

A Corpus Analysis of Polysemy in CEFR-based English Textbooks

Hicham Lahlou¹

¹ Universiti Sains Malaysia

Funding: Universiti Sains Malaysia Short Term Research Grant

Potential competing interests: No potential competing interests to declare.

Abstract

Despite the considerable number of studies on vocabulary and sense relations such as polysemy, they are confined to comparing vocabulary learning methods and strategies and examining the usefulness of certain sense relations in acquiring L2 vocabulary. Until now, no attempt has been made to analyse polysemy in English textbooks used in non-native English contexts. The aim of the current article was to explore the high usage of the most frequently used words from the selected English textbooks in real-life English and to determine the extent to which polysemy is incorporated in the selected English textbooks. The data was collected from two textbooks: *Close-up* (Intermediate Secondary) and *Full Blast Plus 4* (Upper Secondary). The *Sketch Engine* word list tool was used to determine the top 100 most frequently used words in the first sub-corpus, *Close-up* (Intermediate Secondary), and the top 100 most frequently used words in the second sub-corpus, *Full Blast Plus 4* (Upper Secondary). The use of the chosen words in the *British National Corpus* (BNC) was explored using the *Sketch Engine* to determine whether they are the most commonly used in English. *WordNet* was also used to determine the polysemy and frequency of the words under study. The findings revealed the top 100 most frequently used words in *Close-up* and *Full Blast Plus 4* have high frequency levels (ranks) in the *BNC*, yet 65 of the top 100 most frequently used words in both *Close-up* and *Full Blast Plus 4* overlapped despite the level difference. Furthermore, polysemy is only lightly emphasised in the textbooks concerned. The results of the present study have important implications for second and foreign language learning research and education.

Hicham Lahlou*

School of Humanities, Universiti Sains Malaysia, 11800 USM Pulau Pinang, Malaysia

*Correspondence: hicham@usm.my

Keywords: English, L2, vocabulary, polysemy, corpus, central senses, peripheral senses.

Introduction

Humans' encyclopedic or background knowledge is necessary to understand information, concepts and word senses. If a person has no background knowledge of a topic, it will be difficult for them to make sense of it. This prior knowledge is the outcome of a person's physical, social, and cultural experiences. Given that language mirrors reality as humans construe it (Cienki, 2007), a word is an access point to encyclopedic knowledge rather than a container of meaning (Langacker, 1987; Kecskes, 2013). Accordingly, Langacker (1987) asserts that "the entity designated by a symbolic unit can...be thought of as a point of access to a [cognitive] network" (p. 163). When a learner receives a lexical item, their prior knowledge is activated, and so the word evokes all the ideas and thoughts associated with it in the learner's mind. For a better acquisition of a second language (L2), a learner should be mindful of the word associations produced by a native speaker given the possible differences between the word associations of native and non-native speakers. At the same time, the texts used in teaching English should include concepts based on learners' cultural background. This is to make sure they understand them accurately given their prior knowledge and are cognizant of the differences between their culture and other cultures.

The literature on English language vocabulary learning shows that despite the various approaches and methods that have been proposed and implemented, learning vocabulary remains a major challenge for non-native learners of English. Learning vocabulary is a complex process, which embodies semantic, phonological, morphological, syntactic features, and so on (e.g., Kalyuga & Kalyuga, 2008; Wangru, 2016). Studies have shown that acquiring a second language includes the capability of understanding as well as generating groups of words (e.g., Willis, 2003; Kalyuga & Kalyuga, 2008; Nation, 2013). The knowledge of chunks and word multiple senses is as important as or even more important for language use than the knowledge of single words.

Learning or memorizing lists of individual words deprives learners of many experiences of language learning, including metaphorical and metonymic uses of words, and exposure to its linguistic and cultural peculiarities. If students comprehend the role of conceptual metaphor and metonymy, they will easily understand as well as memorize polysemes and idioms (Beréndi, Csábi and Kövecses, 2008; Pérez, 2017). Besides, if learners are cognizant of the relationship between source domains and target domains in metaphorical and metonymic mappings, they will find abstract concepts much easier. Learning chunks and multiple meanings of words is much easier than separate words. For more efficient learning of English, learners should be aware of the mental associations of words, the sense relations among the diverse meanings of a word, and figurative words as well as phrases. Aitchison (2012) argues that "humans do not very often deal with isolated words. We therefore need to find out how words relate to each other in the mental lexicon" (p. 89). In relation to this, much research on vocabulary has centred on multi-word units like collocations, compounds, and idioms, and sense relations like polysemy, homonymy, and synonymy. Despite the significance of these studies, they are limited to comparing vocabulary learning methods and strategies and testing the usefulness of particular chunks and sense relations in learning L2 vocabulary. To date, there has been no attempt to investigate multi-word units and word senses in English textbooks.

While textbooks are not the only resource that teachers use to deliver instruction and assist students in achieving the

intended learning outcomes, they are vital, especially in areas where resources are limited, e.g., rural areas. They are varied alternatives to spare time and money; they consist of graphics that may be more effective than educators' explanations and objects that may not be possible to bring to the classroom (González, 2006). More importantly, they are a common tool to ensure standardizing English learning and teaching in non-English speaking schools and aligning English standards in these schools with international standards. To this end, many European as well as other countries, e.g., Malaysia, Vietnam, and Thailand, for instance, implemented the Common European Framework of Reference (CEFR), an international standard that helps enhance as well as measure learners' English language skills (e.g., Don et al., 2015; Nguyen, 2015). Nguyen (2015) describes CEFR as an ideal standard to embrace. CEFR-based English textbooks for non-English-speaking students, therefore, constitute one of the most important tools as well as sources of knowledge that need an investigation to determine the extent of the incorporation of multi-word units and word senses.

Polysemy, the focus of the current study, is a situation in which a form has multiple meanings that are related (Schmitt, 2010). For example, the word *party* has two different meanings: 'a social event' and 'a political organization'. The current study focuses on polysemy because it is pervasive in the English language and is a vital component, which enhances non-native speakers' use of their second or foreign language. Based on some previous studies, more than 40% of English words have multiple meanings (e.g., Durkin & Manning, 1989; Abou-Khalil, et al. (2019); Shahbazi & Byun, 2020). Importantly, these polysemous lexical units are among the three thousand most frequently used lexical items in English (e.g., Makni, 2013; Ferrer-i-Cancho, 2014; Alnamer, 2017). Polysemy is a crucial aspect of vocabulary, and so it needs to be taken into account in research on vocabulary. The most prototypical senses of words are usually the most frequently used. A study of polysemy does not only incorporate prototypical meanings, but peripheral meanings as well. Thus, a frequency analysis should be on both lexical items and their multiple meanings (Schmitt, 2010).

The current study aims at examining the emphasis on polysemy in English textbooks for non-native students. Overall, the objectives of the current study are:

1. To explore the high usage of the most frequently used words from the selected English textbooks in real-life English.
2. To determine the extent to which polysemy is incorporated in the selected English textbooks.

The findings of the current study have important implications for second language and foreign language education. The aim of the present study is to further broaden current knowledge of one of the major aspects of vocabulary, polysemy, which can be a major obstacle to learning vocabulary and increasing comprehension if not mastered. The study will provide considerable insight into the status of polysemy in the English textbooks used in non-native English countries like Malaysia, Vietnam, and Thailand. The results found about the extent of the incorporation of polysemy in such textbooks will suggest several courses of action to alleviate or rather solve the problem of vocabulary acquisition in in Southeast Asia in particular and all non-English-speaking countries in general.

Literature Review

A polyseme, compared with a monoseme, has multiple meanings. And these meanings are systematically related (Lakoff,

2008; Csábi, 2004; Dölling, 2018). For instance, the word *warm* has multiple meanings: 'temperature' and 'clothing that makes one feel warm'. However, if the senses of a lexical item are not systematically related, they are homonyms. For example, the word *bank* has the meanings of 'place where one puts their money' and 'the edge of a river', which are not systematically related (Lakoff, 2008).

From a cognitive linguistic perspective, polysemy is a 'radial category' of which the related senses are subcategories that exhibit family resemblance but range from prototypical to peripheral members. The most prototypical meaning or category member is the source from which the other meanings are extended (Lakoff, 2008). These semantic extensions, especially figurative senses, are usually motivated by cognitive factors, notably metaphor, metonymy, and image-schema transformation (Lakoff, 2008). Metaphor is a conventional association of one domain with another domain. For example, one of the senses of the category OVER, 'control', is metaphorically extended from another meaning of OVER, namely 'above' (Lakoff, 2008).

In contrast, metonymy is a relationship between a conceptual entity with another that is related to it, that is, domain-subdomain interaction (Lakoff & Johnson, 2003). For instance, *the ham sandwich* in *The ham sandwich is waiting for his check* refers to the person who ordered the ham sandwich (Lakoff and Johnson, 2003). Knowledge of the connection between source domains and target domains in these conceptual mappings will facilitate understanding and memorizing polysemous words (Beréndi, Csábi and Kövecses, 2008).

Previous research findings showed that polysemy constitutes one main obstacle that keeps learners from comprehending texts. Polysemy is common in English as in other languages. In English, however, it consists of a complex relationship between form and meaning, and so it causes difficulties acquiring as well as using vocabulary (e.g., Schmitt, 2010; Mitsugi, 2017). Research showed that it is easier for children to guess the meaning of non-polysemous words (e.g., Saemen, 1970; Nation, 2013). Learners can expand their knowledge of these words effortlessly and independently. Polysemous words, in contrast, are harder to guess from context, especially when the real word form is used in comparison with a nonsense word. The form that is familiar to learners may cause them to understand the meaning, but this meaning may not be appropriate in different contexts (Saemen, 1970; Nation, 2013). Therefore, polysemy is expected to be emphasized in teaching vocabulary and thus in English textbooks. Furthermore, educators' knowledge and application of semantic expansion to vocabulary teaching is critical to enhancing students' understanding and retention of word senses (Boers, 2008).

Methodology

The current paper aims to explore the frequency of the selected polysemous words in English academic texts, to identify the extent to which polysemous words are emphasized in English textbooks used in non-native English contexts. The paper also aims to explore the prototypical and peripheral senses of the most frequent polysemous words in the selected English textbooks.

The study investigates two representative English textbooks that non-English speaking secondary schools currently use and that align with CEFR, the “ideal standard”, to use Nguyen’s (2015) expression. The texts chosen are *Close-up* (Intermediate Secondary) and *Full Blast Plus 4* (Upper Secondary), published by two recognized publishers, National Geographic Learning and MM publications respectively. *Close-up* (Healan & Gormley, 2015) targets English Intermediate Level (B1). It consists of 12 units. *Full Blast Plus 4* (Mitchell & Malkogianni, 2018) targets English Intermediate Level (B1) and contains 8 units.

To gather data for the current study, reliable instrumentation was employed. To accomplish this goal, the *Sketch Engine* word list tool was used to identify the top 100 most frequently used words in the first subcorpus, *Close-up* (Intermediate Secondary), and the top 100 most frequently used words in the second subcorpus, *Full Blast Plus 4* (Upper Secondary). The use of the selected words in the British National Corpus (BNC) was investigated using *Sketch Engine* to determine whether they are the most used in English. The study also used *WordNet* to determine the polysemy and frequency of the words under study.

Results and Discussion

Based on the *Sketch Engine* Wordlist generating the frequency lists of the lemmas identified in the top 100 most frequently used words in *Close-up* (Intermediate Secondary), i.e., verbs, nouns, adjectives and adverbs, showed that their frequency level (rank) in the *BNC* ranges from 1 to 970. The data on the top 100 most frequently used words in *Full Blast Plus 4* (Upper Secondary) showed that their frequency level (rank) in the *BNC* ranges from 1 to 932. This shows the extent to which modern English textbooks, including CEFR English textbooks, adopt the trend of prioritising high-frequency words in language education. This is because frequent words are more useful than less frequent words given that they are encountered more frequently and thus have a higher chance of being learned (e.g., Vilkaitė-Lozdienė & Schmitt, 2020).

The findings on the word frequency also revealed that 65 of the top 100 most frequently used words in both *Close-up* and *Full Blast Plus 4* overlapped (see Figure 1).

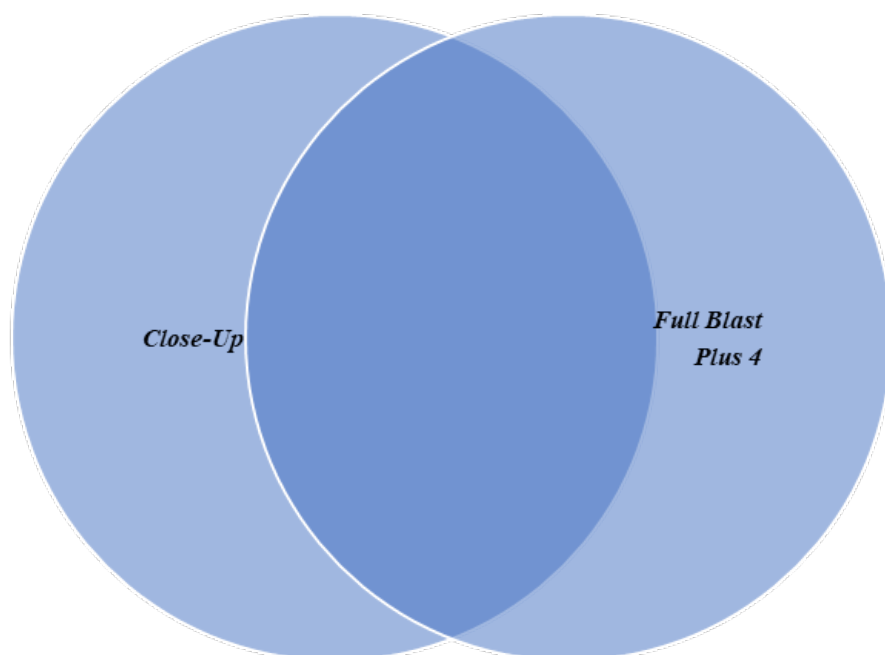


Figure 1. *The Overlapping of the Most Frequently Used Words in Close-Up and Full Blast Plus 4*

Table 1 shows in detail the words shared by *Close-up* and *Full Blast Plus 4* in the top one hundred most frequently used words from *be* to *text*, as well as the ones that differ, i.e., 35 words out of the top 100 most frequently used words. According to these findings, 65% of the top 100 most frequently used words are the same despite the level difference between *Close-up* (Intermediate Secondary) and *Full Blast Plus 4* (Upper Secondary).

Close-Up		Close-Up & Full Blast Plus 4				Full Blast Plus 4	
correct	important	be	look	thing	past	just	mean
action	eat	have	see	ask	come	woman	car
story	different	do	give	try	place	boy	game
remember	underline	use	new	present	happen	back	let
expression	long	word	now	watch	decide	show	letter
sport	house	go	year	need	learn	last	phone
match	check	read	school	find	change	art	note
describe	leave	make	work	day	next	follow	person
love	plan	people	know	play	well	discuss	computer
food	continuous	get	talk	form	way	express	page
partner	picture	think	take	other	week	part	job
teacher	complete	write	student	start	hear	really	even
listen	like	time	help	information	buy	call	home
perfect	question	say	idea	tell	great	film	opinion
family	example	very	want	friend	same	only	man
live	music	good				girl	stop
future	task	answer	first	choose	text	feel	shop
world						money	

Table 1. The Common and Different Most Frequent Words in Close-Up and Full Blast Plus 4

The findings on polysemy in the selected English textbooks showed low emphasis on the multiple meanings of words and thus contextual diversity. Figure 2 compares and contrasts the multiple meanings of the most frequent words in *Close-Up* and their actual meanings in *WordNet*.

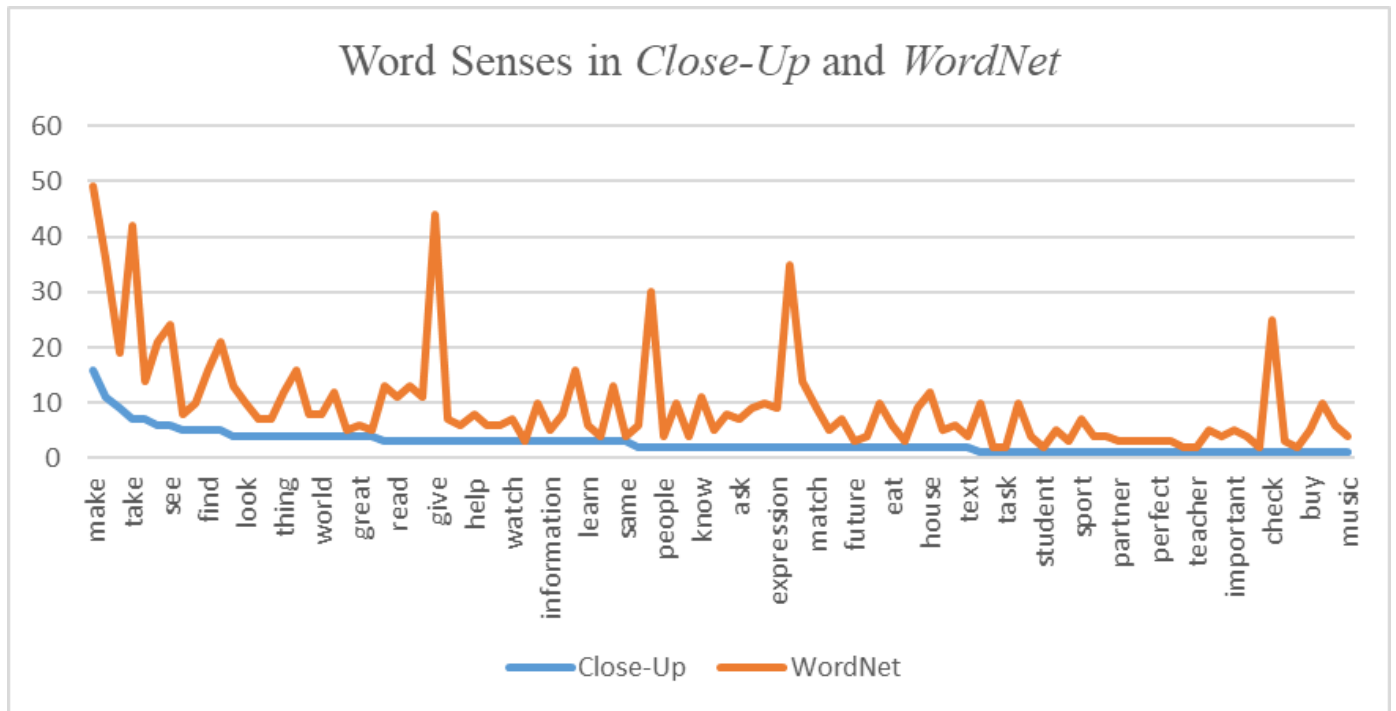


Figure 2. The Number of Word senses in *Close-Up* against *WordNet*

30 of the top 100 most frequently used words in *Close-up* were used with only one meaning though these words correspond to multiple related senses ranging from 2 to 25 meanings. It was also found that 27 of the top 100 most frequent words were used with two meanings despite their multiple meanings ranging from 3 to 35.

Figure 3 compares and contrasts the multiple meanings of the most frequent words in *Full Blast Plus 4* and their actual meanings in *WordNet*. 28 of the top 100 most frequently used words in *Full Blast Plus 4* were used with only one meaning though these words correspond to multiple related senses ranging from 2 to 11 meanings. In addition, 29 of the top 100 most frequent words were used with two meanings despite their multiple meanings ranging from 3 to 28.

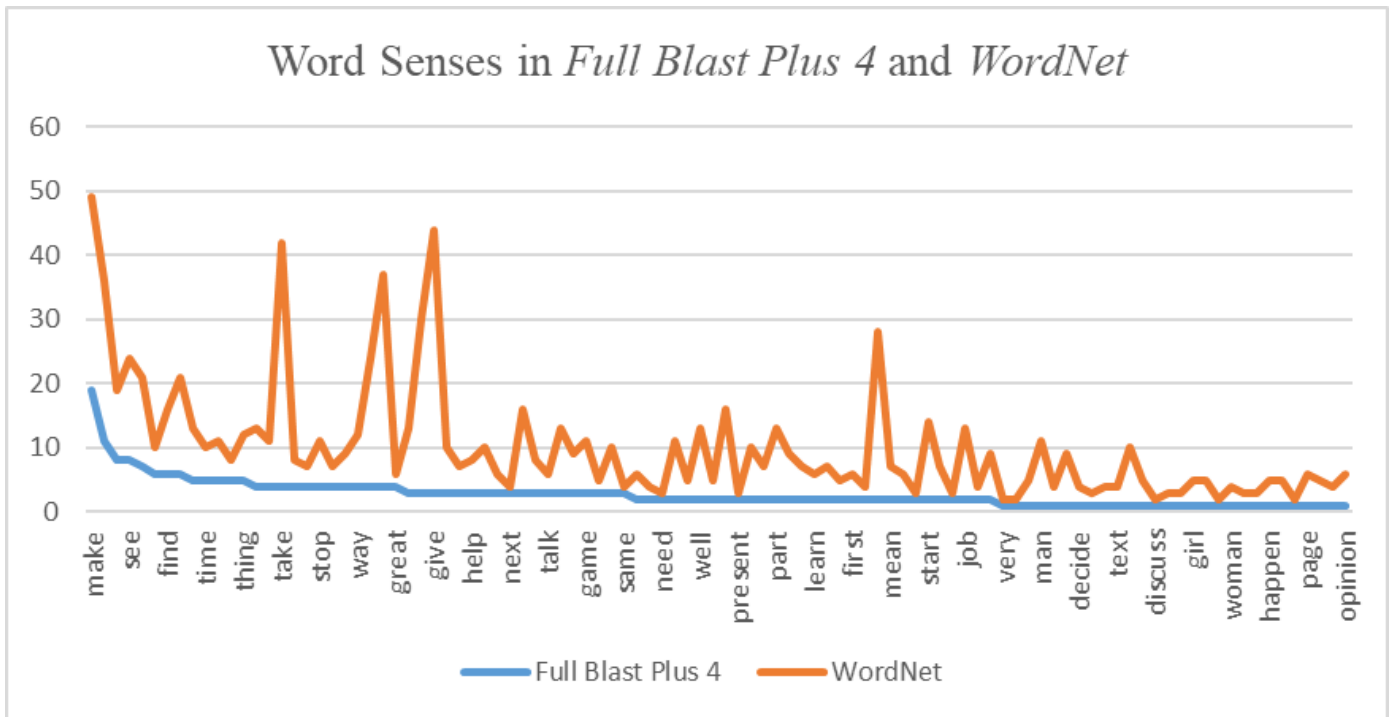


Figure 3. The Number of Word senses in Full Blast Plus 4 against WordNet

As illustrated in Figure 4, 11 of the 65 common words in *Close-up* and *Full Blast Plus 4* mentioned above were used with only one meaning. There were 25 words with the same number of senses in both textbooks, ranging from 2 to 11 senses. In contrast to *Close-up*, 17 words had an increase in the number of senses in *Full Blast Plus 4*, while 12 words had a decrease in *Full Blast Plus 4* in terms of meanings.

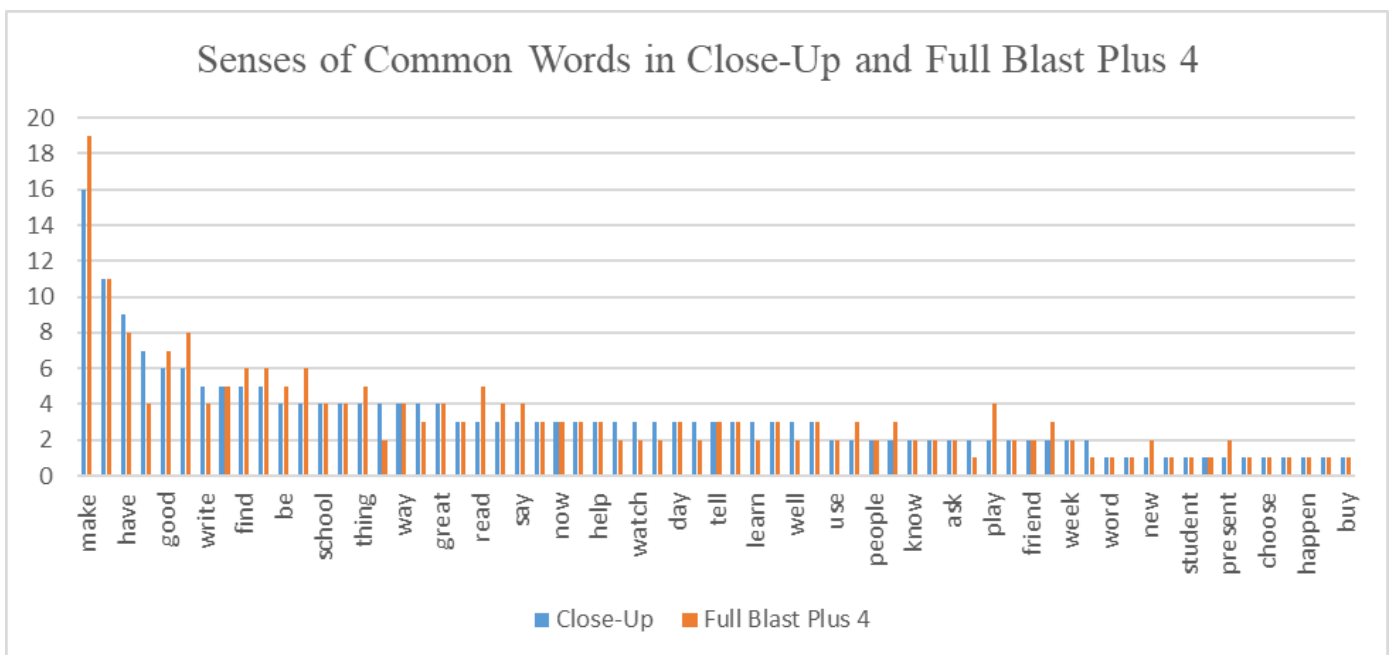


Figure 4. The Number of Senses of the Common Words in Close-up and Full Blast Plus 4

The above-mentioned 11 words used with one meaning in the selected textbooks though their actual meanings range from 2 to 10 according to *WordNet*, were used with the same meaning, the basic literal meaning. 18 of the 25 words with the same number of senses in the selected textbooks were also used with the same meanings, while only 7 words exhibited a variety of meanings and contexts. The 17 words that had an increase in the number of senses in *Full Blast Plus 4* showed a minimal increase in the number of senses, except the verb *to make* which increased by 3 meanings. Some of the 12 words that had a decrease in *Full Blast Plus 4* in terms of meaning number had positively introduced new meanings (semantic extensions), some were repetitions of the same meanings and contexts, others were repetitions of the same meanings with less semantic extensions.

Conclusion

The aim of the current study was to explore the frequency with which the most frequently used words from the selected English textbooks are used in real-life English, as well as to determine the extent to which polysemy is incorporated in the selected English textbooks. The results showed that the top 100 most frequently used words in *Close-up* and *Full Blast Plus 4* have high frequency levels (ranks) in the *BNC*, ranging from 1 to 970 and 1 to 932, respectively. However, word frequency findings revealed that 65 of the top 100 most frequently used words in both *Close-up* and *Full Blast Plus 4* overlapped despite the level difference. The findings on polysemy in the selected English textbooks revealed a lack of emphasis on multiple meanings of words and thus contextual diversity. More than the quarter of the most common words in the textbooks studied were used with only one meaning. Furthermore, despite the level difference between the intermediate and upper-intermediate levels the majority of the top 100 words with multiple meanings and contexts were used with the same meaning.

The current study's findings provide valuable insights into the use of polysemy in the selected English textbooks. They are, however, based solely on a sample of the most frequently used words. Future research on polysemy will need to expand their sampling to a larger number of frequent words as well as other English textbooks in order to obtain more comprehensive research data.

References

- Abou-Khalil, V., Helou, S., Flanagan, B., Chen, M. R. A., & Ogata, H. (2019). Learning isolated polysemous words: identifying the intended meaning of language learners in informal ubiquitous language learning environments. *Smart Learning Environments*, 6(1), 13.
- Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon* (4th ed). Chichester, UK: John Wiley & Sons.

- Alnamer, S. A. S. (2017). On the Awareness of English Polysemous Words by Arabic-Speaking EFL Learners. *Advances in Language and Literary Studies*, 8(2), 112-121.
- Beréndi, M., Csábi, S. and Kövecses, Z. (2008) Using conceptual metaphors and metonymies in vocabulary teaching. In F. Boers and S. Lindstromberg (eds). *Cognitive linguistic approaches to teaching vocabulary and phraseology*, 65–98. Berlin: Mouton de Gruyter.
- Boers, F. and Lindstromberg, S. (2008). How cognitive linguistics can foster effective vocabulary teaching. In F. Boers and S. Lindstromberg (eds.), *Cognitive linguistic approaches to teaching vocabulary and phraseology* (pp.1-61). Berlin: Mouton de Gruyter.
- Cienki, A. (2007). Frames, Idealized Cognitive Models, and Domains. In Geeraerts, D. & Cuyckens, H. (eds.). *The Oxford Handbook of Cognitive Linguistics*. New York: Oxford University Press.
- Csábi, S. (2004). A cognitive linguistic view of polysemy in English and its implications for teaching. *Cognitive linguistics, second language acquisition, and foreign language teaching*, 233, 256.
- Dölling, J. (2018). *Systematic polysemy*. The Blackwell companion to semantics.
- Don, Z. M., Abdullah, M. H., Abdullah, A. C., Lee, B. H., Kaur, K., Pillai, J., & Hooi, M. Y. (2015) *English Language Education Reform in Malaysia: The Roadmap 2015-2025*. Putrajaya: Ministry of Education.
- Durkin, K., & Manning, J. (1989). Polysemy and the subjective lexicon: Semantic relatedness and the salience of intraword senses. *Journal of Psycholinguistic Research*, 18(6), 577-612.
- Ferrer-i-Cancho, R. (2014). The meaning-frequency law in Zipfian optimization models of communication. *arXiv preprint arXiv:1409.7275*.
- González Moncada, A. (2006). On materials use training in EFL teacher education: Some reflections. *Profile Issues in Teachers Professional Development*, (7), 101-116.
- Kalyuga, M., & Kalyuga, S. (2008). Metaphor awareness in teaching vocabulary. *Language Learning Journal*, 36(2), 249-257.
- Kecskes, I. (2013). Intercultures, encyclopedia knowledge, and cultural models. *Language and intercultural communication in the new era*, 3959.
- Lakoff, G. and Johnson, M. (2003). *Metaphors We Live By*. The University of Chicago Press, Chicago. (Original Work Published 1980).
- Lakoff, G. (2008). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago and London: The University of Chicago Press.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical prerequisites* (Vol. 1). California: Stanford University Press.
- Makni, F. (2013). *Teaching Polysemous Words to Arab Learners: A Cognitive Linguistics Approach* (Doctoral dissertation, University of the West of England).
- Mitsugi, M. (2017). Schema-Based Instruction on Learning English Polysemous Words: Effects of Instruction and Learners' Perceptions. *Journal of Pan-Pacific Association of Applied Linguistics*, 21(1), 21-43.
- Nation, I.S.P. 2013. *Learning Vocabulary in Another Language*. Cambridge, UK: Cambridge University Press.
- Nguyen, N. H. (2015). *Vietnam's National Foreign Languages Project 2020: Challenges, Opportunities, and Solutions*

(pp. 62-64). English for ASEAN Integration: Policies and Practices in the Region.

- Pérez, R. G. (2017). Teaching Conceptual Metaphors to EFL Learners in the European Space of Higher Education. *European Journal of Applied Linguistics*, 5(1), 87.
- Saemen, R. A. (1970). Effects of Commonly Known Meanings on Determining Obscure Meanings of Multiple-Meaning Words in Context. Report from the Project on Individually Guided Elementary Language Arts.
- Shahbazi, Z., & Byun, Y. C. (2020). Analysis of Domain-Independent Unsupervised Text Segmentation Using LDA Topic Modeling over Social Media Contents. *International Journal of Advanced Science and Technology*, 29(6), 5993 - 6014.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Springer.
- Vilkaitė-Lozdienė, L., & Schmitt, N. (2020). Frequency as a guide for vocabulary usefulness: high-, mid-, and low-frequency words. In S. Webb (ed.), *The Routledge handbook of vocabulary studies* (pp. 81-96). London: Routledge.
- Wangru, C. (2016). Vocabulary Teaching Based on Semantic-Field. *Journal of Education and Learning*, 5(3).
- Willis, D. (2003). *Rules, patterns and words: Grammar and lexis in English language teaching* Cambridge, England: Cambridge University Press.
- Zheng, S. (2012). Studies and Suggestions on English Vocabulary Teaching and Learning. *English Language Teaching*, 5(5), 129-137.