2
879	2330	shaping	Noun	4	2
880	2333	predominance	Noun	2	2
881	2355	crystallize	Verb	2	4
882	2357	unrecognized	Adjective	4	2
883	2362	regenerate	Verb	2	9
884	2363	allowable	Adjective	2	3
885	2369	framing	Noun	5	2
886	2371	partition	Verb	2	2
887	2378	corrosive	Adjective	2	2
888	2406	dictum	Noun	2	2
889	2407	directed	Adjective	5	2
890	2414	monopolize	Verb	2	2
891	2415	peculiarity	Noun	2	3
892	2434	bottleneck	Noun	2	2
893	2449	invalid	Adjective	2	2
894	2468	reversible	Adjective	2	4
895	2469	spurious	Adjective	2	4
896	2472	equivocal	Adjective	2	3
897	2487	simplification	Noun	2	3
898	2495	inflexible	Adjective	2	4
899	2504	innermost	Adjective	2	2
900	2510	supposition	Noun	2	3
901	2585	distinguishable	Adjective	2	2
902	2610	stricture	Noun	2	2
903	2650	dissociate	Verb	2	3
904	2683	malleable	Adjective	2	2
905	2710	dispensation	Noun	2	3
906	2723	inhospitable	Adjective	2	2
907	2755	disentangle	Verb	3	5
908	2788	opacity	Noun	2	3
909	2827	constancy	Noun	2	3
910	2848	unacknowledged	Adjective	2	2
911	2850	unsound	Adjective	4	6
912	2862	unsupported	Adjective	3	2
913	2885	subservient	Adjective	2	3
914	2953	entangled	Adjective	2	3
915	2956	disposed	Adjective	2	2
916	2965	modelling	Noun	7	2

917	2978	arguable	Adjective	2	2
918	2998	reversion	Noun	2	6
919	3014	tangential	Adjective	2	2

Appendix 2. Meaning definitions of *study* (n) in WordNet and *COBUILD*

No	WordNet	<i>COBUILD</i>
1	Survey; a detailed critical inspection	Study is the activity of studying.
2	Work; applying the mind to learning and understanding a subject (especially by reading); " <i>no schools offer graduate study in interior design</i> "	A study of a subject is a piece of research on it.
3	Report, a written document describing the findings of some individual or group; " <i>this accords with the recent study by Hill and Dale</i> "	You can refer to educational subjects or courses that contain several elements as studies of a particular kind.
4	A state of deep mental absorption; " <i>she is in a deep study</i> "	A study is a room in a house which is used for reading, writing, and studying.
5	A room used for reading and writing and studying; " <i>he knocked lightly on the closed door of the study</i> "	
6	Discipline/field/subject area; a branch of knowledge; " <i>anthropology is the study of human beings</i> "	
7	Sketch; preliminary drawing for later elaboration; " <i>he made several studies before starting to paint</i> "	
8	Cogitation; attentive consideration and meditation; " <i>after much cogitation he rejected the offer</i> "	
9	Someone who memorizes quickly and easily (as the lines for a part in a play); " <i>he is a quick study</i> "	
10	A composition intended to develop one aspect of the performer's technique; " <i>a study in spiccato bowing</i> "	

Supplementary materials: AVL homographs

**AVL homographs**

AVL ID number	AVL item	POS	In <i>COBUILD</i> ?	In WordNet 3.1?
20	policy	noun	Yes	Yes
124	scale	noun	Yes	Yes
167	affect	verb	Yes	Yes
204	test	verb	Yes	Yes
206	tend	verb	Yes	Yes
234	express	verb	Yes	Yes
290	content	noun	Yes	Yes
479	passage	noun	Yes	Yes
481	discovery	noun	Yes	Yes
553	process	verb	Yes	Yes
681	undertake	verb	Yes	Yes
684	graph	noun	Yes	Yes
866	stem	verb	Yes	Yes
890	temporal	adjective	Yes	Yes
915	manifest	verb	Yes	Yes
929	axis	noun	Yes	Yes
948	cue	noun	Yes	Yes
1483	Dominican	adjective	No	Yes
1687	undertaking	noun	Yes	Yes
1745	rigor	noun	Yes	Yes
1845	lag	noun	Yes	Yes
1900	primer	noun	Yes	Yes
2082	MP	noun	No	Yes
2165	predate	verb	Yes	Yes
2182	apex	noun	Yes	Yes
2470	sup	verb	Yes	Yes
2631	forego	verb	Yes	Yes
2888	card	verb	Yes	Yes

Supplementary materials: Figure A

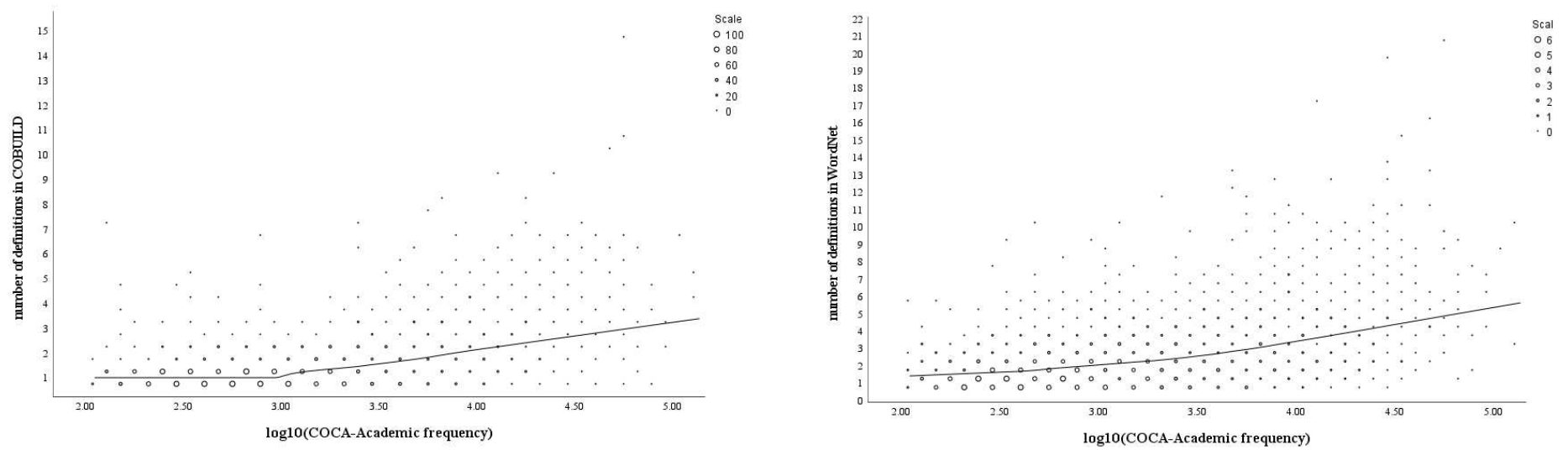
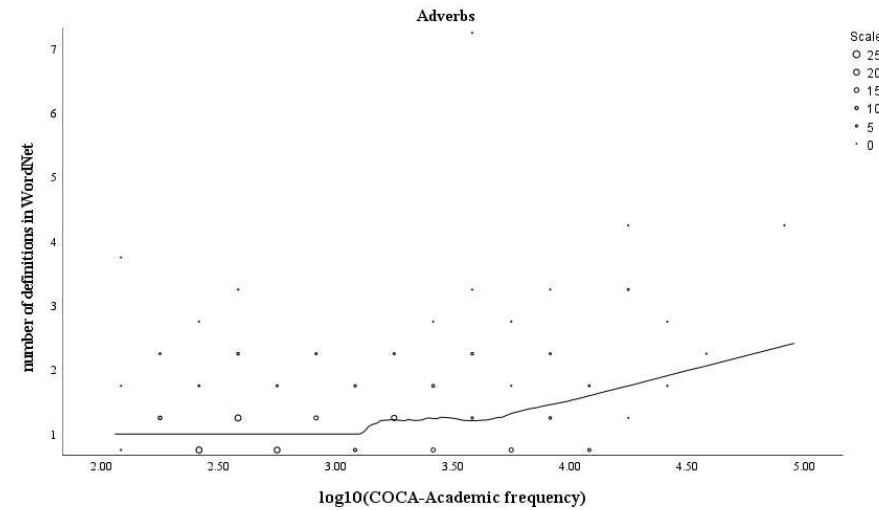
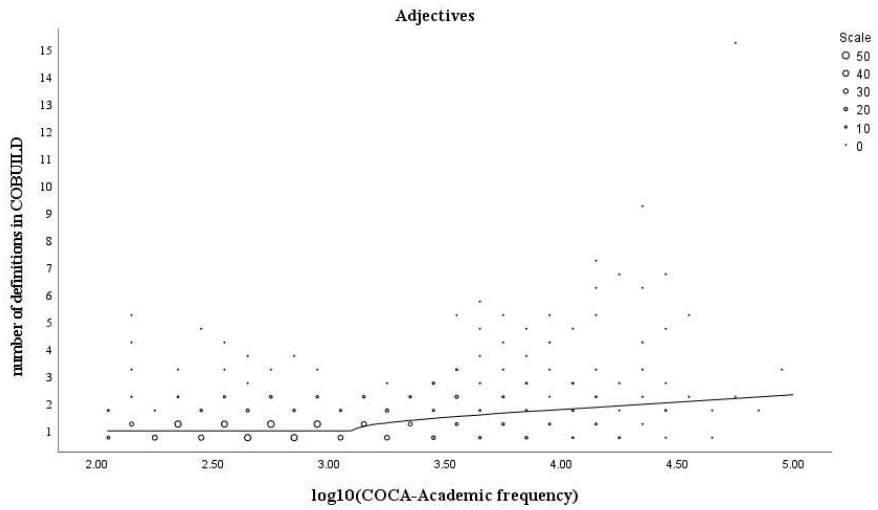
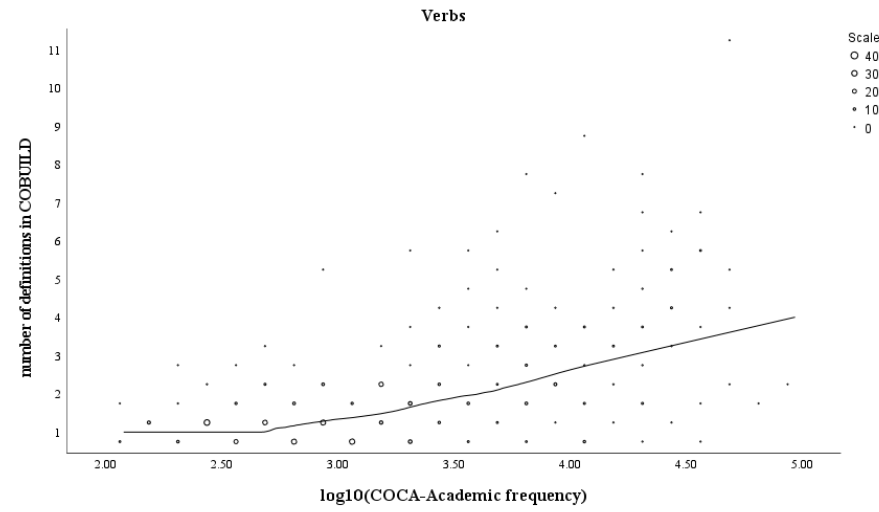
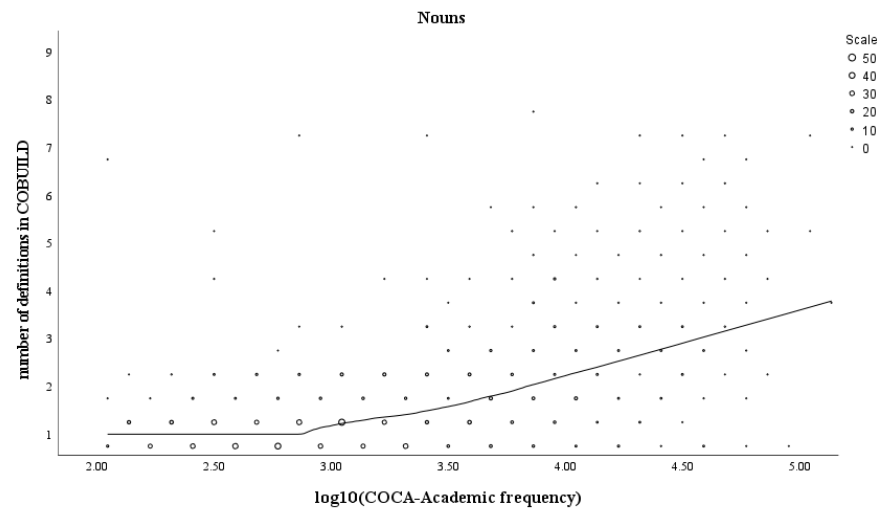


Figure A. Scatterplots of  $\log_{10}(\text{COCA-Academic frequency})$  vs number of definitions in COBUILD and WordNet. The black line is the Loess line. Lemmas are binned to improve legibility.

Supplementary materials: Figure B



**Figure B.** Scatterplots of  $\log_{10}(\text{COCA-Academic frequency})$  vs number of definitions in COBUILD per POS. The black line is the Loess line. Lemmas are binned to improve legibility.



Supplementary materials: Figure C

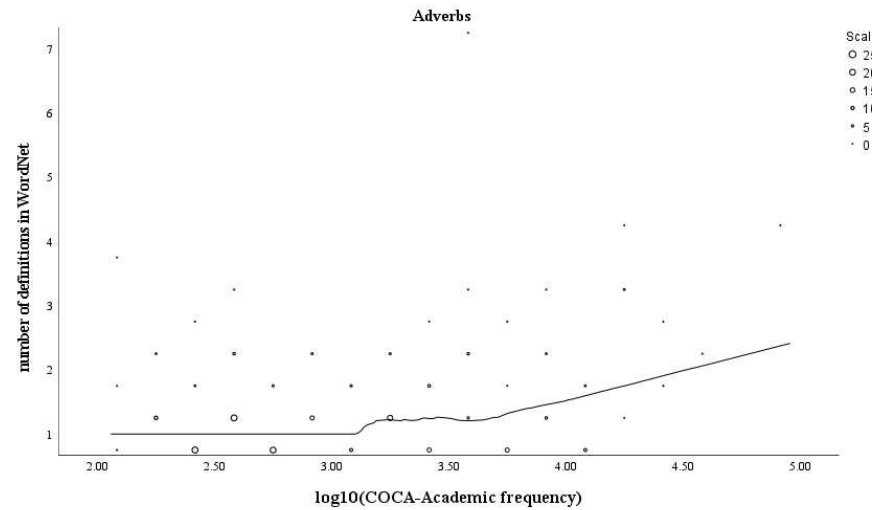
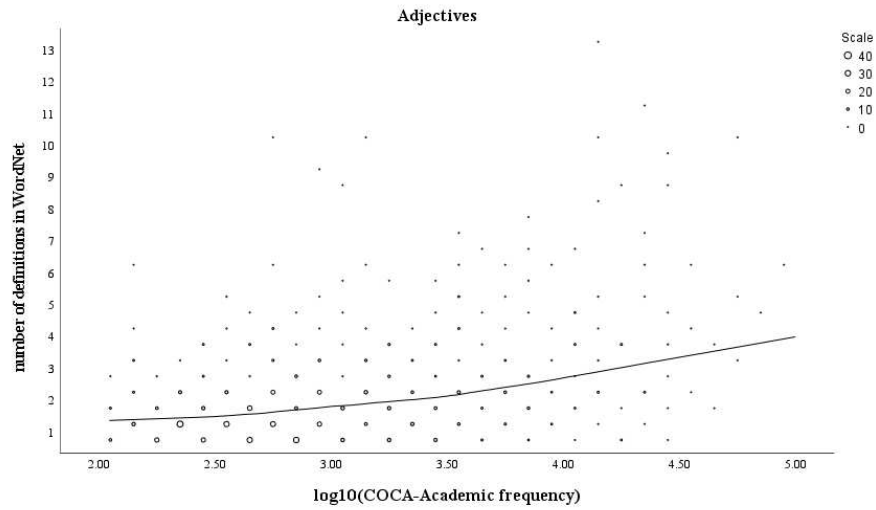
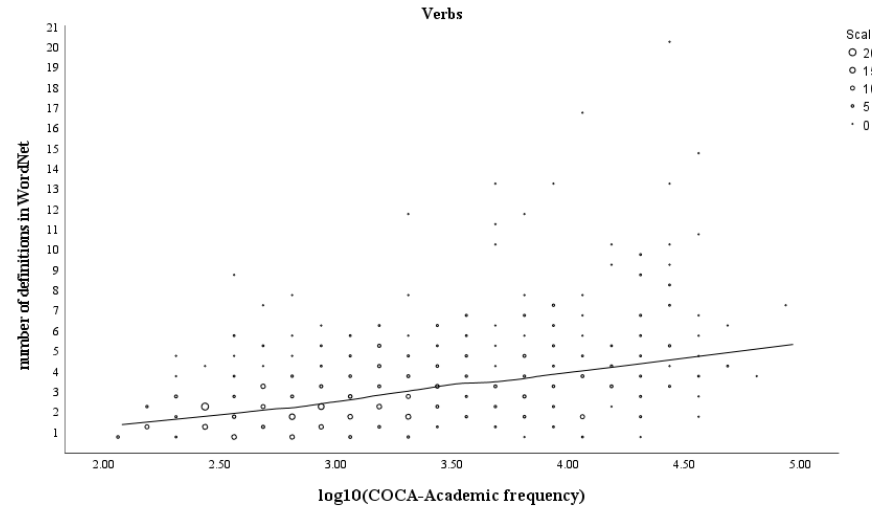
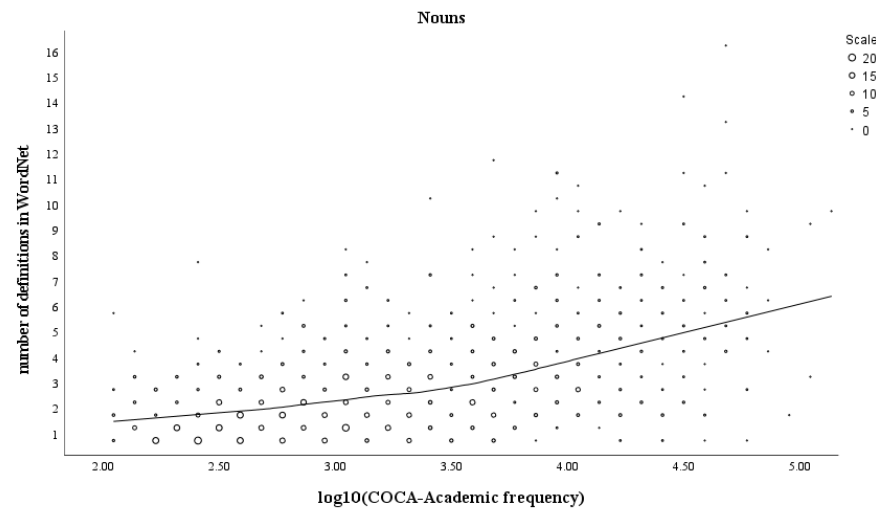


Figure C. Scatterplots of  $\log_{10}(\text{COCA-Academic frequency})$  vs number of definitions in WordNet per POS. The black line is the Loess line. Lemmas are binned to improve legibility.