

**Corpus Analysis of Specialised Vocabulary and Polysemy  
in the Logistics Academic Written English**

**Hongyi ZHAO**

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

University of Leeds

School of Education

January 2026

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Hongyi ZHAO to be identified as Author of this work has been asserted by Hongyi ZHAO in accordance with the Copyright, Designs and Patents Act 1988.

© 2026 The University of Leeds and Hongyi ZHAO

# Acknowledgements

My heart is full of gratitude to express my sincere thanks to those who helped, supported and guided me throughout my PhD journey. Without you, this research and thesis would not be possible to complete.

First, my deep gratitude goes to my supervisors. To my primary supervisor Prof. Alice Deignan, for your constant support and encouragement to my research direction and writings, you guided me to the exploration of word meanings. You are like a lighthouse, guiding me safely to the harbour of research, amidst the mists and frost of “unknown”. And thank you for your kindness and forgiveness to my several delays of submission when I drafted writings, and to my often-forgetting naming documents of my writings. My sincere thanks also go to my co-supervisor Dr. Thi Ngoc Yen Dang. It was you who raised my interest in vocabulary research when I was a Master student at our school couple of years ago (How time flies!). At that time I know I want to explore more on vocabulary, on English language, and on academic field. Thank you for your very detailed comments and suggestions for my writings. It is my great honour to have you both as my supervisors. I learned a lot from your efforts, your research and your experience. I learned how to do PhD works, how to be a qualified researcher, and how to be a tutor. You were, are and still will be my role models.

Second, my gratitude goes to my family members. Thank you my parents, Mr. Junfeng Zhao and Mrs. Wei Wang, for your constant support, on both economic and mental aspects. Without you, I could not go overseas to have further learning and travel after my graduation from undergraduate university. Thank you for your tolerance to my temper and my missing of phone calls. I also would like to say a great, great thank you to my grandma of my mother side, Mrs. Shichun Wang. It was you who guided me in learning English when I was quite young, and you are my model in English teaching.

Third, my gratitude goes to my previous and current academic tutors and peers in Leeds, and those in other places. Thanks to colleagues in the Centre for Language Education Research (CLER) within the School of Education, and its leader, Prof. Maggie Kubanyiova, my previous academic personal tutor, for your supports and experience sharing. Thanks to Dr. Peter Hart and Ms. Rhiannon Price for your outstanding administrative support. Special thanks to Dr. Yunfeng Liu at School of Foreign Studies, Anhui Polytechnic University in China. Thank you for allowing me to be an audit to your sessions of corpus linguistics, even if I was not a student of that university. Thank you for your theoretical guidance on phraseology and hands-on practice in sessions to help me on corpus software. I would like to thank my previous tutors at Chuzhou university in China, who guided me how to be a good teacher. I also would like to say “thank you” for my academic peers: to Dr. Xiaowen Liu, it was you who let me know what a PhD student is; to Dr. Jie Liang, Dr. Zhouhan Jin, Dr. Rasha Mohsen, Dr. Zihao Cao, Yang Liu, Chengyan Li, Airen Zhang, Henry Troy and Xuechun Huang, for great academic communication with

you; to Mingyan Gao, Xueqi Chen and Kenneth Wangyin Li, for sharing academic literature online with me from different countries or regions and universities, hoping you will gain a lot in your PhD journey; to Zhenfeng Zhang, for introducing me to Dr. Yunfeng Liu and wish you a success for your PhD application. Also, thank you to my former IELTS student Xi Long, for your inspirations on the very basic idea of my research proposal of PhD study.

Fourth, I would like to express my gratitude to my friends. Thank you Max Jingyi Li and Carol Binyan Luo, my dear friends at the School of Education when we were Master students. Thank you for your constant support and unwavering belief that I could complete this PhD journey even when I lacked confidence in myself. Also, thank you my net friends “Grapefruit” Wei Li and “Fairy” Juan Lv in China, for your emotional support with beers at the end stage of my thesis writing.

Last but not least, I would like to thank Prof. Philp Durrant and Dr. Thomas Hammond for agreeing to examine this thesis; thank Dr. Richard Badger and Dr. Taguhi Sahakyan for your great suggestions at my transfer and my early stage of PhD; thank Dr. Sarah Daniel for your valuable comments on my thesis.

# Abstract

This research focuses on a specific discipline, Logistics, to explore specialised vocabulary and polysemy in the Logistics academic written context through corpus. Results of this research were generated from a self-built specialised corpus of 5.35 million tokens on Logistics academic written English, covering 720 texts sampled from nine top academic journals in that field.

To select potential words for polysemy analysis, a Logistics specialised academic written English vocabulary list (the LoSAWEL) was developed through keyness, range, dispersion, focusing on lexical words. This list contains 311 individual word types, each classified at different levels with Nation's (2012) BNC/COCA 25 base word lists as references to build connection with their general use frequency.

Word sense identification (collocation analysis, concordance line reading, and general use and discipline-specific dictionary checking) was conducted in the LAWEC (a self-built specialised corpus on Logistics academic written English), the BAWE (a general academic corpus) and the BNC1994 (a general use corpus) respectively. Word sense comparisons were conducted between the Logistics specialised context and general academic context, and between the Logistics specialised context and general use context. Results show that about half of senses had difference in senses in each round of word sense comparison respectively. Adverb and preposition senses were used consistently across corpora. Noun, verb and adjective senses show differences across registers. Shared senses between the Logistics specialised context and general academic context suggest the potential existence of a group of core academic vocabulary, and the difference of senses show disciplinary variation of word sense and word use. Also, register plays a role on differences of word senses.

Results also showed that more than 1/3 individual words in the LoSAWEL are polysemous words. These polysemous words have relations with their levels of general use frequency. The higher general use frequency level a word has, the more likely it can be a polysemous word. These words are also associated with part of speech, with adverbs and prepositions less likely to be polysemous words but nouns, verbs, and adjectives are. This research will contribute to corpus analysis on polysemy in specific context.

## Table of Contents

Acknowledgements .....	ii
Abstract .....	iv
Table of Contents .....	v
List of Figures.....	ix
List of Tables .....	x
List of Abbreviations.....	xiii
Chapter 1 Introduction .....	1
1.1 Background of the research .....	1
1.2 Motivation and rationale of the research .....	5
1.2.1 Personal motivation .....	5
1.2.2 Academic Motivation .....	5
1.3 Context of the research .....	6
1.4 Significance of the research .....	9
1.5 Structure of the thesis .....	10
Chapter 2 Corpus-derived Specialised Vocabulary List.....	11
2.0 Introduction .....	11
2.1 Specialised Vocabulary .....	11
2.1.1 Definition of Specialised Vocabulary .....	11
2.1.2 The importance of specialised vocabulary in research.....	13
2.2 Principles and methods in corpus-derived specialised vocabulary list development ...	15
2.2.1 Unit of Counting.....	16
2.2.2 Selection criteria.....	18
2.2.3 Validation of corpus-derived specialised vocabulary list.....	21
2.2.4 Approaches to developing corpus-derived specialised vocabulary lists .....	23
2.3 Corpus-derived specialised vocabulary lists .....	26
2.3.1 General academic vocabulary lists .....	27
2.3.2 Discipline-specific academic vocabulary lists.....	28
2.3.3 Technical vocabulary lists .....	32
2.4 Further development: Working beyond forms.....	34

Chapter 3 Polysemy .....	35
3.0 Introduction .....	35
3.1 Polysemy and related concepts.....	35
3.2 Polysemy as a linguistic barrier in specialised language learning contexts .....	36
3.3 A corpus linguistic approach to polysemy analysis .....	40
3.4 Corpus analysis of polysemy in specialised contexts .....	44
Chapter 4 Methodology .....	47
4.0 Introduction .....	47
4.1 Corpus Compilation .....	47
4.1.1 Principles underlying corpus compilation .....	47
4.1.2 Actual steps and modifications of corpus compilation in practice .....	53
4.1.3 The evaluation of corpus.....	61
4.2 Logistics specialised vocabulary selection criteria and list development .....	64
4.2.1 Decision on unit of counting.....	64
4.2.2 Software selection.....	65
4.2.3 Specialised vocabulary selection criteria.....	66
4.3 Word sense analysis .....	72
4.3.1 Tool choosing for word sense analysis.....	73
4.3.2 Collocation checking.....	74
4.3.3 Reading concordance .....	78
4.3.4 Identifying and splitting senses with dictionaries .....	78
4.3.5 Pilot study on word senses identification .....	80
4.4 Comparison of word senses between contexts .....	83
4.4.1 Choosing corpora for general academic contexts .....	84
4.4.2 Choosing corpora for general use contexts .....	85
4.4.3 Word sense identification in the BAWE and the BNC1994 NAW .....	86
4.4.4 Pilot study on cross-corpus comparison of word senses .....	89
4.5 Summary.....	93
Chapter 5 Results .....	94
5.0 Introduction .....	94
5.1 Development of the Logistics Academic Written English Word List (LoSAWEL).....	94
5.1.1 The specialised words in Logistics academic written English.....	94

5.1.2 Statistical analysis on frequency, range and dispersion on LoSAWEL words .....	97
5.1.3 Lexical coverage of the LoSAWEL .....	99
5.1.4 Comparison between the LoSAWEL and other corpus-derived word lists .....	100
5.2 Comparison of LoSAWEL item word senses in Logistics academic written context, general academic written context and general use context .....	103
5.2.1 Word senses of LoSAWEL items in Logistics academic written context .....	103
5.2.2 Word sense of LoSAWEL items in the general academic written context .....	109
5.2.3 Word senses of LoSAWEL items in the general use context.....	112
5.2.4 Word sense comparison between the LAWEC and the BAWE .....	115
5.2.5 Word sense comparison between the LAWEC and the BNC1994 NAW.....	131
5.2.6 Summary for word sense identification and comparison in the LAWEC, the BAWE, and the BNC1994 NAW.....	146
5.3 Polysemy in the Logistics Academic Written Context.....	147
5.3.1 polysemous LoSAWEL words in the LAWEC .....	147
5.3.2 LoSAWEL polysemous words and general use word frequency .....	148
5.3.3 LoSAWEL polysemous words, part of speech and register-specific collocations	151
5.3.4 Polysemy categories in the LAWEC .....	154
5.3.5 Summary of polysemy in the LAWEC .....	161
5.4 Summary of Results .....	162
Chapter 6 Discussion .....	163
6.0 Introduction .....	163
6.1 Specialised vocabulary in the academic written Logistics English context .....	163
6.1.1 Specialised vocabulary across vocabulary categories and general use frequency levels.....	163
6.1.2 Specialised vocabulary focussing on the most frequent 2000 general use vocabulary.....	166
6.1.3 Interdisciplinary feature of Logistics specialised vocabulary.....	168
6.2 Word sense analysis and comparison across corpus .....	169
6.2.1 Word form as a more suitable counting of unit for word sense analysis in specialised context.....	169
6.2.2 Core academic vocabulary, senses and register-specific collocations.....	171
6.2.3 Word senses varying in different registers.....	171
6.3 Polysemy in Academic Written Logistics English Context.....	173

6.3.1 Polysemy and specialised Logistics vocabulary .....	173
6.3.2 Polysemy and general use high frequency words in discipline-specific contexts	174
6.3.3 Polysemous word senses, part of speech and register-specific collocations .....	176
Chapter 7 Conclusion.....	179
7.1 Summary of the research.....	179
7.2 Contributions of the research .....	180
7.3 Implications of the research .....	181
7.4 Limitations of the research .....	182
7.5 Directions for future research.....	183
References.....	184
Appendix A: Function Word List .....	204
Appendix B LoSAWEL and Word Sense Identification in the LAWEC, the BAWE and the BNC1994 NAW .....	208

## List of Figures

Figure 1.1 Word frequency of "thrombocytopenia" in different registers of the COCA .....	3
Figure 2.1 Number of published corpus-derived narrow-view discipline-specific individual vocabulary lists (from 2000 to 2025).....	31
Figure 3.1 Semantic related scale used in Wang and Nation (2004) (adopted from Wang and Nation, 2004, p.297) .....	38
Figure 4.1 Input codes for random issue selection .....	58
Figure 4.2 Examples of random issue selection.....	58
Figure 4.3 Input code for random paper selection .....	59
Figure 4.4 Examples of random paper selection.....	60
Figure 4.5 Screenshot of part of collocates of <i>model</i> .....	75
Figure 4.6 Concordance lines examples of the node <i>model</i> and the collocate developed.....	75
Figure 4.7 The top 10 collocates of <i>model</i> and their occurrences in concordance lines.....	76
Figure 4.8 Samples of concordance lines of node <i>model</i> and its collocate for (number = 200) .....	77
Figure 4.9 Samples of concordance lines of node <i>model</i> and its collocate for (number = 300) .....	78
Figure 4.10 Examples of concordance lines for collocation “mathematical model”.....	80
Figure 4.11 Collocate list of <i>paper</i> in +/-5 span and with LogDice $\geq 9$ .....	82
Figure 4.12 Extract concordances of “judicial review” in the BAWE.....	88
Figure 4.13 The metadata of an extract of concordance of “judicial review” in the BAWE...	89
Figure 4.14 Examined collocates of <i>model</i> in the LAWEC .....	90
Figure 4.15 Examined collocates of <i>model</i> in the BAWE .....	90
Figure 4.16 Examined collocates of <i>model</i> in the BNC1994 NAW .....	91
Figure 5.1 Correlation between the number of word senses and the number of items (with corresponding percentages) .....	104
Figure 5.2 Percentage of monosemous and polysemous words at each LoSAWEL level.....	149

## List of Tables

Table 2.1 A summary of broad-view discipline-specific individual vocabulary lists .....	30
Table 4.1 Information about the ten chosen journals .....	52
Table 4.2 Suggested maximum word count of journals .....	54
Table 4.3 Original and current corpus design on number of texts .....	56
Table 4.4 Original and current plan on sample selection .....	56
Table 4.5 Coding Example .....	57
Table 4.6 The size of the Logistics Written Academic English Corpus .....	61
Table 4.7 Top 30 specialised selected items (ranked by Cohen's d from high to low) .....	72
Table 4.8 The number of collocations generated with different LogDice thresholds .....	81
Table 4.9 Probabilities of two senses of paper occurring in the LAWEC .....	83
Table 4.10 The amount of collocates generated with different LogDice threshold in the BAWE .....	86
Table 4.11 The amount of collocates generated with different LogDice threshold in the BNC .....	87
Table 4.12 Word senses of <i>model</i> in three corpora (emphasis added) .....	92
Table 5.1 Keywords, Dispersion, and Range criteria applied for word selection .....	95
Table 5.2 Criteria of lexical words applied for word selection .....	95
Table 5.3 The top 10 frequent specialised words in the LoSAWEL .....	96
Table 5.4 The top 10 specialised words in the LoSAWEL based on Range .....	96
Table 5.5 The top 10 evenly distributed specialised words in the LoSAWEL .....	97
Table 5.6 Frequency of the LoSAWEL words .....	98
Table 5.7 Range of the LoSAWEL words .....	98
Table 5.8 Dispersion of the LoSAWEL words .....	99
Table 5.9 Lexical coverage of the LoSAWEL .....	99
Table 5.10 The LoSAWEL words in the Nation's (2012) BNC/COCA 25 base word lists ..	101
Table 5.11 General academic written words in the LoSAWEL .....	102
Table 5.12 An example of word sense identification in the LAWEC .....	103
Table 5.13 The number of word senses and items .....	104
Table 5.14 Part of speech categories of word senses in the LAWEC .....	105
Table 5.15 Four preposition word senses and their word types in the LAWEC .....	106
Table 5.16 Examples of general list collocations in the LAWEC .....	107
Table 5.17 Examples of disciplinary collocations in the LAWEC .....	108
Table 5.18 An example of word sense identification in the BAWE .....	109
Table 5.19 Part of speech categories of word senses in the BAWE .....	110
Table 5.20 Examples of general list collocations in the BAWE .....	111
Table 5.21 Examples of disciplinary collocations in the BAWE .....	112
Table 5.22 An example of word sense identification in the BNC1994 NAW .....	113
Table 5.23 Part of speech category of word senses in the BNC1994 NAW .....	114
Table 5.24 Five selected examples of general list collocations in the BNC1994 NAW .....	115
Table 5.25 Comparison of word senses identified in the LAWEC and the BAWE at each LoSAWEL level .....	116

Table 5.26 Comparison of word senses in part of speech categories in the LAWEC and the BAWE .....	117
Table 5.27 Comparison of word senses in register-specific collocation categories in the LAWEC and the BAWE.....	117
Table 5.28 Number of senses shared or not shared between the LAWEC and the BAWE ...	118
Table 5.29 Part of speech and LoSAWEL levels of senses found only in the LAWEC .....	118
Table 5.30 Examples of senses in the LAWEC but not in the BAWE.....	120
Table 5.31 Number and LoSAWEL levels of register-specific collocations identified in the LAWEC but not in the BAWE .....	121
Table 5.32 LAWEC general list collocations not identified in the BAWE .....	122
Table 5.33 LAWEC disciplinary collocations not identified in the BAWE.....	123
Table 5.34 Part of speech of LAWEC-BAWE shared senses (D).....	124
Table 5.35 LAWEC-BAWE shared register-specific collocations (D) .....	125
Table 5.36 Word sense classification in part of speech and register-specific collocation categories in LAWEC-BAWE word sense comparison .....	125
Table 5.37 Collocates of <i>lead</i> in the LAWEC and the BAWE, ordered by LogDice .....	127
Table 5.38 Collocates of <i>per</i> in the LAWEC and the BAWE, ordered by LogDice .....	129
Table 5.39 Collocates of <i>order</i> in the BAWE (LogDice $\geq$ 9) .....	130
Table 5.40 Collocates of <i>order</i> in the LAWEC (LogDice $\geq$ 9).....	130
Table 5.41 Comparison on word senses identified in the LAWEC and the BNC1994 NAW at each LoSAWEL level.....	132
Table 5.42 Comparison of word senses in part of speech categories in the LAWEC and the BNC1994 NAW .....	133
Table 5.43 Comparison of word senses in register-specific collocation categories in the LAWEC and the BNC1994 NAW .....	133
Table 5.44 Number of senses shared or not shared between the LAWEC and the BNC1994 NAW .....	134
Table 5.45 Word sense classification in part of speech and register-specific collocation categories in LAWEC-BNC word sense comparison .....	134
Table 5.46 Part of speech and LoSAWEL levels of senses found only in the LAWEC but not in the BNC1994 NAW .....	135
Table 5.47 Examples of senses in the LAWEC but not in the BNC1994 NAW .....	136
Table 5.48 Number of register-specific collocations in the LAWEC, BNC1994, and those only in the LAWEC but not in the BNC1994 NAW .....	138
Table 5.49 Examples of general list collocations found in the LAWEC but not in the BNC1994 NAW .....	139
Table 5.50 Disciplinary collocations in the LAWEC but not found in the BNC1994 NAW.	140
Table 5.51 Number of LAWEC-BNC shared senses (D) in each part of speech category and each LoSAWEL level.....	141
Table 5.52 Examples of LAWEC-BNC shared senses (D).....	142
Table 5.53 Examples of LAWEC-BNC1994 NAW shared register-specific collocations (D) – general list collocations.....	143
Table 5.54 Two rounds of word sense comparison .....	146

Table 5.55 polysemous words of the LoSAWEL classified by the number of senses identified. .....	147
Table 5.56 An example of LoSAWEL polysemous word in the LAWEC .....	148
Table 5.57 Number of polysemous words at each LoSAWEL level.....	149
Table 5.58 examples of polysemous words at each LoSAWEL level.....	150
Table 5.59 Polysemous word senses and part of speech categories .....	151
Table 5.60 Senses of <i>set</i> identified in the LAWEC.....	153
Table 5.61 Collocates of <i>state</i> in the LAWEC .....	154
Table 5.62 Word senses of <i>ability</i> in the LAWEC .....	155
Table 5.63 Word senses of <i>high</i> in the LAWEC .....	156
Table 5.64 Word senses of <i>step</i> in the LAWEC .....	156
Table 5.65 Word senses of <i>quality</i> in the LAWEC .....	157
Table 5.66 Word senses of <i>transportation</i> in the LAWEC .....	157
Table 5.67 Word senses of <i>shows</i> in the LAWEC.....	158
Table 5.68 Word senses of <i>value</i> in the LAWEC.....	159
Table 5.69 Word senses of <i>function</i> in the LAWEC .....	159
Table 5.70 Word senses of <i>based</i> in the LAWEC .....	161
Table 6.1 Examples of specialised Logistics vocabulary at four levels.....	165
Table 6.2 Word sense of <i>lead</i> in three corpora.....	172

## List of Abbreviations

ASWL	Academic Spoken Word List
AVL	Academic Vocabulary List
AWL	Academic Word List
BNC	British National Corpus
COCA	Corpus of Contemporary American English
COBUILD Dictionary	Collins Birmingham University International Language Database Dictionary
EAP	English for Academic Purposes
EFL	English as a Foreign Language
ESL	English as a Second Language
ESP	English for Specific Purposes
IELTS	International English Language Testing System
LAWEC	Logistics Academic Written English Corpus
LoSAWEL	Logistics Specialised Academic Written English Word List

# Chapter 1 Introduction

## 1.1 Background of the research

Vocabulary is closely related to almost every aspect of our life and communication. As one of the classical quotations states: “Without grammar little can be conveyed; without vocabulary nothing can be conveyed.” (Wilkins, 1972, p.111). Similarly, vocabulary is highly related to language learning and the four language skills: listening, reading, speaking and writing, where vocabulary is the bricks for construction of the plaza of language (Webb and Nation, 2017). This is because we all need vocabulary to convey meanings for expressing our mind.

Vocabulary knowledge covers a wide range of aspects of vocabulary in use, which have been classified as nine aspects in three functions: form, meaning and use (Nation, 2001, 2013). Of these, the form-meaning connection, or what meaning is conveyed through a certain word form, is a central aspect (Nation, 2022). One of the short-cuts on form-meaning connection is vocabulary lists. Corpus-derived vocabulary lists are a promising resource for vocabulary teaching, learning, testing and research because they contain typical items that provide considerable lexical coverage in texts and are derived from authentic language analysis, instead of intuition (Dang, 2020). With these, learners are likely to get the best return of their effort, because they know which words are most important to meet their learning needs (Durrant, Siyanova-Chanturia, Kremmel and Sonbul, 2022; Nation, 2022). However, most vocabulary lists focus on form (Gardner, 2007), and they do not provide sufficient information about word meaning. Also, the form-meaning connection sometimes is not an one-to-one relationship. A certain word form may have different senses of related meanings in different contexts, which is defined as polysemy (McCarthy, 2023). For example, “see” has different senses in:

- (a) I can see the lake through my window.
- (b) I see what you mean.
- (c) I will see you this afternoon.

Through looking up “see” in *the COBUILD dictionary* (2023, pp.1342-1343), it refers to “noticing something with eyes” in sentence (a), which seems to be the basic sense of this word. Differently, in sentence (b) it means “to understand”, and it is typically used in spoken language, often in the multi-word unit “I see”. In sentence (c), it has the sense of “meeting or visiting someone”. These are only some examples of “see”, and the COBUILD dictionary provides several senses for this word. It may have other different senses in specific contexts, since the sense of a word depends on the context where it occurs (Aitchison, 2012).

Polysemy is a common linguistic feature and most words frequently used in everyday language have several senses (McCarthy, 2023). Researchers and practitioners have noticed the existence of polysemy in English language teaching and learning. Studies on polysemy have been conducted from different perspectives, including language description (e.g., Berdicevskis, 2020), polysemous word acquisition (e.g., Alnamer, 2017; Corssley, Salsbury and McNamara, 2010; Lee, 2021; Wei and Lou, 2015), cognition processing (e.g., Csábi, 2004; Sultanova, 2025; Williams, 1992) and computational linguistics (e.g., Almeman and Schockaert, 2024; Gulhayo, 2024). However, most studies have focused on polysemy in everyday language.

Polysemy also exists in specialised contexts, for example, general academic English or discipline-specific English. Hyland and Tse (2007) questioned the existence of core academic vocabulary and called for studies on discipline-specific contexts. They noticed that sometimes the meaning of a word form varies in different discipline-specific contexts. Their research suggests the difference of vocabulary in specific contexts. It also indicates the variation of word senses of the same word form depended on context where they occur. Therefore, discipline-specific vocabulary may cause difficulty in communication between professionals and laymen, or between experts and novices (Coxhead, Parkinson, Mackay, and McLaughlin, 2020), because laymen or novices are not as familiar with disciplinary specific context as experts are. Polysemous words with both general use senses and technical senses is one of the potential difficulties, when compared to that pure technical vocabulary, or terminology, which has only one sense. This is because a technical vocabulary item with only one sense, a monosemous word, occurs in a limited range of contexts (McCarthy, 2023), and its sense is consistent across topics. However, polysemous words with both general use senses and technical senses may occur in varied contexts, where the sense can be quite different from each other. For example, “thrombocytopenia” is a medical technical vocabulary item. It is very unlikely to occur in everyday language or to be known by people with less medical knowledge. The search through Corpus of Contemporary American English (COCA) (Davies, 2008-) provides evidence. The COCA provides word frequency in different registers, shown in Figure 1.1. This word is more likely to occur in academic contexts, but much less in registers related to everyday topics such as spoken, fiction, magazine and news contexts. This word, however, may not be as difficult to medical professionals as it is to non-specialist people, since it is related to their field and has only one sense across contexts.


SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
FREQ	89	6	10	7	0	0	4	1	61
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8
PER MIL	0.09	0.05	0.08	0.05	0.00	0.00	0.03	0.01	0.51
SEE ALL SUB-SECTIONS AT ONCE									

Figure 1.1 Word frequency of "thrombocytopenia" in different registers of the COCA

“case” is an example of polysemous words with both general use sense and technical sense by referring to the COBUILD dictionary (2023, pp.221-222):

- (d) ... a black case for his spectacles.
- (e) ... a case of champagne.
- (f) ... the nominative case. [grammar]
- (g) It can be difficult for public Figures to win a libel case. [law]
- (h) The couple were made official suspects in the case. [crime]
- (i) ... the case of a 57-year-old man who had suffered stroke. [medicine, health and disease]

In (d) and (e) without specific contexts given, “case” refers to “a container or box to hold or protect something”, and in (e) it specifically refers to “a box for alcoholic drinks, usually twelve bottles”. These senses may be frequently used in everyday language. For the rest of the example sentences, although the COBUILD dictionary marks up on the specific contexts where they often occur, it is a general use dictionary which may not as professional as those technical dictionaries. Therefore, I double checked example sentences from (f) to (i) in relevant technical dictionaries – *the Oxford Concise Dictionaries*, which have different volumes covering different disciplines. “case” is listed as an entry of technical vocabulary in the volume of English Grammar (Aarts, Chalker and Weiner, 2014), Law (Law, 2022), Crimes and Justice (Davis, 2020) and Medicine (Law and Martin, 2020) respectively. “case” in each of four example sentences of specific context has technical sense and is different from each other. In (f) it refers to “a grammatical category which indicates the function of nouns or pronouns within a sentence”. In (g) it has the sense of “trial”; in (h) it is “a crime that police are investigating”. In (i) it refers to “a problem that doctors are dealing with”. By referring to Nation’s (2012) BNC/COCA 25 base word lists, “case” is one of the most frequent 1000 English words. However, the examples shown above suggest that “case” is also shared across disciplines, but has different technical senses related to discipline-specific contexts. Words like “case” can be traps for EFL learners or novices with less professional

knowledge, because of their unfamiliarity to the specific contexts and their taking it for granted to believe that they know such general high-frequency vocabulary.

Studies have shown that polysemous words can pose difficulties for learners in the specific contexts (e.g., Deignan, Semino and Paul, 2019; Le and Miller, 2023; Mićović and Beko; 2022). Based on interviews with middle school students with English as first language (L1), Deignan, et al. (2019) noticed that students believed they have acquired some words but actually they did not accurately paraphrase their technical senses in specific subjects, due to polysemy. The issue was also found in learning of English as a Second Language (ESL) or English as a Foreign Language (EFL) learners (Mićović and Benko, 2022). However, few studies have systematically explored polysemous words in a specialised context (Skoufaki and Petrić, 2021). It suggests the gap of exploring polysemous words in specialised contexts, and having comparison between general use senses and technical senses of polysemous words in description. With corpus analysis (Deignan, Candarli and Oxley, 2023) and dictionary checking (Le and Miller, 2023; Skoufaki and Petrić, 2021) scholars have explored polysemous words in the general academic context and medical discipline-specific context respectively. A corpus is a representative large computer-readable text collection, which captures natural language use and provides authentic language examples in contexts. This has advantages for exploring word sense and word use through a large sum of concordance lines rather than a handful of intuition-based examples (Hunston, 2022; Sinclair, 1991). Nevertheless, corpus-driven analysis of polysemous words in a discipline-specific context remains rare. This suggests much room for exploring polysemous words and word senses in a specific context.

In my research, I followed previous studies such as Skoufaki and Petrić (2021), Le and Miller (2023) and Deignan et al. (2023) to explore polysemous words in a specialised context through corpus analysis. Logistics, a specific discipline, was set as the specialised context in this research with a special focus on written academic English texts. The rationale of discipline selection will be discussed later in section 1.3. A specialised corpus, **Logistics Academic Written English Corpus (LAWEC)**, was compiled for data analysis. My research firstly investigated what individual words are most frequently used and evenly distributed in the Logistics written academic contexts, and what formal features they have (e.g., part of speech). A specialised vocabulary list on Logistics was developed. It then explored word senses of individual words in the list in three corpora respectively: (1) the LAWEC, (2) British Academic Written English Corpus (BAWE), and (3) British National Corpus (BNC). Collocation analysis, concordance line reading and discipline-specific technical dictionary checking were involved in word sense identification. Comparisons were made between the LAWEC and the BAWE, and between the LAWEC and the BNC respectively. This aims to see to what extent the word senses in the LAWEC are different to those in the general academic written context or the general use context. Finally, I focused on the LAWEC, and explored to what extent the words in my specialised vocabulary list are polysemous words. My three research questions are as follows:

RQ1: *What are the most frequently-used and evenly-distributed specialised Logistics words, and what are their formal features?*

RQ2: *To what extent are the words identified used in ways specific to Logistics?*

*a) is their use different from the general academic context?*

*b) is their use different from general English?*

RQ3: *What is the extent and nature of polysemy in the words identified?*

In this thesis, when presenting examples of individual words and word senses, words *in italics* are items in my specialised vocabulary list on Logistics. For items not included in the list, double quotation marks “” are presented. Capitalization is for examples of LEMMA.

## 1.2 Motivation and rationale of the research

### 1.2.1 Personal motivation

The motivation of specialised word list development in this research originates from my experience of an IELTS teacher in a language centre, mainland China, and a personal communication with one of my students. In 2021, I worked as an IELTS teacher on reading and writing courses, where I used the first nine lists of Nation’s (2012) BNC/COCA 25,000 word lists and Coxhead’s (2000) the Academic Word List (AWL) as additional materials for vocabulary instruction. I also introduced corpus-derived field-specific lists available from the “Vocab” part of the EAP foundation website (<https://www.eapfoundation.com/vocab/academic/other/#field>) (Smith, 2021) to address specialised vocabulary that is highly related to a specific topic in IELTS reading texts. These vocabulary lists in my courses received positive feedback from students. They felt that they have a clear learning goal set based on these lists and gradually developed their awareness of vocabulary in registers.

One of the students came to me after sessions, and raised some questions on vocabulary lists in English learning. She asked whether there is a discipline-specific vocabulary list about Logistics, a discipline for her postgraduate study. She also asked how to deal with items with multiple meanings in the lists (e.g., which sense should be leaned in priority). It was through the communication with this student that I noticed the potential gap in discipline-specific vocabulary research. I had a hypothesis before searching for related publications: there are still room for discipline-specific vocabulary research, where some less-explored specific disciplines related to our life deserve attention.

### 1.2.2 Academic Motivation

Based on that hypothesis, searching for and reviewing related studies were conducted. Studies (e.g., Dang, 2020; Dang and Webb, 2025) have shown that corpus-derived specialised vocabulary

lists witnessed more sophisticated methodology in the last two decades, and the contexts that lists focus on became more specific and technical. However, most of lists are on hard science disciplines such as Computer Science, Engineering, and Medicine. Within the recent decade, attention has shifted to more technical contexts, for example, Carpentry, Fabrication, Rugby, and Radiotelephony. Dang (2020) called for more studies on some under-researched areas of corpus-derived vocabulary lists. Some under-explored subjects or disciplines which are related to modern life, such as Logistics, can be one of potential topics.

Moreover, initial reading and questions raised by my student reminded me of polysemy as a research topic. Research on polysemous words was mainly on general language, with few exceptions on specialised contexts through corpus analysis (e.g., Deignan et al., 2023; Le and Miller, 2023; Skoufaki and Petrić, 2021). As discussed in section 1.1, students like L1 and L2 learners may experience difficulty when encountering polysemous words in specialised English contexts. Studies on word senses may help setting clear learning goals for learners. However, studies of systematic analysis on polysemous words and word senses in specialised contexts or in specialised vocabulary lists instead of providing intuition-based examples are in scarcity. Further, word senses of polysemous words in specialised contexts are seldom compared to those in general academic and general use contexts. Addressing these questions is also in line with the call of Hyland and Tse (2007) to recognize different ways of language use in different contexts.

### 1.3 Context of the research

The context of this research is Logistics, with a special focus on written disciplinary academic texts. No participants were involved in this research, but the potential target users of the specialised vocabulary list on Logistics developed in this research are mainly undergraduate or postgraduate students majoring in Logistics in universities of China. It may also suit language teachers in this field, and professionals who are interested in Logistics worldwide. In this section, the rationale for choosing Logistics as the specialised context is firstly discussed, and then for the rationale of potential target users of the list.

Logistics refers to the whole process of planning, transporting and controlling goods or information from a place to another (Quyet and Phung, 2023). In other words, Logistics is the service that plans and transports goods from producers to its customers with optimal management. The concept of Logistics originates from military, providing and distributing weapons and wartime necessities to soldiers for their operations, but it plays a more civilian role after the World War II (ibid). It is now interdisciplinary, closely related to disciplines such as Business, Management, Economics, Mathematics, Psychology and Sociology (Delfmann et al., 2010;

Swanson, Goel, Francisco and Stock, 2017). Logistics integrates various business and trade activities into a single system by using approaches or theories from aforementioned related disciplines (Milenkov, Sokolović, Milovanović, and Milić, 2020). None of business and trade activities would be considered as a success, if the production could not be delivered or arrive to customers on time. Logistics, therefore, is recognized as the middleman between producers and customers in the whole business process. Logistics is also related to everyday life, which flourishes due to globalization, just-in-time delivery services, advancement of technology on computer sciences and transportation, and the expansion of Internet shopping (Risberg, 2023). E-commerce is one of the fields promoted by Logistics (Sardarabady and Durst, 2024). For example, Amazon relies on its powerful logistics network (Andreoli-Versbach and Gans, 2024; Hassel and Sieker, 2023). Similarly, in China, evidence has shown the success of the Alibaba Group and Jingdong, two China e-commerce giants, results from their well-organized self-built logistics systems (Luo, Rong, and Zheng, 2020; Wang, Y. and Zhao, J. 2022), which changed the labour market of China. From life to work, Logistics influences different aspects of daily life. The Logistics industry employs millions of people worldwide, from warehouse workers to truck drivers and delivery staff. A survey reported that around 5-10% working force is at this sector, and around 17.8 million people in China were employed in Logistics in 2023, This amount of working people is the largest among six countries in the survey (Contract Logistics Observatory, 2023). According to *the Logistics Market Share, Size and Trends 2025 to 2034* (Precedence Research, 2025) (available from: <https://www.precedenceresearch.com/logistics-market>), the global logistics industry is to reach \$11.23 trillion in value by 2025, which is mainly driven by e-commerce and technology advancement. It is estimated to reach around \$23.14 trillion by 2034 with digitalization and AI as the main driving force, and it is also estimated the Asia-Pacific region as the main expanding market (around 45% market share) (ibid). All these suggest that Logistics is a key sector closely related to global economy and life, and a significant topic in research.

Due to the significance of Logistics in research, technology development and daily life, Logistics as a discipline also receives attention in China in the last two decades. By 2022, there are 337 universities in China listed on the website (<https://www.dxsbb.com/news/11463.html>) which set Logistics as an independent major, with thousands of students admitted or graduating every year. It now increases to 519 universities and colleges by September 2025 (ibid). To promote international competence in global economy, Logistics students in China are required to have professional Logistics disciplinary knowledge and English competence whose demand have been increasing since the initiation of the “Belt and Road” policy (Chen, 2019) and the outbreak of Covid-19 (Gong and Song, 2022). Logistics English has been set as a compulsory professional course of English for Specific Purposes (ESP) for the 3<sup>rd</sup>-year Logistics undergraduates (Pan,

2008; Wang, K., 2022). Some universities also have similar courses for postgraduates. This course usually lasts for a whole academic year with 30-40 sessions (45 mins each; two sessions per week) (Pan, 2008). Each university has autonomy in choosing textbooks, and these textbooks are considered to be a major language input channel (Li, 2018; Zhao, B., 2018). Previous studies have shown that the students have learned general English at least for 8 years before having ESP courses like Logistics English (Wang, K., 2022).

However, there are three issues related to Logistics students and Logistics English in China. First, although there are no exact data particularly for Logistics students in China on English language proficiency, practitioners report that students have low English proficiency based on teachers' intuition, in-class quizzes or English for General Purpose (EGP) test results such as Gaokao (Wang, K., 2022). There is other evidence to show Logistics students' English proficiency. Lu and Dang (2022) noticed that vocabulary size of postgraduate students from one of the top universities is not as large as expected, only having acquired the most frequent 1,000 word families on average. These students are from different disciplines of a top university in China, showing the overall low vocabulary size of students. Based on Lu and Dang's (2022) findings, it is reasonable to have an estimation that Logistics students in China also have such issue. Vocabulary size is a key predictors of English proficiency (Nation and Coxhead, 2021; Webb and Nation, 2017). Moreover, tutors report that students do not have a high level of autonomy and motivation in Logistics English learning, probably because of language barriers caused by technical vocabulary and its specialised senses in textbooks (Dai, 2020), inadequate either linguistics or Logistics disciplinary knowledge (or both) of tutors (Dai, 2020; Wu, N., 2016; Zhang and Wang, Y. H., 2019), or impractical uninteresting learning tasks. Further, according to a questionnaire survey on Logistics undergraduates, more than 40% of student participants believe that they are unable to have a fair grasp on Logistics English knowledge (e.g., vocabulary) or having fluent professional communication in English (Xie, Y., 2022). These all suggest that both tutors and students in this discipline may not have a clear understanding of the related linguistic knowledge (e.g., specialised vocabulary, phraseology and senses) that worth teaching and learning, which raises my interest on further investigation to address their specific language learning needs, and, to some extent, relieve their difficulties in language learning.

Second, Logistics English, although it is set as a compulsory course, does not receive much attention, especially in vocabulary studies. Although the Committee of Logistics Majors in Higher Education was established in 2018 (CLPP, 2019), there is no clear report about Logistics English and how to develop students' language proficiency in regulations on their official website. Moreover, by the time of writing, of all 404 journal articles on "Logistics English in China" available on the website of the China National Knowledge Infrastructure (CNKI), the biggest online academic database in China, few focus on specialised vocabulary in Logistics English, for

example, lexical bundles in Logistics English textbooks (Zhou, 2019), and the interdisciplinary and polysemous nature of Logistics English vocabulary (Chen, 2019; Xie, Z. and Xiong, 2012). They only provide a small number of examples based on intuitions instead of corpus data.

Third, apart from the disciplinary development and needs of professionals with both Logistics and English proficiency, the increasing number of Chinese EFL students, studying, or planning to study, overseas during or after undergraduate study also needs consideration. In China, there are 15 universities with collaborative agreements in Logistics majors with universities in Australia, Austria, Finland, Germany, South Korea, Thailand, the UK, and the USA (Ministry of Education, 2022). The Logistics students in these universities are in “2+2” or “3+1” programs (two-or-three-year study in China and one-or-two-year study abroad), and many students fund themselves to study abroad on an individual basis. No matter where they go, English is the language most likely for them to be exposed to because more than 80% of academic articles are published in English around the world (O’Neil, 2018). Being able to read disciplinary academic articles is a basic requirement for academic success. Therefore, focusing on Logistics written academic context and Chinese EFL undergraduate/postgraduate Logistics students as main target may bring benefits to a large group but neglected learners.

## 1.4 Significance of the research

I hoped to produce a systematic analysis of specialised vocabulary in Logistics written academic English, to know what individual words are most frequently used and evenly distributed across this context, and investigate polysemy of these items and to what extent their word senses are different from those in general academic and general use contexts.

My research aimed to produce results that may have a positive impact on understanding of polysemy in a discipline-specific context, and on teaching pedagogy of specialised vocabulary and its word senses by corpus linguistics. First, my results of specialised Logistics vocabulary list generated from a self-built specialised corpus will help setting language learning goals for relevant stakeholders such as undergraduates, postgraduates, ESP teachers and novices in the Logistics field. Second, through collocation analysis, concordance line reading and discipline-specific dictionary checking, the word sense comparison between specialised context and general academic context, and that between specialised context and general use context may provide opportunities for knowing to what extent the word senses are different across contexts. Finally, this research may provide insights into word sense analysis and contribute to awareness-raising of polysemy analysis in a specialised context. It may help promoting language practitioners’ and students’ awareness of vocabulary use in specific contexts.

## 1.5 Structure of the thesis

After the introductory chapter of this thesis, Chapter 2 is the first part of Literature Review. It discusses one of the foundational concepts in this study: specialised vocabulary. It also describes criteria and approaches to developing corpus-derived specialised vocabulary lists, and reviews relevant studies.

Chapter 3 is the second part of Literature Review. It discusses what polysemy is, and why polysemy is a linguistic barrier in language learning of specialised contexts. Methods for distinguishing word sense and conducting polysemy analysis through corpus linguistics are reviewed. It also reviews relevant studies on polysemy analysis with corpus methods in specialised contexts. These two chapters explain the inspirations and methodology for conducting this research.

Chapter 4 describes research design, methods and research process. First, the specialised corpus compilation as data collection is explained. Second, corpus-driven selection criteria of specialised Logistics vocabulary list development are presented, with trials for decision of cut-off points of criteria. Third, word sense identification process is tested with examples and pilot study, including collocation analysis, concordance line reading and sampling, dictionary checking, and the calculation of occurrence of word senses in citations. Fourth, steps of word sense comparison between specialised context and general academic context, and those for comparison between specialised context and general use context are demonstrated respectively.

Chapter 5 presents findings, consisting of three sections in accordance with three research questions of this thesis: specialised Logistics vocabulary list development, word sense comparison results between specialised context and general academic context and those in comparison with general use context word sense analysis, and polysemous word identification in the Logistics written academic context.

Chapter 6 presents discussion based on findings of this research on four aspects: (1) nature of specialised Logistics vocabulary and disciplinary vocabulary features, (2) similarities and differences of specialised Logistics word senses between Logistics written academic contexts and general academic context, (3) those in comparison with general use context, and (4) polysemy of specialised Logistics vocabulary.

Chapter 7 is the last chapter, which summarizes major findings of this research. Main contributions, relevant pedagogical implications, limitations, and recommendations for future research are also covered.

# Chapter 2 Corpus-derived Specialised Vocabulary List

## 2.0 Introduction

This chapter is the first part of Literature Review. This chapter, with a special focus on lists of individual words in written context, consists of four sections. Section 2.1 discusses the definition of specialised vocabulary, and the importance of specialised vocabulary in English language teaching and learning. Section 2.2 reviews principles and methods in corpus-derived specialised vocabulary lists. Section 2.3 reviews studies on corpus-derived specialised vocabulary lists and their pedagogical implications in English language teaching and learning. Finally, section 2.4 summarizes and discusses potential issues in corpus-derived specialised vocabulary lists which links the next chapter of Literature Review.

## 2.1 Specialised Vocabulary

### 2.1.1 Definition of Specialised Vocabulary

Different approaches exist in defining specialised vocabulary. A narrow view (University of Warwick, 2017) defines specialised vocabulary occurring in soft science disciplines like Arts, Humanities and Social Sciences. These vocabulary items and units are related to disciplinary knowledge and have specialised meanings in disciplines. This view clearly distinguishes specialised vocabulary from technical vocabulary, where technical vocabulary, in this narrow view, refers to words with specialised meanings particularly in the fields of hard science disciplines such as Sciences, Engineering, and Medicine, or trades. However, Liu and Lei (2020) point out that such distinguishing may not be much helpful since the boundary of disciplines is not a clear cut. In contrast, a broader view recognizes specialised vocabulary to be used interchangeably with technical vocabulary which may occur in any discipline (Nation et al., 2016). These items can be those used exclusively in a particular discipline, and can be those from general language but with disciplinary meanings (Liu and Lei, 2020). This view of interchangeable definition between specialised vocabulary and technical vocabulary can also be seen in other studies where similar terms are used, for example, “specialised lexis” (Baker, M., 1988, p.91), “specialized vocabulary” (Robinson, 1980, p.71) and “specialist terms” (Woodward-Kron, 2008, p.239).

A more recent definition of specialised vocabulary covering two categories of vocabulary: academic words and technical words (Coxhead, 2018; Nation, 2013, 2022), because these two categories of words are both more likely to occur in a particular range of topics for particular rhetorical functions or with topic-related meanings, when compared to everyday language. For academic words, they are frequently used in academic contexts (e.g., academic textbooks, journal articles, seminars, lectures, or conference presentations) and shared across different disciplines (Nation, 2022). For technical words, they may

only occur in technical registers, or are with professional senses that may be quite different from senses used in everyday language (ibid). My research follows this recent definition. In sum, specialised vocabulary refers to words much frequently used in particular contexts, suggesting that specialised vocabulary has a narrow range of use in particular contexts. These words may also occur in everyday language but may have different senses, as polysemous words. People within a particular field may have sufficient knowledge of these words; they may know and use such vocabulary accurately, fluently and appropriately. However, for those outsiders or novices of a particular context, they might have, or, they believe that they have some knowledge of specialised vocabulary but sometimes they may encounter difficulty or misunderstanding in communication (Coxhead, 2012). Specialised vocabulary is important for learners with special needs in language learning, for example, academic literacy development in secondary or tertiary education, technical studies, or professional communications in trades (Coxhead, 2018; Coxhead et al., 2020; de Chazal, 2014).

Specialised vocabulary has two typical features. First, specialised vocabulary can cut across different general use frequency levels (Nation and Gu, 2019). In other words, specialised vocabulary can be general high-, mid- or low-frequency words due to its nature of polysemy in different contexts where they occur (Schmitt and Schmitt, 2020). Some specialised vocabulary items are frequently used in everyday language but also with technical senses. For example, in general use, “price” and “cost” are classified as general high-frequency words because they are in the 1<sup>st</sup> 1,000 word list of Nation’s (2012) BNC/COCA 25,000 word lists. They are frequently used in everyday topics. However, these two words are also highly related to business or economic contexts with professional meanings (e.g., “Price Economics”, a sub-branch of Economics investigating product values and value changes in the market). In contrast, “acute”, a specialised word related to medical language (e.g., “acute medical emergencies”), is in the 4<sup>th</sup> 1,000 BNC/COCA word list and classified as a general mid-frequency word. In the Language and Linguistics field, “affix” is frequently used as a term with discipline-specific meaning; however, it is not as frequently used as in everyday topic and is classified in the 9<sup>th</sup> 1,000 BNC/COCA word list as a general low-frequency word. These examples indicate that the general use word frequency level may not be the sole criterion to identify specialised vocabulary. What matters is the context where it often occurs and the sense it conveys in the context.

Another feature of specialised vocabulary is that, although academic words and technical words are under the umbrella of specialised vocabulary in conception, there is a blurred boundary between these two categories of vocabulary, when compared to those in general use frequency levels. Such blurred boundary exists especially in a discipline-specific or professional-domain context. This is because some specialised vocabulary items may have specific word senses in different contexts. For example, Hyland and Tse (2007) noticed that “volume”, a general academic vocabulary item shared across several disciplines and an item included in the AWL, has different specific senses in science disciplines (mostly referring to “quantity”) and social science disciplines (mostly referring to “books”). Moreover, when

looking it up in dictionary on Physics – Oxford Dictionary of Physics (eighth edition) (Rennie and Law, 2019, p.927), “volume” can also be used as a term with the sense of “the space occupied by a body or mass of fluid”. Such phenomenon leads to the difficulty in classifying specialised vocabulary. Therefore, considering such blurred boundary and following Nation’s (2022) conceptualization, I use the term “specialised vocabulary” across this thesis if without specific statements, instead of either academic or technical words. Also, “specialised context” is used in this thesis referring to academic and/or technical contexts.

### 2.1.2 The importance of specialised vocabulary in research

Those above show the nature of specialised vocabulary: relatively frequently used in particular specialised contexts, cutting across the general use frequency levels and possibly having specific senses different from those in daily topics. Focusing on specialised vocabulary in research is important for at least three reasons.

First, in specialised contexts, both academic words and technical words cover a considerable amount of running words, either in written or spoken language. If these words are not handled properly in language teaching and learning, they could be a difficulty for learners. For example, the 1,741 items in the Academic Spoken Word List (ASWL) (Dang et al., 2017) account for about 90% lexical coverage in a 13-million-token academic spoken corpus. This suggests that a relatively small number of words is especially frequently used in such context and useful for language learners with specialised language learning needs. It also indicates that academic spoken communication would be less efficient if students do not acquire vocabulary knowledge of these words due to the high lexical coverage they provide. In written technical contexts, technical words usually cover about one in three of all running words. The specific percentage depends on the nature of target discipline (Chung and Nation, 2003). Chung and Nation (2004) used a four-step rating scale (see below) to identify technical words manually.

- (1) A word having a meaning that has no particular relation to the specific target field, for example, function words or other common words;
- (2) A word having a meaning that has minimal relation to the specific target field;
- (3) A word having a meaning that is closely related to the specific target field but the meaning is also used in general language or may occur in other fields as non-technical terms;
- (4) A word having a meaning that is particular to the specific target field and less likely to be known in general language.

They found that about 16% running words in an Applied Linguistic text and 37% words in an Anatomy text have technical sense. Such discrepancy suggests that the lexical coverage provided by specialised vocabulary may vary according to the context or discipline where it occurs. Other studies

also reported that technical words covering around a third of all running words in texts, for example, 35% in Hsu (2014) on Engineering textbooks, 30% in Lei and Liu (2016) on medical corpora, 40% in Lu (2018) on Chinese Traditional Medicine texts, and 30-38% in Coxhead et al. (2020) on four professional trades. Given such high lexical coverage of specialised vocabulary in texts, this vocabulary category is necessary for English for Academic Purposes (EAP) and English for Specific Purposes (ESP) learners to reach smooth communication in particular contexts.

Another reason is that specialised vocabulary conveys (1) academic conventions and (2) expresses discipline/technical knowledge. They may also (3) suggest underlying disciplinary epistemology and professional concepts (Hyland and Tse, 2007). These are essential elements for community inclusion and could also help with establishing identity of community members. Regarding (1) conveying academic conventions, one specialised vocabulary example is “acknowledgement”, an item included in the AWL. On searching in the COCA (Davies, 2008-), I noticed that this word is most likely to occur in academic register among all eight registers of COCA. It may result from the convention of expressing author’s gratitude to those who support his or her research in academia. The example of specialised vocabulary which (2) expresses discipline-specific knowledge and professional concepts is “function”. This word is also included in the AWL, and listed as an entry in the Middle School Mathematics Word List (Greene and Coxhead, 2015). Although these authors did not clearly provide what sense this word has in the AWL and the Middle School Mathematics Word List, this word may have different senses in the general academic context and Mathematics context. Specifically, The Concise Oxford Dictionary in Mathematics (6<sup>th</sup> edition) (Earl and Nicholson, 2021, p.418) defines “function” as a term with professional concept: “an expression for the relationship between one variable and another”. For specialised vocabulary (3) suggesting underlying disciplinary epistemology, “volume” discussed above can be an example. As discussed earlier, most likely, this word has different senses in sciences and social science disciplines. It could be the consequence that, the description on natural world is at the heart of sciences while, in social science, the focus is usually on human behaviours (Hyland and Tse, 2007). It therefore may lead to different senses in preference between sciences and social science disciplines: sense of “quantity” in sciences referring to measurement of natural elements for description and sense of “books” in social sciences referring to human intellectual production in social communication and society. These examples show that specialised vocabulary can be a carrier of academic or disciplinary conventions and knowledge. Without these, communication in specialised contexts would be greatly negatively affected. In general academic contexts, Hyland (2022) commented that the identity shift occurs from novice writers to experienced writers in academic writing, and the academic language they use (e.g., the use of specialised vocabulary) is one of the key facilitators. In more discipline-specific fields, Coxhead et al. (2020) summarized that both professional tutors and learners participating in their interviews on specialised vocabulary in trades believe that these words with technical senses play a great role in professional learning and trade talk, with a heavy demand on technical concepts. Coxhead

et al. (2020) also pointed out that these items help with identity shift of learners in moving from novices who often use general high-frequency words to those who think and behave like experts with appropriate use of specialised vocabulary in professional communication. One key phenomenon is the increased use of specialised vocabulary with technical sense, which suggests that learners are trying to use more accepted ways to participate communication activities in their fields.

Moreover, here is a third reason. Although specialised vocabulary has been recognized as a key element for learners' success in academic or professional fields, in research and teaching practice specialised vocabulary knowledge receives less attention when compared to general high-frequency vocabulary. Knowledge of academic vocabulary is considered as a key prediction on students' academic success (Ackerman and Chen, 2013; Coxhead, 2021, 2024). Similarly, knowledge of technical vocabulary also has its role in professional development for vocational students in English-medium context, and for those in technical fields (Coxhead et al., 2020; Liu and Lei, 2020). However, studies have found that EAP and ESP practitioners may not pay as sufficient attention to specialised vocabulary as researchers do in expectation (Coxhead, 2011; Zamfir, 2022). Moreover, classroom instructions, textbooks and materials in EAP and ESP programs may not always provide learners with sufficient opportunities on specialised vocabulary knowledge, due to lack of adequate instruction time, insufficient specialised vocabulary knowledge of teachers and heavy workload of teachers to manage tasks (Enesi, Vrapı and Trifoni, 2021; Fraser, Higa and Davies, 2025; Havwini, Ratnasari, Feby, Fadlilah and Firdus, 2024; Lu and Dang, 2022). Teachers' insufficient knowledge of specialised vocabulary is one of serious issues in EAP or ESP programs. Xhaferi (2010) pointed out that, even for language teachers who are not in a particular scientific field, they may have much difficulty in handling specialised vocabulary than those who are. This, therefore, may lead to students' insufficient specialised vocabulary knowledge and inadequate skills for professional communication or academic achievements. Scholars in related studies (e.g., Benson and Coxhead, 2022; Coxhead et al., 2020; Xhaferi, 2010) have suggested that language teachers in ESP courses for a particular disciplinary context need to be familiar with core specialised vocabulary in such context. Therefore, identifying specialised vocabulary that fits specialised language learning needs of students is a necessary step. In practice, providing a corpus-derived vocabulary list is one of the ways to identify specialised vocabulary and helps with planning teaching and learning (Durrant, Siyannova-Chanturia, Kremmel and Sonbul, 2022; Nation, 2022).

## 2.2 Principles and methods in corpus-derived specialised vocabulary list development

Corpus (pl: corpora) refers to a collection of text of naturally occurring languages, selected with principles, and stored and processed by electronic devices (Hunston, 2022; Sinclair, 1991). Due to its

nature of large size of data collection and its focus on natural language use representing target fields, corpus analysis helps much on vocabulary studies (Oakey, 2022). With corpus analysis and selection criteria set, a relatively small number of frequently-used key specialised words related to a particular context would be identified.

### 2.2.1 Unit of Counting

A core step in developing word list is the decision of unit of counting, that is, how to count a word (Dang, 2020). Such decision is usually dependent on the aims of research or students' language proficiency (Dang, 2020; Nation, 2016; Schmitt and Schmitt, 2020; Webb, 2021). Different units of counting can lead to different items included in vocabulary lists, generating various formats of word lists, and thus there is no one-fits-all unit of counting that is suitable for learners at all proficiency levels (Nation, 2016). There are five units of counting for individual words: token, type, lemma, flemma and word family (Nation, 2013), of which the last four are frequently used in vocabulary list development (Dang, 2020; Nation, 2016, 2022).

#### *Type*

Type, or word type, refers to counting each non-repeated word for only one time in the text (Dang, 2020). Still using the sentence "You like bananas and I like apples" as an example, "like" is repeated twice but considered as one word type. This sentence has six non-repeated words and therefore six word types. Chung and Nation (2003) argues that word type, rather than word family, is a more suitable unit of counting in identifying specialised vocabulary such as technical words. However, before 2010, very few early specialised vocabulary lists take word type as unit of counting. One of the exceptions is the Applied Linguistics Academic Word List (Vongpumivitch, Huang and Chang, 2009), which includes 475 AWL word types and 128 non-AWL word types. In recent years, word types gradually receive a wider acceptance in corpus-derived specialised vocabulary list development, for example, Lu (2018), Benson and Coxhead (2022), and Drayton and Coxhead (2023). Dang and Webb (2025) also reported the increasing use of word type, as a unit of counting in specialised texts. It has its advantage to "narrow down" learning goals targeting individual word forms which are particularly frequently used in specialised contexts, rather than the whole word family. It also lightens workload in vocabulary teaching and learning, especially for users whose word part knowledge is not well developed. Webb (2021, p.942) argues that "smaller lexical units appear more sensible with less proficient learners". However, due to its focusing on individual word forms, it may neglect users' word part knowledge development (e.g., related forms in inflectional and derivational systems), another important aspect of vocabulary knowledge (Nation, 2016).

### *Lemma*

Lemma refers to counting the base word and its inflections (Francis and Kucera, 1982). Another criterion of lemma is that all items included in the same lemma should be with the same part of speech (Dang, 2020). For example, the singular form of the noun “apple” and its plural form “apples” should be considered as one lemma, although they are two individual word types. Similarly, “walk” as a verb and its inflections, “walks”, “walked”, “walking” should be recognized as one lemma only, instead of four. This unit of counting is believed to be more suitable for young and low proficient learners who have less knowledge on affixes (Brown, Stoeckel, McLean and Stewart, 2022; Nation and Webb, 2011). Compared to larger units such as word family, lemma has advantages in corpus analysis in identifying and counting items (Gablasova and Brezina, 2021) and research practice (Webb, 2021). It is also argued that lemma involves less discussion than word family on what inflections and derivations should be included (Kremmel, 2021; Schmitt, 2010). However, similar to word type, lemma has limitations in developing learners’ word part knowledge from the perspective of pedagogical value (Webb, 2021). The New Academic Vocabulary List (AVL) (Gardner and Davies, 2014) and the New Medical Academic Word List (Lei and Liu, 2016) are examples of corpus-derived specialised vocabulary list research using lemma as unit of counting.

### *Flemma*

Flemma is a similar but larger unit of counting when compared to lemma. The difference between flemma and lemma is that the former one does not distinguish different part of speech of words but the latter one does (Dang, 2020). For example, flemma counts both “walk” (*n.*) and “walk” (*v.*) together as one item. This is because learners may infer the meaning of either one of these two in texts without too much difficulty, if one of items is known by learners. However, in corpus-derived specialised vocabulary list studies, flemma is not as widely used as other units of counting, the exception being Dang’s work (2017, 2018a, 2018b) on general academic, soft science academic, and hard science academic spoken vocabulary.

### *Word Family*

Word family (or Level 6 word family) is the largest and most widely-used unit of counting in corpus-derived vocabulary list development (Dang, 2020). It includes a headword which is usually the base word form, inflections, and frequently-used and transparent derivatives of the base word form up to Level 6 in Bauer and Nation’s affix taxonomy (Nation, 2022). The full list of affixes included in a word family is divided into seven levels, according to affix frequency and transparency in meaning (Bauer

and Nation, 1993; Bauer and Nation, 2020). This unit of counting was developed based on the assumption that if a learner knows a member of items in a word family then s/he knows all members of items in this word family (Nation, 2013; Vilkaite-Lozdiene and Schmitt, 2020). A typical example of word family as unit of counting in corpus-derived specialised vocabulary list is the AWL (Coxhead, 2000). This list contains 570 word families frequently used in general academic context. However, this unit of counting has been disputed by some scholars (e.g., Laufer, 2021; Snoder and Laufer, 2021). Through assessment, these researchers noticed that there is no strong evidence to guarantee the aforementioned assumption. Also, they (e.g., Laufer, 2021) believed that word family as unit of counting in vocabulary studies may overestimate learners' vocabulary size or provide much heavier workload for learners in vocabulary acquisition. It leads to questioning the suitability of word family in vocabulary list development, and, therefore, the debate on whether to use a smaller unit of counting (e.g., lemma or flemma) in vocabulary acquisition and corpus-derived vocabulary list development studies (e.g., Laufer, 2021; Webb, 2021). However, Nation (2021) suggests that the comparison between word family and lemma or flemma and the discussion of which one is the most suitable unit of counting may not be useful, since they each represent different levels of learners' language proficiency in development. He comments that both lemma and flemma with the affix knowledge they represent in proficiency can be classified within the 6-level word family framework by Bauer and Nation (1993) (e.g., lemma at Level 2; flemma between Level 2 and 3). In this sense, Nation (2022) further suggests that the use of word family as a unit of counting is not a "whether or not" question. Instead, it is a choice for which one is appropriate in a given purpose. In corpus-derived vocabulary list development, Nation (2016, 2022) recommends developing vocabulary lists with the unit of counting that is appropriate for both research aims and learners' proficiency level. He also recommends developing the lists in different formats with different units of counting in order to meet needs of a wide range of audiences. Some existing studies have adapted such method. For example, the AVL (Gardner and Davies, 2014) uses lemma but it is also available in the word family format.

### 2.2.2 Selection criteria

Another principle in developing corpus-derived specialised vocabulary list is the selection criteria of word items. Studies can be classified into two main categories: (1) specialised vocabulary lists with only corpus-driven criteria, and (2) specialised vocabulary lists with both corpus-driven criteria and human involvement criteria. Most of existing studies on corpus-derived specialised vocabulary list are in the first category, using corpus-driven criteria solely (Dang, 2020).

### *Frequency*

The widely-used corpus-driven criteria in corpus-derived specialised vocabulary list development are frequency, range and dispersion (Dang, 2020). Frequency refers to the number of times of a particular item occurring in the whole corpus or sub-corpora. It is one of the dominant criteria in vocabulary list development through corpus, and therefore the corpus-derived vocabulary lists are also called frequency lists or frequency-based lists (Durrant et al., 2022). This is because vocabulary generally follows Zipf's law (Zipf, 1949) that a relatively small amount of vocabulary occurs many times and a large number of items only occur once. Frequency can either be presented through raw frequency or normalised frequency. Raw frequency is the total number of a particular item occurrence in the entire corpus, and normalised frequency refers to how many times an item occurs in a given number of words (e.g., per thousand words or per million words). For example, when developing the AWL, Coxhead (2000) set the criterion that members of a word family should occur at least 100 times in her whole corpus (28.57 times per million). The cut-off point of frequency in corpus-derived specialised vocabulary list is not a fixed threshold, which should depend on the size of target corpus and the research aims. The frequency criterion in the AWL was also adopted by some other studies, due to the great influence of the AWL. It may also help with word list comparison or validation, if similar criterion was used. For example, to develop a medical word list, Wang, Liang, and Ge (2008) compiled a specialised medical corpus with around 1 million tokens. Since the size of their corpus is a third of that developed by Coxhead, they set the frequency threshold as 30 times of occurrence of a word family member in their entire corpus, compared to that of 100 times in Coxhead's (2000) AWL.

### *Range*

Range refers to the number of different texts or subcorpora where the target item occurs (Dang, 2020). For the corpus-derived specialised vocabulary list research, range as a selection criterion helps selecting core vocabulary items that widely spread across the whole particular context. Still taking the AWL as an example, to be included in this list, a member of a word family needs to occur at least 10 times in every subcorpus. Also, the target item needs to occur in at least 15 of 28 disciplines included in the whole corpus (more than 50% of all disciplines included). For her research, range helps to ensure that the selected items are not with limited range of use in the general academic context. Similar to frequency, the threshold of range varies according to research aims and the number of texts or subcorpora included in the whole specialised corpus. For example, Wang, Liang, and Ge (2008) used the same criterion of range on disciplines/areas as Coxhead (2000) did for their medical word list, and had adjustment based on their corpus feature. Specifically, they required that a member of a word family needs to occur in at least 16 of 32 areas in their corpus (50% of all areas included). Likewise, Khani and Tazik (2013) set the criterion of range on 6 of 12 academic journals (50% of all journals selected) included in their corpus

for specialised vocabulary list development on Applied Linguistics. Nevertheless, range is not able to test whether the item is evenly distributed across the whole corpus. This is because when a target item occurs in a specific text or subcorpus section, it only counts once for range, regardless of how many times of occurrence this item has (Gries, 2021). Therefore, it raises the needs of the criterion of dispersion.

### *Dispersion*

Dispersion refers to the evenness of the target item distribution across the whole corpus (Dang, 2020). This criterion, unlike frequency or range with sole calculation method, has different ways in calculation. There are several dispersion calculation techniques in studies on corpus-derived specialised vocabulary list development, for example, Jullian's D, DP and  $DP_{norm}$ , but they all have values ranging from 0 (extremely unevenly distribution) to 1 (extremely evenly distribution). These techniques are used in different studies. For example, Jullian's D was used by Dang et al. (2017), DP was by Barraza et al., (2024) and  $DP_{norm}$  by Deignan et al. (2023). Although dispersion like Jullian's D has receives attention in developing general use vocabulary lists (Dang et al., 2017), dispersion is relatively less used in corpus-derived specialised vocabulary list development research. Gries (2021) also noticed that there are very few studies reporting relations on frequency data and dispersion. Here below are some examples of studies on using dispersion as a criterion in specialised vocabulary list development, where the authors tested different thresholds of dispersion before the final decision. Dang et al. (2017) had nine pilot studies to test Jullian's D in order to find the best cut-off point for her research on general academic spoken vocabulary list, the difference being cut-off point of Jullian's D from 0.1 to 0.9. Through this she noticed that 0.6 as the threshold provides the balance among size and lexical coverage of the vocabulary list in the whole corpus and subcorpora. Barraza et al. (2024) also tested frequency and lexical coverage of items before the decision of dispersion threshold when developing the Applied Linguistics Academic Word List (ALAWL), where they used the Deviation of Proportions (DP) proposed by Gries (2008).

### *Keyness*

Another corpus-driven criterion is keyness, or keyword analysis. It aims to know the "aboutness" of a target corpus (Scott, 1997). Through this, the items with quite different frequency and significance of statistics in occurrence can be identified by checking with a reference corpus (usually a large general use corpus) (Baker, P., 2023). However, compared to those three aforementioned criteria, keyness is much less widely used in specialised vocabulary list development. Deignan et al. (2023) used Cohen's d

and BNC Baby+ for keyword analysis as one of the selection criteria to analyse disciplinary vocabulary of UK Key Stage 2-3.

Relying on corpus-driven selection criteria to develop specialised vocabulary list has two strengths (Dang, 2020; Nation, 2016). One strength is that they are transparent and thus allow replication. With such strength, it allows comparison between vocabulary lists based on different selection criteria and corpora. The AWL is one of the specialised vocabulary lists that is frequently chosen in comparison. It may be the reason that the AWL has great influence on corpus-derived specialised vocabulary list development, whose criteria were adopted by other late studies. It may also be related to the research aim to compare difference between general academic vocabulary and disciplinary-specific vocabulary. The comparison between the AWL and other corpus-derived specialised vocabulary often aims to see how many items overlap in two lists and how many items only occur in the other specialised vocabulary list instead of the AWL. For example, by using similar selection criteria as Coxhead (2000) did, Wang, Liang, Ge (2008) developed the medical academic word list and had comparison of lexical coverage, number of items in the lists and overlapped items between two lists. Another study on technical word list of Computer Science (Chen, H. and Lei, 2019) found that 333 of 769 word types in their list overlap with those in the AWL. These examples show the difference of vocabulary use between general academic and discipline-specific contexts. The other strength is to help with dividing the specialised vocabulary list into sublists through frequency, which sets clear goals for language teaching and learning as a short cut (Dang, 2020). For example, Coxhead's AWL was divided into 10 sublists based on item occurrence in the corpus where it derived. Despite their strengths, vocabulary lists developed with only corpus-driven criteria rely much on the nature of corpus where the lists are derived (Dang, 2020). Thus, it may miss some important items that deserve teacher instructions or students' self-learning.

### 2.2.3 Validation of corpus-derived specialised vocabulary list

Another principle in corpus-derived specialised vocabulary list is the validation of vocabulary lists (Dang, 2020). This is because each corpus-derived vocabulary list is affected by the nature of its corpus, that is, the size of corpus, the texts contained in the corpus and methods of sampling in corpus compilation. Testing vocabulary lists through another independent corpus is a valid assessment for relevant studies because such validation method is not biased towards the corpus where the vocabulary lists are derived (Dang, 2020; Nation and Webb, 2011).

The validation of corpus-derived specialised vocabulary list should meet three criteria (Dang, 2020):

(1) the vocabulary list needs to be validated in another independent corpus, rather than the corpus where the list is derived;

(2) the corpus for validation should be in similar size as that of the corpus for vocabulary list development; and

(3) the corpus for validation should have the same source of texts and sampling method as that of the corpus for vocabulary list development.

This suggests that researchers need to compile two similar corpora: one for vocabulary list development and the other for vocabulary list validation. In the process of validation, the items in the lists are checked in the validation corpus, mainly for lexical coverage, range and dispersion. In other words, it aims to test whether a certain item in the list provide similar lexical coverage in two corpora and whether the whole list provide similar vocabulary use in statistics. Since the two corpora are developed based on similar sampling methods, with the same source of texts and in similar size, the vocabulary list is considered as a good representation if lexical coverage and other quantitative tests of the vocabulary list are similar in comparison. However, based on Dang's (2020) and Dang and Webb's (2025) review, around half of existing studies have not reported the validation of corpus-derived vocabulary lists. This suggests that this principle is more on theoretical aspect but may be neglected much in research practice. One of the possible reasons is the time-consuming of corpus compilation. Due to the fact that the corpus in analysis become larger in size with the advancement of technology and computer science, researchers may need to spend more time and energy in compiling two corpora, compared to that in compiling one for vocabulary list development only. Another possible reason is the lack of source of texts, especially for specialised vocabulary list development. There could be no sufficient resource available for compiling two specialised corpora in similar size of millions of tokens to generate reliable results for both vocabulary list development and validation (e.g., Drayton and Coxhead, 2023).

Among those studies which have the validation process, most do not meet all the aforementioned criteria (Dang, 2020). They either used a smaller corpus for validation, compared to that corpus where the vocabulary list was derived, or they did not validate the list with corpora of the same genre. For example, the AWL was developed based on a 3.5-million-token general academic corpus (Coxhead, 2000); however, it was validated in a corpus with a size of 678,000 tokens, which is much smaller than the corpus for vocabulary list development. Similarly, the corpus for developing the AVL was 120 million tokens in size, while its validation corpus is from 16 to 83 million tokens in size (Gardner and Davies, 2014). One of the few exceptions is Dang et al.'s (2017) ASWL, which was developed and then validated in two independent corpora with similar size of 13 million tokens. Compared to general academic vocabulary list validation through corpus, corpus validation in discipline-specific or technical vocabulary list research is much in scarcity.

An alternative method for validating specialised vocabulary lists is to test them in general use and/or general academic corpora, if the resource of specialised texts is not sufficient or available. For example,

when analysing radiotelephony on technical vocabulary, relevant resource is rare for availability, which poses difficulty for compiling another specialised corpus in the similar size in such context (Drayton and Coxhead, 2023). Therefore, in order to have validation for their technical vocabulary list, Drayton and Coxhead (2023) used three corpora of other genres as alternatives (a written fiction corpus, a general academic written corpus, and a general spoken corpus). Their lexical coverage analysis results of the radiotelephony technical vocabulary list in these three corpora showed a significant difference (around 3-5%), compared to that in their radiotelephony corpus for vocabulary list development (more than 88%). They thus suggest that such technical words are less likely to occur in written fiction, written academic or general spoken English.

However, the test on lexical coverage is only one of the aspects of vocabulary list validation, especially for the pedagogical value of vocabulary lists. Nation (2016) put forward a framework on vocabulary list validation, which covers eight aspects. Purpose of the vocabulary list development, as the first aspect, focuses on who the target learners can be and what the vocabulary knowledge this list covers. Unit of counting, main list and additional list attached are the following three aspects, discussing whether and to what extent the list is relevant to such target learners. The validation also includes evaluating the corpus and criteria used for vocabulary list development. Finally, developers need to self-criticise their lists and discuss availability of lists, which requires to acknowledge limitations of lists and be open for further evaluation. By using this framework, Drayton and Coxhead's (2023) study evaluated their technical vocabulary list on radiotelephony. However, this evaluation was conducted by the researchers themselves, which may have some bias and may not suitably reflect thoughts of users (e.g., practitioners, learners, textbook developers and policymakers) (Dang and Webb, 2025). Having these stakeholders involved in vocabulary list evaluation is one of the ways to address this issue (Dang et al., 2022). For example, He and Godfroid (2018) invited ESL teachers to rate difficulty and usefulness of items in a list occurring in COCA-academic corpus to identify words with priority in teaching. Nevertheless, stakeholder involvement in vocabulary list evaluation is still in scarcity in research of recent years.

#### 2.2.4 Approaches to developing corpus-derived specialised vocabulary lists

The selection criteria and techniques for developing corpus-derived specialised vocabulary lists have been more systematic and sophisticated in the past two decades. After consideration of selection criteria, another issue in corpus-derived specialised vocabulary lists is the decision of approaches. Currently, there are three approaches in related studies, and they are all form-focused approaches.

### *Developing vocabulary lists based on other ready-made lists*

The first approach is to develop vocabulary lists based on other general use vocabulary ready-made lists. In other words, a reference vocabulary list is chosen, and all items in this list will be excluded when developing the specialised vocabulary list. This approach is based on the assumption that students have acquired general high-frequency vocabulary (Dang and Webb, 2025). Commonly, the reference vocabulary list is general use list with well-developed criteria or good reputation in research or pedagogy (e.g., the GSL, BNC2000, BNC/COCA 2000). This approach, taking learners' vocabulary knowledge into consideration, targets learners with at least basic vocabulary proficiency and knowledge (e.g., students who have acquired the most frequent 2000 words), and thus, there is no need for them to learn these words again in limited classroom time for instructional efficiency. Vocabulary lists developed based on this approach aims to avoid repeated teaching and learning for saving time (Dang, 2020). In addition, Martinez et al. (2009) commented that some items in the GSL may have interference on general academic vocabulary acquisition, since these items have similar rhetorical functions. With such items excluded, teachers and students in academic context can allocate more attention to vocabulary that are relatively new to them and frequently used in target contexts. The typical example is the AWL by Coxhead (2000). This list did not include items occurring in the GSL by West (1953), since the AWL targets university students with learning needs in academic contexts who have acquired the most frequent vocabulary. This approach is also used in disciplinary corpus-derived academic vocabulary list development (e.g., Liu and Han, 2015; Vongpumivitch et al., 2009; Wang, Liang, and Ge, 2008; Yang, 2015), but is less used in technical word list development.

However, there are some limitations in excluding items in the GSL or other general high-frequency vocabulary. General high-frequency vocabulary may be polysemous words (McCarthy, 2023; Stubbs, 2001), having multiple senses, and senses used in specific contexts may be different. As discussed in the previous section, early studies (e.g., Chung and Nation, 2003; Hyland and Tse, 2007) have noticed that general high frequency words may also have discipline-specific or technical senses. This suggests that, if the general high-frequency words are excluded from specialised vocabulary list by focusing on form, the list may miss items that are important for learners in specialised contexts. Also, the quality of ready-made vocabulary lists will affect the quality of vocabulary lists that are developed based on previous ready-made lists. Taking the GSL as an example, although it is less likely for a huge change of general high-frequency words within a relatively short period of time (McCarthy, 2023; Nation, 2022), the GSL indeed does not contain some modern general high-frequency vocabulary such as “television” and “victim”, with some out-of-dated items included (e.g., “shilling”) (Nation, 2013, 2022). This indicates the out-of-date feature of the GSL to a certain degree. Therefore, bias may exist if vocabulary lists are developed based on the GSL by excluding its items. Furthermore, recent studies on students' vocabulary proficiency shows that learners have insufficient knowledge on general high-frequency vocabulary, such as the most frequent 2000 words, especially for EFL learners who may lack

of sufficient input (Al-Masrai and Milton, 2012; Dang, 2020; Kitano, 2018; Lu and Dang, 2022; McLean, Hogg and Kramer, 2014). In this sense, taking the assumption that learners have acquired general high-frequency vocabulary for granted and excluding such vocabulary from lists may lead to bias in specialised vocabulary list development.

### *Developing lists from scratch*

To address these issues, Gardner and Davies (2014) put forward the second approach for developing corpus-derived vocabulary lists. They developed the AVL from scratch (Dang, 2020). In other words, they did not develop the AVL based on any other ready-made vocabulary lists. Any items that meet their corpus-driven criteria are included in the AVL, regardless of being general high-frequency vocabulary or presented in other vocabulary lists. This approach does not take learners' vocabulary knowledge for granted, and therefore, may be suitable for a wide range of users. Moreover, as this specialised vocabulary list does not rely on other ready-made vocabulary lists, its quality is not affected by the that of other lists.

However, the limitation of this approach is that it may cause repeated teaching or learning in practice. Because developing vocabulary lists does not rely on any other ready-made lists and has no assumption on students' vocabulary knowledge, a number of items may have been learned or repeated in general use vocabulary lists (Gong, et al., 2025). Gong et al. (2025) found that 351 AVL items are repeated in both the GSL (West, 1953) and the New General Service List (NGSL) (Browne et al., 2013). This, therefore, may lead to less sufficient teaching or learning time allocated to those that students have learned partially or not learned in previous stages. Lei and Liu (2016) adapted this approach in developing Medical Academic Vocabulary List, but they had a double check on word meaning of general high frequency words. Specifically, they developed their vocabulary list from scratch with corpus-driven criteria (e.g., frequency, range, dispersion). Then they compared the list with the Brezina and Gablasova's (2015) New General Service List. If any item occurred in both lists, they had disciplinary dictionary checking on the meaning of that item in two medical dictionaries. If that item had disciplinary meanings in two medical dictionaries, the item would be kept. They are based on the assumption that their list users should have learned general meaning of these words. In this way, they called for attention to general high-frequency words kept in their medical academic vocabulary list, because these words have meanings directly related to medical disciplines. Hence, repeated teaching or learning can be largely reduced.

### *Developing lists with a combined approach*

As each approach has its own advantages to balance their limitation, Dang et al. (2017) put forward the third approach. They developed the ASWL from scratch so that the list does not miss any potential important items. After that, they classified items in the ASWL into four levels according to their general use frequency so that the list can accommodate students' language proficiency levels. Specifically, the BNC/COCA 25 base word lists (Nation, 2012) were used as reference lists representing general use vocabulary frequency. Each of the base list contains 1,000 word families according to frequency. Items in the Level 1-3 of the ASWL were classified based on the first, second, and third 1,000 BNC/COCA base word lists, which represented items that are frequently used in the academic spoken contexts and are also general high-frequency vocabulary in general language. The Level 4 contains all other items outside the 1<sup>st</sup> – 3<sup>rd</sup> 1,000 BNC/COCA word lists. This approach links specialised vocabulary and general language by general use vocabulary frequencies. Through this, teachers and students can shift their attention to levels that are suitable for their proficiency, based on pre-session vocabulary size/knowledge tests. For example, if a pre-session vocabulary test shows that the vocabulary size of learners is about 2,000 word families, they could quickly go through the Level 1 and 2 to check the acquisition of vocabulary knowledge of these items. They then could allocate more attention to Level 3 and 4 which are beyond their current proficiency levels. This approach helps with avoiding repeated teaching for saving time and promoting adaptability for learners with different proficiency levels. This approach was also adopted by Lu (2018) to develop technical vocabulary list on Traditional Chinese Medicine. In her study, she had three levels and nine sublists, which also had reference to BNC/COCA 25 base word lists, classified at Level 1 and 2 respectively. Differently, she had loan words at Level 3. This is reasonable since Traditional Chinese Medicine context has hundreds of loan items with cultural-based features such as “qi” and “dao” which are not used in general use English. It balanced adaptability of vocabulary lists for learners, technicality of items in the list and general use vocabulary frequencies.

## 2.3 Corpus-derived specialised vocabulary lists

Specialised vocabulary list development has a long history which can be dated back to second half of last century. Early specialised vocabulary list research mainly focused on academic context, through either observation of students' learning difficulty of certain words (Ghadessy, 1979; Lynn 1973) or frequency calculation in a small corpus of university textbooks (Campion and Elley, 1971, cited in Nesi, 2002; Praninskas, 1972). The advancement of corpus analysis software such as AntConc (Antony, 2024) and LancsBox (Brezina et al., 2021) helps corpus compilation and development of more representative specialised vocabulary lists (e.g., discipline-specific contexts, vocational contexts or technical contexts). Corpus-derived specialised vocabulary list studies can be classified into three categories: general academic vocabulary lists (section 2.3.1), discipline-specific academic vocabulary lists (section 2.3.2)

and technical vocabulary lists (section 2.3.3). Again, only specialised vocabulary lists on individual words in written context are reviewed here, since this search has a special focus on individual words. Spoken lists are only mentioned when they provide methodological insights for this research.

### 2.3.1 General academic vocabulary lists

Most early general academic vocabulary lists were not developed based on large size corpus with a wide range of academic texts. For example, Champion and Elley (1971) and Praninskas (1972) developed academic vocabulary lists from a corpus of several available university textbooks respectively, with limited disciplines and register involved. Another issue of these early general academic vocabulary lists is the unclear statement of unit of counting. Also, these lists are not widely available in research and teaching practice. Another example of early general academic vocabulary list is the University Word List (UWL) (Xue and Nation, 1984). This list has 836 word families included and was adapted from four previous lists: Champion and Elley (1971), Praninskas (1972), Lynn (1973), and Ghadessy (1979), of which only the first two were developed from corpora. The UWL provides around 8.5% lexical coverage in the Learned and Scientific section of the Lancaster-Oslo/Bergen (LOB) corpus (Xue and Nation, 1984). However, since the UWL was adopted from four ready-made lists and not directly generated from corpus data, selection criteria were not consistent across four lists.

The AWL in Coxhead (2000) replaces the UWL. This list consists of containing 570 word families, selected from a 3.5-million-token academic corpus. The corpus consists of four subcorpora (Arts, Commerce, Law and Science) covering 28 subsumed disciplines. Frequency (100 times of item occurrence), range (15 of 28 disciplines) and specialisation (exclusion of the GSL items) were key selection criteria of the AWL development. For teaching practice and pedagogical implications, the AWL was divided into 10 sub-lists based on vocabulary frequency in the corpus to set learning goals with priority. It is one of the most widespread general academic vocabulary lists because of its strengths on free availability and easy applicability (Dang, 2020; Coxhead, 2016; Nation, 2011). However, this list also receives criticisms on its exclusion of the GSL items (Gardner and Davies, 2014), not every member of a word family being frequently used in academic contexts (*ibid*), lack of further exploration on disciplinary vocabulary (Hyland and Tse, 2007), and some bias towards Humanity-based disciplines in corpus compilation.

After the AWL, there were some more updated general academic vocabulary lists. Paquot (2007) argued that keyness and evenness of distribution of items should be considered in corpus-derived vocabulary list development. This is because “keyness not only extracts highly frequent words in the academic corpus but also gives prominence to discipline or topic-related vocabulary which is not necessarily frequent in the academic corpus” (*ibid*, p.129). This was not included by Coxhead (2000)

in developing the AWL. Paquot (2007) also argued that the exclusion of the GSL items may lead to missing some items with different uses in general language and academic contexts. With criteria of keyness, frequency, range and evenness distribution applied, Paquot (2007) developed the Academic Keyword List (AKL) containing 838 keywords. However, the corpus size used to develop the AKL (1 million tokens) is much smaller than that for the AWL (3.5 million tokens). Also, the AKL only listed the keywords, rather than word families or flemmas or lemmas. Another word list is the New Academic Word List (NAWL) (Browne, Culligan and Phillips, 2013a). It is derived from a 288-million-token academic corpus covering different academic registers such as journal articles, students' essays, and bestselling academic textbooks published in recent decades. This corpus also includes academic spoken discourse from the Michigan Corpus of Academic Spoken (MICASE) and the British Academic Spoken English (BASE) corpus, which is different from that of the AWL and the AKL specifically focusing on written academic discourse. The NAWL contains 957 lemmas selected by frequency, dispersion and exclusion of items in the New General Service List (NGSL) (Browne, Culligan, and Phillips, 2013b). It provides a much higher lexical coverage (92%) than that of the AWL in the corpus where the list was derived, and has one English definition for each word. However, the NAWL was also developed based on a ready-made list whose quality may affect the quality of the NAWL, although such ready-made list was derived from updated materials. In addition, the development and validation of the NAWL were not clearly described and lack transparency.

Recognizing the potential limitations of developing general academic vocabulary list based on ready-made lists, Gardner and Davies (2014) put forward a new approach of developing lists from scratch (also see section 2.2.4). The development of the Academic Vocabulary List (AVL) based on a 120-million-token academic corpus in the COCA. Also, it considered frequency ratio (the occurrence of items in academic corpus should be 1.5 times higher than that in the non-academic corpus), range, dispersion, and discipline measure (occurrence of items should be no more than three times of expected occurrence in any one of nine disciplines). Frequency ratio is used to remove general high-frequency words that do not meet their high frequency in general academic English (Lei and Liu, 2016), and discipline measure is to remove discipline-specific or technical words because their focus is on general academic vocabulary. Another difference between the AWL and the AVL is that the AVL used lemma as unit of counting, with 3,015 lemmas included, which addresses one of the criticisms of the AWL that not every member of a word family is frequently used in the academic contexts.

### 2.3.2 Discipline-specific academic vocabulary lists

Based on the call for research by Hyland and Tse (2007) that a general academic vocabulary list may not be suitable for students from different disciplines in language learning for specific purposes,

discipline-specific academic vocabulary list research flourishes. There are two sub-categories of discipline-specific academic vocabulary list research, due to the difference of definition on “discipline”.

The first sub-category of research has a broad view in defining disciplines. I shall call it broad-view discipline-specific academic vocabulary list in the following paragraphs. It defines discipline as a group of related academic subjects (Dang, 2020), for example, subjects like Mathematics, Engineering, Physics and Chemistry are subsumed into the umbrella of Sciences. This is based on the assumption that these subjects are closely related to each other and may have shared knowledge, and thus the vocabulary in use may also be shared across subjects. For example, “factor” and “method”, as Baker, M. (1988) states. Most of the broad-view discipline-specific academic vocabulary lists are on written registers such as textbooks and academic articles, the only two exceptions being Dang (2018a, 2018b), as the summary in Table 2.1 below. This is reasonable since the spoken corpus compilation is more time-consuming with less data available (Knight and Adolphs, 2022), so is specialised spoken corpus compilation (Koester, 2022). Apart from dominance of written academic registers in these lists, three more features are observed through these studies: corpus size, data source and research focus, and word selection criteria.

First, on average, the corpus size for developing broad-view discipline-specific academic vocabulary lists ranged from 0.4 to 6.5 million tokens, which is smaller than that for general academic vocabulary lists. This may be the availability of texts for corpus compilation, since the discipline-specific academic vocabulary lists cover a narrower range of disciplines when compared to those in general academic contexts. Data availability is one of the difficulties for corpus compilation (McEnery and Brookes, 2022). Second, most studies available and summarized in the Table 2.2 above focus on disciplines on hard sciences (e.g., mathematics, physics, statistics, engineering, biology and chemistry), although there exists difference on specific disciplines covered in each study. Only two (Chanasattru and Tangkiengsirisin, 2016; Kwary and Artha, 2017) have a special focus on social sciences or soft sciences. This may be a bias toward research focus. Third, word selection criteria for developing these vocabulary lists have a preference on pure corpus-driven selection based on corpus data. The only exception was It-ngam and Phoocharoensil (2019), which invited a panel of three experts in science disciplines to rate the items selected by corpus-driven criteria, with a rating scale adapted from Chung and Nation (2004). In addition, three vocabulary lists were developed based on ready-made lists, for example, the exclusion of items in the GSL (Veenstra and Sato, 2018), and the exclusion of items in both the GSL and the AWL (Coxhead and Hirsh, 2007; It-ngam and Phoocharoensil, 2019). This method, as discussed earlier, may have missing of important items which are frequently used in both general use or general academic contexts, and the vocabulary list quality is affected by that of previous lists used in selection process.

Table 2.1 A summary of broad-view discipline-specific individual vocabulary lists

Name of Word Lists	Counting Units	Methods Adopted	Corpus Size, List Availability and Other Notes
EAP Science Word List (Coxhead and Hirsh, 2007)	Word family	Corpus-based	1.5 million tokens
Social Science Word List (SSWL) (Chanasattru and Tangkiengsirisin, 2016)	Word family Word types	Corpus-based	0.4 million tokens
Academic Article Word List for Social Science (AAWL-SS) (Kwary and Artha, 2017)	Word family	Corpus-based	1 million tokens
Science Textbook Word List (Veenstra and Sato, 2018)	Word family	Corpus-based	0.7 million tokens
Science Academic Word List (It-ngam and Phoocharoensil, 2019)	Word family	Combined	5.5 million tokens
Scientific Research Article Word List (SRAWL) (Kanzaki, 2019)	Lemma	Corpus-based	1.7 million tokens only focus on titles and abstracts including four sublists
Ontario High School Science Word List (OHSWL) (Mahfouz, 2021)	Lemma	Corpus-based	3.23 million tokens

A second sub-category of discipline-specific vocabulary lists has a narrower definition on discipline, which focuses on a specific discipline, or subject. Therefore, they are also mentioned as subject-specific for field-specific vocabulary lists in literature (e.g., Dang, 2020). The underlying idea is based on the call of Hyland and Tse (2007) that vocabulary may vary in specific discipline contexts and senses may also be different. To have a comparison of the concept of broad-view discipline-specific vocabulary lists, in this review I named them as narrow-view discipline-specific vocabulary lists. These vocabulary lists have a special focus and are more suitable for a particular group of learners in ESP field. One of the earliest research projects is the Medical Academic Word List (MAWL) (Wang, Liang and Ge, 2008). Their list contains 623 word families selected from a 1-million-token specialised corpus on medicine

research articles. The selection criteria were similar to those in Coxhead (2000), the differences being specific context focused and cut-off points. More than 50 narrow-view discipline-specific vocabulary lists on corpus are available in the last two decades. The number of published studies in each year is presented in Figure 2.1 below.

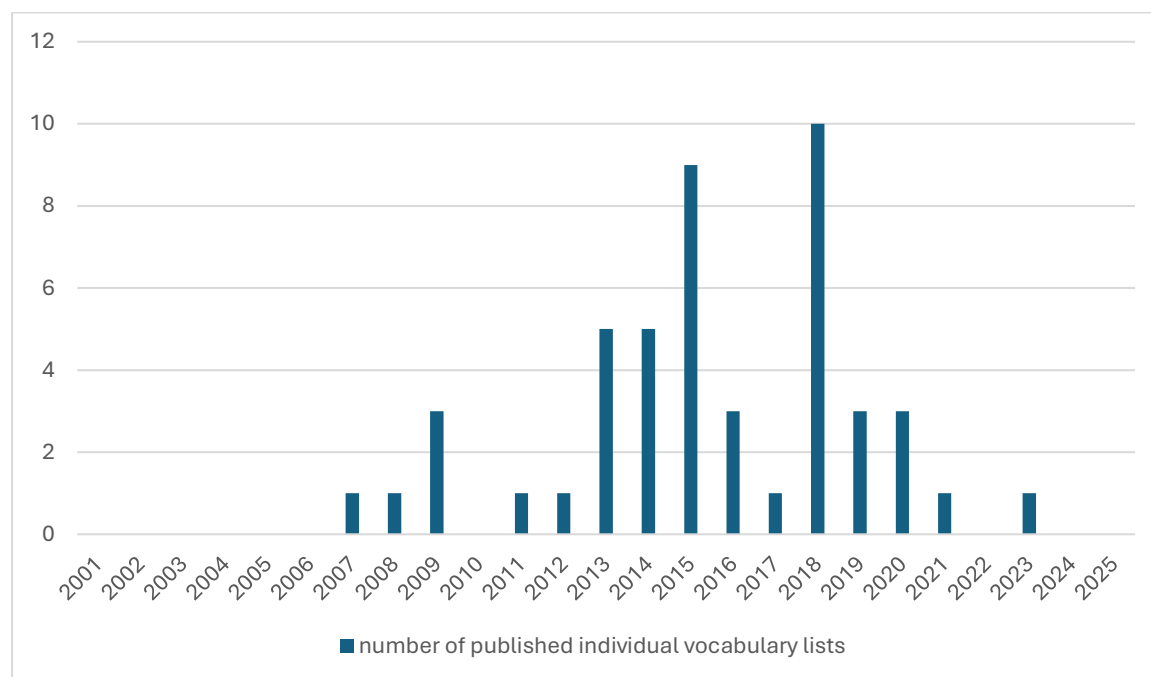


Figure 2.1 Number of published corpus-derived narrow-view discipline-specific individual vocabulary lists (from 2000 to 2025)

The data shown in this Figure indicates that the number of studies increases from 2007 to 2018, reaching a summit in 2018 ( $n = 10$ ), and decreases till now. Existing studies cover a wide range of disciplines, including but not limited to Accounting, Business, Computer Sciences, Chemistry, Education, Engineering, Food Science, Geography, Linguistics and Applied Linguistics, Medicine, Nursing, Pharmacy and Zoology. Although they cover more than a dozen of specific disciplines, studies mainly focus on Business, Computer Science, Engineering and Medicine. Each of four disciplines has at least three or four related lists available.

Narrow-view discipline-specific vocabulary list research also has similar issues as those of broad views mentioned above: written discourse in dominance and corpus-driven selection criteria in preference. Almost no related spoken vocabulary lists available. It can be the reason of lack of data availability in corpus compilation, or similarities of such discipline-specific vocabulary shared between spoken and written contexts. In addition, few studies have human experts' involvement for word selection (e.g., Martinez, Beck and Panza, 2009; Kruawong and Phoocharoensil, 2020; O'Flynn, 2019;

Valipouri and Nassaji, 2013; Watson-Todd, 2017). This suggests the majority of related studies using corpus-driven criteria for word selection. Different from broad-view discipline-specific vocabulary lists, the bias of focusing on science disciplines is reduced to some extent in studies of narrow-view discipline-specific vocabulary lists. There are studies on different social science disciplines, for example, Business (Hsu, 2011; Hu, Chen, and He, 2019; Konstantakis, 2007; Tangpijaikul, 2014); Applied Linguistics (Khani and Tazik, 2013; Vongpumivitch et al., 2009), Education (Freund, 2014; Mozaffari and Moini, 2014), and Linguistics (Lee and Kim, 2020). Another difference is the size of corpus. Corpus size varies in developing narrow-view discipline-specific vocabulary lists, ranging from less than 0.5 million tokens (e.g., Freund, 2014; Sadeei, Khany and Tazik, 2023; Ward, 2009) to more than 10 million tokens (Hsu, 2013; O'Flynn, 2019; Xodabande and Xodabande, 2020). This may also be the reason of data availability, since there could be different amount of data published and available in different specific disciplines. For example, Business and Medicine, two major disciplines closely related to human life, have more texts available, which contributes to larger specialised corpus compilation (e.g., 15 million tokens in Hsu, 2013 for Medicine). However, these studies have a particular focus on students or readers in higher education, where their text sources are mostly from textbooks or academic journal articles. Comparatively, they have a less focus on younger learners who have learning needs in specific disciplines (e.g., middle school subjects). Only two series of research (Green and Lambert, 2018; Greene and Coxhead, 2015) have a special focus on middle school subjects for vocabulary lists, where they collected textbooks and materials used in several subjects at schools for corpus compilation.

### 2.3.3 Technical vocabulary lists

Research on extracting technical vocabulary through electronic corpus dates back in 1980s. For example, Yang (1986) and Baker, M. (1988) had trials on sub-/technical vocabulary extraction in specialised corpora. Corpus-derived vocabulary list development also witnessed a trend of shifting focus from general use or general academic contexts to more specialised contexts (Green and Lambert, 2018). Since technical vocabulary lists are related to terminology, they, therefore, may have a higher requirement to professional knowledge and target to a specific group of learners or users, usually in professionals or trades. In this sense, existing studies have a much more various topics when compared to those in discipline-specific vocabulary lists, for example, Culinary course (Nordin et al., 2013), Wine taster (Hou, 2014), Traditional Chinese Medicine (Lu, 2018), Summer Olympics (Riccobono, 2019), Equine Veterinary (Safari, 2019), Basketball (Riccobono, 2020), Tourism (Tabrizi and Jamalzadeh, 2020), Carpentry (Coxhead and Demecheleer, 2018; Coxhead et al., 2020), Airlines, Hotels and Tourism Business (Laosrirattanachai and Ruangjaroon, 2021), Rugby (Benson and Coxhead, 2022), Radiotelephony (Drayton and Coxhead, 2023). One of the typical examples is Coxhead et al.'s (2020) research on four vocational trades: Carpentry, Fabrication, Automative and Plumbing. They collected

written materials (e.g., prints and notes) and spoken data (e.g., classroom audios and field work recordings) as data source for compiling a 1.6-million-token specialised corpus, and selected words with both corpus-driven and human-involvement criteria. Their research was one of the studies that covers both written and spoken data, and corpus-driven and human-involvement criteria in corpus-derived vocabulary list development. Their vocabulary selection approach was also followed by recent studies (e.g., Lu, 2018; Benson and Coxhead, 2022; Drayton and Coxhead, 2023).

Similar to other categories of corpus-derived specialised vocabulary lists, corpus-derived technical vocabulary lists also have limitations of written text in dominance and variation in corpus size. However, there are three points in difference: unit of counting, having expert's judgement in word selection as a common technique, and inclusion of multi-word units in vocabulary lists. First, the decision of unit of counting has a shift from word family to word type. As discussed in section 2.2, word family is one of the most popular units of counting in corpus-derived vocabulary list research. In corpus-derived technical vocabulary lists, researchers have a preference of using word type as unit of counting. This is because not every member of a word family or lemma conveys technical sense in a specific context (Coxhead et al., 2020), and the technical senses were usually conveyed through specific word form (Drayton and Coxhead, 2023). For example, when analysing vocabulary in radiotelephony, Drayton and Coxhead (2023) noticed that "cleared" was frequently used in such spoken context as a part of technical multi-word unit, rather than other word form such as "clear" or "clearing" in the same word family. Word type, in technical contexts, may be more suitable than word family as the unit of counting to explore the technical feature of vocabulary. Second, compared to specialised vocabulary list development in general academic or discipline-specific context, researchers are more likely to invite experts from the target context or trade to have judgements on word selection. More than half of aforementioned studies have this process, which shows their difference in word selection. This may be the reason that researchers are experts in language fields but not professionals in the target technical context. In order to select representative technical vocabulary for particular groups of learners, professional knowledge and experts' assistance are needed (Coxhead et al., 2020; Ha, 2015; Quero, 2015). Third, corpus-derived technical vocabulary list in the recent few years (e.g., Benson and Coxhead, 2022; Coxhead et al., 2020; Drayton and Coxhead, 2023; Lu, 2018) also have a focus on technical multi-word units. Most general academic and discipline-specific vocabulary lists focus on individual vocabulary, making it and multi-word unit vocabulary lists as two separated independent studies. Different from those, these technical vocabulary list studies analyse vocabulary from both individual words and multi-word units in the same research. For example, Coxhead et al.'s (2020) research on four trades has individual vocabulary lists, and provided a list of technical lexical bundles for each trade respectively. This is because, in specialised technical contexts, meaning can be expressed through both individual words and multi-word units, and the multi-word units are, as a whole, in expression, whose meaning is not the sum of meaning of the components.

## 2.4 Further development: Working beyond forms

This chapter reviews corpus-derived specialised vocabulary lists, starting with a discussion of what specialised vocabulary is. The concept of specialised vocabulary covers academic and technical vocabulary; however, there is no clear boundary between these two categories: an academic word may convey discipline-specific or technical senses in a specific context whose sense may be quite different from other contexts. Due to the high lexical coverage of specialised vocabulary and its difference in specific contexts, specialised vocabulary is of significance in research, especially being useful for learners with learning needs in specialised contexts. It is a key element for communication in specialised context and community identity establishment. This chapter then reviews principles and techniques for corpus-derived specialised vocabulary development. It covers issues of unit of counting, how to address other words (e.g., abbreviations and loan words), corpus-driven criteria (frequency, range, dispersion, keyness), and three approaches for vocabulary list development in existing research. Finally, it reviews existing studies in the last two decades, based on categories of general academic, discipline-specific (broad view *vs.* narrow view) and technical vocabulary lists.

The review shows that corpus-derived vocabulary list research has covered a wide range of topics or disciplines, with a trend from more general contexts to specific contexts (Green and Lambert, 2018). Another issue that widely exists in almost all corpus-derived vocabulary lists is the focus of form. The criterion of deciding to include general high-frequency words by checking whether they have disciplinary meaning through discipline-specific dictionary in Lei and Liu (2016) provides a potential direction for going beyond word forms in corpus-derived specialised vocabulary list development. Thanks to the advancement of computer science technologies and corpus linguistic research, corpus-derived vocabulary lists can be generated within a short time when the corpus is input on software. However, few existing studies on corpus-derived vocabulary lists consider other aspects of vocabulary knowledge (e.g., word meaning and word use), especially for polysemous words having different senses in contexts (It-ngam and Phoocharoensil, 2019). Deignan et al. (2023) pointed out that such items can be a threat to corpus-derived vocabulary lists, since the word sense in a specific context can be quite different from that in general use and academic contexts. Polysemy, a pervasive linguistic phenomenon, may pose challenges to learners (Deignan et al., 2023). This will be reviewed in Chapter 3 Polysemy, the second part of literature review.

## Chapter 3 Polysemy

### 3.0 Introduction

This chapter reviews studies of polysemy in a specialised context, with a special focus on corpus linguistics as methodology in analysis. It consists of four sections. Section 3.1 discusses the definition of polysemy, distinguishes polysemy from homonymy, and sense from meaning. In Section 3.2, I discuss the issue of polysemy as a linguistic barrier of language learning in specialised contexts. Section 3.3 reviews corpus linguistic methodology and approaches in investigating polysemy, followed by review of studies on analysing polysemy in specialised contexts through corpus in Section 3.4.

### 3.1 Polysemy and related concepts

In semantics, meaning and sense of words are two closely related but distinguished terms. Meaning refers to the concept denoted by the language expression (Curse, 2002), while sense “is a discrete representation of one aspect of the meaning of a word” (Jurafsky and Martin, 2025, Appendix G, p.1). In other words, meaning of words is the definition of words, and sense of words is the interpretation of meaning of words based on the specific context where it occurs. Words that have more than one related sense are described as polysemous words (McCarthy, 2023). Some words have quite different and unrelated senses but represented in the same word form, which are considered to be homonyms (Nation, 2022). These unrelated senses usually come from different language sources due to historical reasons (ibid). For example, the homonyms “bank” meaning “financial organization and branches” in the phrase “the national bank”, and “bank” meaning “raising ground near waters” in the phrase “bank of a river” originates from different languages. Although they are both nouns, the first is from Latin or Old Italian in late 15<sup>th</sup> century or even earlier, while the second one is from Old Norse or Old Danish which can be dated back to 12<sup>th</sup> century, based on etymology (Oxford English Dictionary, available from: [https://www.oed.com/dictionary/bank\\_n1?tab=etymology](https://www.oed.com/dictionary/bank_n1?tab=etymology)).

“house” is polysemous in the following sentences:

- (1) It is a small house at LS6.
- (2) Harewood House is near to Leeds.
- (3) The stranger’s shouting at midnight woke up the whole house.
- (4) This building housed thousands of staff of MI6 during the WWII.

Through looking up the COBUILD dictionary (2023), “house” in each of four sentences above has different senses in contexts. In (1), it refers to “a building”. In (2), it means “an expensive private home”, usually in British English. In (3), it is “the whole family living in the same building”. In (4), although

it is a verb, its meaning is “to stay or to live in”. Although the senses of “house” are different in these sentences, they are related to the concept of “building”, to some extent.

However, although in linguistic theories polysemy and homonymy are different concepts, the remainder of this thesis will not distinguish them, and will use “polysemy” for multiple meaning regardless of etymology. There are three reasons for this. First, the origins or language sources of some word sense may not be clear enough due to the long passing of time. The cut between relatedness and unrelatedness of word sense in concepts of polysemy and homonymy is much clearer in theory than in practice. Mojela (2007) writes that the decision of relatedness and unrelatedness of senses is a rather subjective issue, which depends on vision of knowledge of decision-makers such as lexicographers. Second, one of aims in this thesis is to identify different word senses of specialised Logistics vocabulary, and have comparisons between these senses and those in general academic or general use contexts. Distinguishing whether a word is a polysemous word or a homonym is not relevant to the aim of this thesis. Third, from a more practical issue as an EFL teacher and learner, EFL teachers and learners who do not study Linguistics may not be interested in where the word comes from and how its senses develop through history. What they are concerned with is whether word senses would cause confusion in context.

### 3.2 Polysemy as a linguistic barrier in specialised language learning contexts

Polysemy is a common linguistic phenomenon due to the economy of expression (Falkum, 2015), and the law of least effort (Crossley et al., 2010; Deutchman, 2021; Gyori, 2002; Powers, 1998). Most of everyday words in English have multiple senses (McCarthy, 2023), and only very few have one sense, as monosemous words in very technical context (Sinclair, 1991). Due to the large amount of polysemous words and senses, Thornbury (2002, p.53) commented that it is “a complete headache for students”.

Acquisition of any word is a cumulative process, rather than one-time-and-go process (Nation, 2022; Schmitt and Schmitt, 2020). It also applies to word sense knowledge acquisition. Therefore, students sometimes may be confused when encountering unfamiliar word senses in context, due to either partial linguistic knowledge or unfamiliarity with specific contexts or both. Parent (2009) summarized two possible outcomes of students’ encountering unfamiliar senses of polysemous words in contexts. First, they may misinterpret the sentence, because they believe that they know the word senses, and the senses they know work in that sentence, at least to some degree. Laufer (1997, p.26) also reported that “it often happens that students know one meaning of a polyseme or a homonym, and are reluctant to abandon it, even when, in a particular context, its meaning is different. [...] The mistaken assumption of the learner [...] is that the familiar meaning is the only meaning.” Second, the communication may break down completely because the sense that students know does not work at all within the context, and the context

does not provide sufficient clues for them to infer the meaning of words. These two outcomes are based on the issue that students know, or partially know, the word form and its word sense. Another situation is that students know neither the word form nor the word sense. They may look up the word in dictionaries, but dictionaries may provide dozens of entries of senses for a word form (Parent, 2009). Students then need to check the senses provided by dictionaries and contexts, but there is no guarantee of students going through every sense entry or sufficient contextual clues for solving the problems (Parent, 2009). Therefore, for their difficulties on polysemous word sense in specific contexts, some may be solved, while some may be in either of two outcomes mentioned above. This also happens especially when students are in unfamiliar specialised contexts, and when general use dictionaries do not clearly list technical senses of words.

### *Polysemy recognized as an issue in specialised contexts*

Polysemy as an issue has been recognized in specialised contexts (Nagy and Townsend, 2012). Baker, M. (1988) listed different categories of polysemy in specialised context. First, polysemous words can be items with specialised senses in one or more disciplines whose senses are different from those in everyday use. For example, “bugs” in computer science meaning “errors in computer programs, usually used in plural forms and informal situations”. This sense is different from that commonly used in everyday language, meaning “small insects or creatures”. Second, polysemous words can be items with specialised senses in different disciplines, such as “morphological”, which “means different things to linguists and botanists” (p.92). Each category of polysemous words may pose challenges to learners in specialised context language learning, which has been shown in studies. This will be discussed further in next sub-sections.

An early study related to words having multiple senses in specialised contexts is by Wang, K.T. and Nation (2004), where they specially focused on homonyms rather than polysemy. Although they did not discuss polysemy, their research provides methodological insights. Specially, they checked all 570 items in the AWL manually and found 60 word families (around 10%) are homonyms. Of these 60, 21 occur frequently in the corpus where the AWL was developed. Among these 21 word families with homonyms and frequent occurrence, 6 were worth being clearly distinguished on word senses to create new word families, because of the high frequency of homonyms, and another 3 should be excluded, because of the frequency of homonyms did not meet the inclusion criteria of the AWL. They argued that students may not have much difficulty on homonyms because of such low percentage of AWL homonyms when using the AWL in vocabulary teaching and learning. It seems that polysemy/homonym is not a big problem in academic contexts. However, 10% of word families in homonyms may still be a lot for students, especially for those with low vocabulary proficiency. Also, there are two other issues deserving further discussion when looking at results of Wang, K.T. and Nation’s (2004) research. First,

the AWL was developed based on exclusion of items in the GSL which contains general high-frequency words. General high-frequency words have been shown to be more likely to have multiple senses (e.g., Kuiper, Fromont and Gerhard, 2018; McCarthy, 2023; Skoufaki and Petrić, 2021; Stubbs, 2001). There are general high-frequency words also frequently used in academic contexts but not included in the AWL and checked in specific contexts. In this sense, the issue of polysemy is actually underestimated if only the AWL items were checked. Second, although they used a semantic relatedness scale (see Figure 3.1), dictionary checking and concordance line reading, the dictionary they used was a general use dictionary, which may not particularly list the specialised senses specific to a discipline (Dorset and Reijnerse, 2015). This could lead to omission on some specialised senses.

Polysemous words in discipline-specific contexts are also reported. Greene and Coxhead (2015) developed several discipline-specific vocabulary lists for middle school students in the USA, where they noticed some items overlapping in lists but having different senses (e.g., “function” in the mathematic list and chemistry list). However, they did not have report on how many items in the list are polysemous words and to what extent the senses are different from those in general use languages. Form-focused lists which may not guide users to think more about vocabulary knowledge in depth, for example, word senses in a specific context (Deignan et al., 2023; Deutchman, 2021). Researchers and language practitioners have reached consensus that polysemy poses challenges to learners in language learning of specific contexts. This will be discussed further in next sub-sections.

Semantic relatedness level	Description of the degree of relatedness
0	The meaning is the same as the base meaning.
1	The meaning is only slightly different from the base meaning.
2	The meaning is related to the base meaning with some changes.
3	The meaning is substantially different from but is still related to the base meaning.
4	The meaning is very distantly related and almost totally different from the base meaning.
5	There is no relationship at all between this meaning and the base meaning.

Figure 3.1 Semantic related scale used in Wang and Nation (2004) (adopted from Wang and Nation, 2004, p.297)

*Polysemous words with specialised senses that are different from those in general language*

As discussed in section 2.1, there is no clear boundary between specialised vocabulary and general use vocabulary. Some specialised vocabulary that are frequently used in specialised contexts may also be used, or frequently used, in everyday language. If these senses vary in different contexts, students (both L1 and L2 learners) may be confused when encountering these words. For example, Jacob (1989) found that first-year undergraduates in physics were unable to interpret physical terms accurately, when the terms are polysemous words and have senses different from everyday senses. Hulstijn and Laufer (2001) observed that even for advanced L2 learners rarely can know all senses of polysemous words in specific contexts. Parent (2009) noticed the problem caused by everyday sense of “volume” and its discipline-specific sense for L2 learners. Similarly, Deignan et al. (2019) also found this word as an issue for L1 learners. Fraser (2012) commented that learners believed they knew these words already, because the general senses they know may interfere specialised senses; however, this is an erroneous belief. Watson-Todd (2017) stated that polysemous words are a great problem for his engineering students at university. He noticed that the word “value”, whose engineering-specific sense is quite close to its general use sense, was difficult for his students to use it correctly in context. Alnamer (2017) found that Arabic students were able to guess primary sense of polysemous words, the one they are familiar with in everyday life, but faced difficulties in guessing senses in specific contexts. By testing L2 learners with at least B2 language proficiency levels, Mićović and Beko (2022) found that the correct guessing rate of polysemous word sense was quite low (ranging from 11.43% to 25.71%) when the word senses were closely related to a specific discipline.

*Polysemous words with different specialised senses in different disciplines*

A second category of polysemous words are those with different specialised senses in different disciplines. For example, Hyland and Tse (2007) noticed that “volume” tends to have different senses in social sciences (“books”) and sciences (“quantity”), as discussed in previous chapters. They also noticed that the same word form tends to have different collocates in different disciplines, and thus is associated with difference of word senses across disciplines. Collocations, like vocabulary, are associated with topics or disciplines (Baker, P., 2010; Deignan et al., 2023; Stubbs, 2001), which are a useful approach in distinguishing word senses (Firth, 1957; Hunston, 2022; Oakey, 2022; Sinclair; 1991). However, few studies have compared polysemous word senses in different disciplines with multiple examples. For example, as discussed above, although Greene and Coxhead’s (2015) vocabulary lists for middle school subjects have some word forms overlapping across subjects, they did not provide definitions for polysemous word senses in specific subjects. Quigley (2018) noticed “variable” has different specialised senses in computer science, science and mathematics, but this was not based on corpus data, with only some examples provided. These suggest that, although researchers

and practitioners have noticed the existence of polysemy and challenges it poses to students in specialised language learning contexts, there have been few corpus-driven studies on polysemy in specialised contexts. Through interviews with L1 middle school students, Deignan et al. (2019) found that students cannot paraphrase the sense of “concentration” correctly, when the word collocates with “camp” (“concentration camp”) in English, and “liquid” (“concentration of a liquid”). This word is a problem for students because it has different senses in specific disciplines, and the senses are different from that in everyday use. Further, Deignan et al. (2023) conducted a corpus analysis on polysemy in five UK middle school subjects, compared word senses in different subjects and provided examples to call for attention to polysemous words and senses in specific contexts: polysemy is an issue even for L1 students. Deignan and Jones (2025) provided a list containing 57 polysemous words whose senses in specific subjects are different. They provided an example sentence for each different senses, extracted from their concordance lines, for example “function” (p.132),

“The components of a cell each have different functions. Year 7 Science textbook.

Complete the table and draw the graph for each function. Year 8 Maths worksheet.

Draw a line to match each sentence to its correct function. Year 7 English assessment.

You’re better off going to this hotel cos they’ve got a better function room. BNC2014 Spoken.”

From studies discussed above, two issues can be seen. First, polysemous word senses, like vocabulary, vary according to specific contexts where they occur, especially in some discipline-specific academic contexts. This, therefore, suggests that polysemy analysis should be focused on specialised contexts, and have comparisons with word senses in other contexts to Figure out differences. The analysis needs to provide more detailed examples of polysemy in specialised contexts. However, and third, few studies mentioned above provide a comprehensive analysis, and most are with a small number of selected cases.

### 3.3 A corpus linguistic approach to polysemy analysis

Polysemy has been studied from different perspectives such as cognitive linguistics, computational linguistics, corpus linguistics, lexicography, psycholinguistics, philosophy, semantics and translation (Gries, 2019). Each perspective has offered different insights. Although senses, or “meaning is a subjective experience” (Moon, 1987, p.86) which needs human judgement, it is a product of context which can be realized or presented through a large amount of concordance lines (Moon, 1987). This is the true advantage of corpus linguistics and corpus analysis software. In this section, I discuss how corpus linguistic approaches help sense distinction in polysemy analysis. Gries (2019) summarized two main approaches to polysemy analysis through corpus linguistics methods. The first one is to use corpus

as resources to retrieve a large amount of examples for analysis. These examples are retrieved based on the concept of co-occurrence. Traditionally, word senses are considered to be independent and as a unique feature of word forms (Moon, 1987). Therefore, senses are recorded in isolation or decontextualization or with few contextual clues in dictionaries (ibid). However, linguists like Firth (1957) claim that word senses need to be viewed in a syntactical way. In other words, senses of words are contextual, influenced by other words or items co-occurring. For example, his example of “dark night” suggested the sense of “dark” and that of “night” affected each other (Firth, 1951). He made his classical statement that “you shall know a word by the company it keeps” (Firth, 1957, p.179), and called for identifying meaning by collocation.

Collocation is defined as two or more words often co-occurring together within a specified span of text (Sinclair, 1991). Thanks to the advancement of computer science technology and pioneering research on collocation analysis (e.g., Sinclair, 1966; Sinclair et al., 1970), electronic corpus and collocation are used to distinguish senses in lexicography. With their experience of developing the COBUILD dictionary through corpus study, Moon (1987) commented that, in contexts, the interpretation is restricted which helps with disambiguation of word senses. Through observation and concordance line reading, she provided specific examples on word senses of “skate”. When “skate” means a sporting activity, this word often collocates with “ice” and “roller”. Additionally, these two collocations also distinguish from each other, suggesting two different sporting activities. When “skate” refers to a kind of fish, it collocates with words such as “fish” and “water”. Another example is the noun “gap”, which has four distinguishing senses. To refer to sense of “physical space or distance”, “gap” collocates with “mountain”, “teeth” or “between”. For the sense of “time intervals”, this word collocates with “year” and “of”. It collocates with “fill” referring to “deficiency”, and with words such as “generation”, “narrow” or “reduce” referring to “discrepancy”. In this way, different word senses can be identified through different sets of collocations which are closely associated with that word form, and are decided within contexts. Collocation, therefore, is recognized as a formal criterion of lexicalised situational context to disambiguate senses, which sometimes may be the only criterion (e.g., sense distinction of delexical verbs) (Barnbrook, Mason and Krishnamurthy, 2013; Moon, 2008). Sinclair (1991) found different senses of “yield” through different collocations and contexts of this word. He also noticed that different forms of “yield” such as “yields”, “yielded” and “yielding” have different senses, which suggests that word forms, collocations and contexts are all factors in distinguishing word senses.

Due to the development of quantitative analysis methods, sense distinction through collocations has more complex statistical techniques. Corpus analysis software or websites like AntConc, LancsBox or Sketch Engine provide various techniques for collocation strength calculation like t-score, MI, LogDice and others (Szudarski, 2022). These are objective methods to identify repeated collocations for further analysis of meaning. For example, Deignan and Love (2021) identified different senses of “ice” through

comparing its top collocations with log-likelihood in a self-built science corpus and the Spoken BNC2014. Apart from function words in collocations, they noticed that the top collocations of “ice” in the Spoken BNC2014 are all things produced by humans such as “chocolate”, “skating” and “cube”. In their science corpus, the top collocations such as “sea”, “melting” and “arctic” are more related to natural phenomenon which are outside human control. It supports the workable idea that different senses can be identified through different collocations because collocations are closely associated to the contexts (e.g., Baker, P., 2006; Sinclair, 2004). Nevertheless, Ellis (2002) argued that high frequency of collocations may not directly imply these collocations having identifiable and distinct meanings or functions. These frequent collocations can be high-frequency words or function words. Therefore, the frequent collocations do not necessarily suggest they make sense in psycholinguistic salience. Brezina and Fox (2021) also call for qualitative analysis on collocation, together with statistical methods.

Those above indicate that sense distinction based on collocations in corpus analysis still needs assistance of subjective judgement of researchers. This is because the decision of splitting senses is a subjective process (Deignan et al., 2021; Moon, 1987). Hoey (2005) provides psycholinguistic explanation on collocation analysis that collocations are stored in brains, and that words are primed when other words often collocating with them are set as prompts. However, whether senses should be distinguished when each different set of collocation occurs depends on research aims. Gries (2019) also summarized that the sense distinction of polysemy ranging from extreme lumpers to extreme splitters. The former refers to strict monosemic bias (e.g., Ruhl, 1989) which emphasizes that each word has a core meaning shared across all senses it has, and the latter is extreme polysemy – splitting word senses with fine and tiny features. For example, Wang, K.T. and Nation (2004) followed Ruhl’s (1989) Monosemic Bias to structure their framework of four-level semantic relatedness scale for rating word sense relation for distinguishing, and then checked the word using a corpus. Wei (2006) conducted a series of analysis on “experiment”, based on data from Jiaoda English of Science and Technology (JDEST) corpus (Yang, 1985), one of the earliest electronic specialised corpora in China and Asia. He split collocations of “experiment” whose z-scores are beyond threshold ( $>2$ ) as two clusters. The first one has collocations like “performed”, “conducted”, “controlled” and “reported”; the second one has words such as “demonstrate(s)”, “show(s)”, “indicate(s)” and “suggest(s)”. His reason was that these collocations perform different grammatical functions with “experiment”: (1) “conducted an experiment” and (2) “an/the experiment shows”. In (1), “experiment” is considered as a research activity performing as the objective of verbs, while in (2) it is recognized as a part related to writer’s comments, which realizes a different discursal function compared to that in (1). This splitting of sense is more like extreme splitters mentioned in Gries (2019). Differently, Deignan et al. (2023) used splitting narrowly when analysing polysemy in UK middle school subjects. One of their research aims was to identify what polysemous words and what senses they have may pose challenges to students, when compared to senses in other disciplines or everyday topics. Therefore, they did not split senses as what Moon

(1987) did in developing the COBUILD dictionary in the way that Wei (2001) did mention above. In addition, polysemy analysis through corpus also needs checking words used in contexts, that is, concordance lines. Fox (1987) emphasized the importance of using real language examples to provide guidance on collocation in sense distinction, which can be realized through concordance lines obtained in corpus. Hunston (2022) claims that most corpus analysis starts from quantitative methods but finally often ends with qualitative investigation of language items (e.g., individual words or multi-word units) through concordance lines. Through different rounds of sorting concordance lines, the repeated collocations or structures contributing to word senses can be observed (Hanks, 2013; Oakey, 2022; Sinclair, 1991; Stubbs, 2001). For example, Deignan et al. (2023) studied concordance lines available from their corpora, and then grouped concordance lines according to word sense and use. This supports the claim of Hunston (2022) that meaning and form are closely associated with each other, both of which can be observed through concordance lines.

A second approach summarized in Gries (2019) on polysemy analysis is more related to computational linguistics. This approach has a big difference which does not involve (semi-) human annotation, for example, natural language processing (NLP) or large language model (LLM) analysis. It aims to use pre-trained deep-learning language models to capture word senses in contextualization, for example, investigation on general word meaning frequency (e.g., Liu, Gong, Shi and Guo, 2025), and automatic detection of metaphors (e.g., Reimann and Scheffler, 2025). In other words, this approach involves much computer assistance and computer language models to handle extensive data with a pre-determined framework. In this section, my focus is relevant studies in specialised context (e.g., Geng and Liang, 2024; Gong, Liu, Shi, and Guo, 2025), although much research existed on general use English. Geng and Liang (2024) compiled two corpora, one for soft and hard sciences each, with 28,000 research article abstracts and 4.2 million tokens in total, and created part-of-speech (POS) tags through Python. They then used a pre-trained language model BERT to capture word sense of nouns in two disciplinary contexts. The difference of sense of polysemous noun words was presented through a statistical test ANOSIM, and a polysemous word list was generated. To validate their results of language models, they conducted a collocation analysis on selected items. For example, “stress” in soft sciences is more related to mental health conditions while in hard sciences refers to physical pressure, which is in line with findings of Deignan et al. (2023). However, their research focuses on polysemy in a broad view of disciplines, rather than word senses in use in a specific discipline, for example, to what extent the word senses in a specific discipline are different from those in other contexts. Relying on computers only may not manage that, since computers cannot yet understand languages as humans do (Deignan and Jones, 2025). Gong et al. (2025) used BERT (Bidirectional Representations from Technology model) to identify word senses in general academic contexts, and developed a meaning-based academic vocabulary list. They firstly developed an 1550-lemma academic vocabulary list based on an 85-million-token general academic corpus covering 9 disciplines from the COCA, by following Gardner

and Davies's (2014) approach. They examined senses of these academic words by BERT with semantic annotations based on entries from Oxford English Dictionary (OED), finding that most of lemma are polysemous. Senses not meeting their selection criteria (e.g., range, dispersion and frequency ratio) were excluded. For example, they excluded the sense of sheet in lemma "paper", but kept the sense of essays; the sense of fictions of lemma "novel" was excluded but that of "new or unusual" was retained. This list has pedagogical implications that general academic word senses were provided, with exclusion of general use senses. However, similar to that of Geng and Liang (2024), their research did not provide sufficient information of word senses in a specific discipline, or have comparison on word senses across disciplines.

### 3.4 Corpus analysis of polysemy in specialised contexts

This section reviews studies using the first approach of corpus analysis on polysemy. Not much research exists on polysemy in specialised contexts. Wang, K.T. and Nation (2004) is one of the early studies; however, as discussed above, they focused on homonyms using a sense relatedness scale, with some sampled concordance lines checked. Corpus analysis on polysemy in specialised contexts has received more attention in recent years. Deignan et al. (2019) observed through both corpus and interviews that UK middle school students may encounter polysemous words and their different senses in scientific session materials, some of which students may not accurately comprehend (e.g., "concentration"). Further, by using BNC 2014 Spoken corpus as a reference, Deignan and Love (2021) identified a list of polysemous words that are frequently used in the UK middle school science subject (climate change topic), through collocation analysis and concordance line checking. Their word sense comparison between specialised science context and everyday spoken context shows the difference of word senses in contexts (e.g., "ice" as a natural phenomenon in scientific context and collocates with food and drinks in everyday spoken context). Their methods show that collocation can be one of the techniques to identify polysemous words and word sense difference. They also hypothesized that polysemy may widely exist in other disciplines. Based on this, Deignan et al. (2023) collected classroom materials of English, Science and Mathematics used in KS2 and KS3 stages for corpus compilation and had analysis on a series of linguistic features. With a keyword list of each KS3 subject generated, they checked the collocations and concordances of each word types in the lists through their corpora of KS2, KS3 and a general corpus respectively. Their findings show that polysemy exists in different disciplines, and the word senses may be different from those in everyday use or those in other disciplines. Five categories of polysemous words were summarized: (1) word sense difference due to contextual/topic difference, (2) fine-grained difference of senses, (3) distinct word sense difference, (4) difference of lexicogrammatical feature, and (5) word frequency difference. They believe that these polysemous words may pose challenges to students at this stage, since students may not be familiar with such senses partially

due to insufficient input in daily life and unfamiliarity to specific contexts. The keyword list with polysemous word sense comparison and example sentences is presented in Deignan and Jones (2025). This series of research suggests that polysemy deserves attention in a specialised context and further word sense comparison, where collocation analysis and concordance checking for contextual clues are key to polysemy analysis.

Apart from studies on polysemy for middle school students, there are studies on higher education academic contexts. Skoufaki and Petrić (2021) examined word senses of all items in a corpus-derived specialised vocabulary list to identify polysemy in general academic contexts. They checked items in the AVL to see how many items are polysemous words with the help of COBUILD dictionary and WordNet. Through assistance of these lexicographic tool, three major findings were generated. First, around 1/3 lemmas in the AVL are polysemous words. Second, most of polysemous lemmas are at the first 1,000 frequency band of the AVL, suggesting a close relation between word frequency and polysemy. The higher frequency a word has, the more likely it is a polysemous words. It is in accordance with polysemy research results in general use contexts. Also, polysemous words at a higher frequency band of the AVL tend to have more senses than those at lower frequency band. Third, polysemous words also correlate with part of speech, nouns, verbs and adjectives being more likely to be polysemous words. Their research has a comprehensive analysis of word senses of the AVL items, and answers to what extent these items are polysemous words. Part of their findings are in line with those in general use contexts (e.g., Nerlich and Clarke, 2003; Parent, 2009), suggesting that polysemy is a pervasive linguistic phenomenon and associates with word frequency and part of speech. Moreover, this research also shows that lexicographic tools, such as well-known dictionaries, can also be useful in identifying polysemy (Kilgarriff, 1992, 2006; MacAurthur, 2015; Pantel and Lin, 2002). However, they did not check collocation profiles of these items in the AVL nor the concordance line reading. Collocation is one of useful ways to disambiguate word senses in contexts (Sinclair, 2004; Stubbs, 2001). In addition, they are in general academic context, but Hyland and Tse (2007) argued that word senses may be different in specific disciplines. Without checking word senses in a discipline-specific context, it is difficult to answer this.

Le and Miller (2023) explored polysemous words in medical context. Similar to Skoufaki and Petrić (2021), they examined items of a corpus-derived specialised vocabulary list with assistance of lexicographic tools. Specifically, they checked headwords of Hsu's (2013) Medical Word List (MWL) which contains 595 word families frequently used in medical contexts, through the Oxford English Dictionary (OED). 302 polysemous headwords were identified, accounting for more than 60% of all items in the list. This percentage is quite different from that in Skoufaki and Petrić (2021), suggesting more exploration is needed in other discipline-specific contexts. The difference may result from disciplinary features, or different ways of understanding polysemy. Different from Skoufaki and Petrić (2021), they also invited three experts (two in EAP and one in medical field) to identify core meaning

of each polysemous word. Additionally, they noted that the MWL was developed out of the GSL and the AWL. They also noted that, in this way, it could lead to omission of some items with medical senses in the GSL and AWL. They therefore argued that the percentage of polysemous words could be even higher if these items can be checked for polysemy. However, they, like Skoufaki and Petrić (2021), did not conduct collocation analysis and concordance checking. In addition, although they had the OED as a tool to explore word senses, a discipline-specific medical dictionary was not included. Referring to discipline-specific dictionaries can be a supplementary to general use dictionary, since general use dictionaries like the OED may not include medical technical word senses (MacArthur, 2015). Although Le and Miller's (2023) research invited three experts, human judgement is based on one's working experience and professional knowledge which may differ from that of others (Coxhead et al., 2020; Schmitt, 2010). In this sense, the discipline-specific dictionary can be an assistant.

It can be observed through findings of the above studies (e.g., Deignan et al., 2023; Deignan and Jones, 2025; Le and Miller, 2023; Skoufaki and Petric, 2021; Wang, K. T. and Nation, 2004) of corpus analysis on polysemy in specialised contexts: polysemous words and word senses, like specialised vocabulary, vary according to the context where they occur. However, to the best of my knowledge, there are very few studies on polysemy analysis through corpus focusing on a discipline-specific context. Also, not much research employed multiple tools in polysemy analysis for disciplinary word sense identification (e.g., involving collocation analysis, concordance line checking, and discipline-specific dictionary for disciplinary sense identification). This raises my interest in exploring polysemy in Logistics discipline-specific context with multiple tools involved, and have further word comparisons with other contexts. I hope this research could have contributions to corpus analysis on polysemy in a discipline-specific context, and methodology in polysemy analysis with multiple tools. The next chapter presents procedures of the compilation of Logistics Academic Written English Corpus (LAWEC), and different tools for word sense analysis.

## Chapter 4 Methodology

### 4.0 Introduction

This chapter discusses the methods used to find the answer to the four research questions:

RQ1: What are the most frequently-used and evenly-distributed specialised Logistics words, and what are their formal features?

RQ2: To what extent are the words identified used in ways specific to Logistics?

a) is their use different from the general academic context?

b) is their use different from general English?

RQ3: What is the extent and nature of polysemy in the words identified?

### 4.1 Corpus Compilation

This section discusses the compilation of the Logistics Academic Written English Corpus (LAWEC). This corpus is compiled to attempt to answer RQ1. This section consists of three parts. Section 4.1.1 focuses on principles underlying the development of the LAWEC, and my original plan for corpus development based on these principles. Section 4.1.2 describes the actual steps I took when applying those principles in corpus compilation. Section 4.1.3 is the evaluation of the quality of the self-built corpus against the principles discussed in Section 4.1.1.

#### 4.1.1 Principles underlying corpus compilation

##### *Representativeness*

Representativeness refers to the extent to which a corpus provides a good generalisability on the domain it targets to (Egbert, Biber and Gray, 2022). For this principle, it is recommended by Biber (1993) and Coxhead (1998) that a corpus should select a wide range of materials which are representative to the context it focuses on. With principled sampling method and representative data, the corpus itself can be the basis of language analysis generalization (Biber, 1993; Reppen and Simpson-Vlach, 2020) and help with more valid results. Specifically, Biber (1993) suggests that this principle can be realised through three stages:

(a) defining the target population represented in the corpus,

- (b) selecting samples representing the target population, and
- (c) evaluating the extent to which the samples could represent the target population.

This section is mainly on first two stages, and the last one will be discussed later in section 4.1.3. In terms of the first step (defining the target population), the composition of the corpus needs firstly to be figured out, that is, what text types are (not) included (Egbert, Larsson and Biber, 2020). This involves two aspects: boundaries of the target population, and internal categories of the target population (Egbert et al., 2022). Both of these are shaped by research aims and have impact on linguistic features that the corpus contains (Egbert et al., 2020; Reppen and Simpson-Vlach, 2020). As discussed in section 1.2, this research mainly aims to identify specialised individual words and their senses in a corpus representing Logistics academic written English. Therefore, the created corpus LAWEC focuses on academic written language in Logistics as the boundary of target population. For the inner categories of the target population, registers are in consideration (Biber, 1993; Egbert et al., 2022). In the LAWEC, academic journal articles are chosen as the register as inner categories. Compared to other academic written registers (e.g., students' essay, manual report or textbook), academic journal articles may provide a better picture to examine specialised vocabulary and word senses for polysemy analysis for three reasons. First, academic journal articles written by experts are more informative (Conrad, 2001), and have a wider range of lexical diversity, compared to other aforementioned academic written registers. With more diversified lexical items, academic journal articles may contain more specialised vocabulary as candidates to be chosen for specialised word list development and word sense analysis. Second, academic journal articles are “the primary means” to convey disciplinary knowledge (Jin, 2021, p.1), and are the typical, widely-used representative texts in academic contexts (de Chazal, 2014; Lu, 2018). This feature implies the diversity of audiences of academic journal articles, and therefore analysing specialised vocabulary and word senses in this register may bring benefits to a wide range of stakeholders in both academic and professional field. By focusing on this register, it could help undergraduate or postgraduate L2 students in Logistics on Logistics specialised vocabulary, the target users of the vocabulary list developed in this research. It could also provide materials for teachers or professionals in Logistics field. Third, academic journal articles are more available to be accessed to, through the Internet and libraries in the digital publishing era. The open accessibility and availability of academic journal articles may lead to fewer ethical issues on copyright or confidentiality (McEnery and Brookes, 2022). Moreover, free online availability suggests that the articles are more likely to be read by target beneficiaries.

In terms of the second step (selecting representative samples) it requires to consider three aspects: sampling unit, source of raw materials and time span of texts in collection. Sampling unit refers to the individual texts collected in the corpus, and it is often aligned with the title of inner categories (Egbert et al., 2022). For this corpus (LAWEC), the sampling unit of the academic journal articles register (inner category) is the whole text of individual academic journal articles. Texts to be collected need to meet

two specific criteria. First, they should be academic journal research articles. It means that all non-research articles published in a journal (e.g., editorial preface, acknowledgements, content of issues, notice of call for papers and case reports) will not be collected because of their less relevance to disciplinary knowledge. In my study, academic journal research articles were decided based on information provided by journals in sections. Specifically, texts in sections of “Articles”, or “Research Articles”, or “Original Articles” were included, and those in sections of “Editorial”, “Case Report” were excluded. Second, the whole main body texts (presented in words rather than in tables or graphs) were included rather than extract of texts. Evidence has shown that differences of lexis are found in different sections of academic journal articles (Thompson and Hunston, 2019). Thus, focusing on completeness of written texts attempts to reduce risks of section-biased specialised vocabulary on word sense analysis and polysemy, rather than collection of fragments. Moreover, collecting whole main body texts presented in words is currently a common way on sampling unit for corpus compilation (Egbert et al., 2022; McEnery and Brookes, 2022; Sinclair, 1991), with tables, formulas, footnotes, endnotes, disclosure statements, acknowledgements, contributions of authors, author biography, funding issues and reference/bibliography excluded. This research would follow the convention in corpus compilation.

Apart from sampling unit, selecting representative samples also requires consideration on the sources of raw materials. Logistics is an interdiscipline closely related to disciplines like Business, Economics, Engineering, Management, Mathematics and Sociology (Swanson, Goel, Francisco and Stock, 2017). For setting clear disciplinary boundaries in searching Logistics journals collected in the LAWEC, it consists of academic journal articles from 10 top key journals on Logistics (see Table 4.1), by searching ‘logistics’ in discipline/field of journals in the Journal Citation Report (JCR). These journals are listed in the JCR and recommended to experts or individuals who are interested in the field of Logistics. The JCR evaluates journals in a multidimensional method where it provides basic information of listed journals, including issues published each year, Impact Factors, Journal Citation Indicators and journal ranking (Clarivate, 2022). It is an objective, selective and transparent list updated annually, helping with identifying journals that are crucial to a certain field and discovering the best-fit journals for the researchers (ibid). Moreover, one of professionals confirms that she is familiar with some of these journals when she did her Master’s learning on Logistics in a university at Hong Kong (2022-2023), and they are also in her reading lists (personal communication, 2023). Therefore, the 10 top key Logistics journals recommended by the JCR 2022 can be regarded as representative texts, and may be candidates of the top choice in this field.

Furthermore, time span of published texts should also be in consideration for selecting representative samples. For the LAWEC, by the time of data collection in May 2023, only texts published within recent 5 years (2018-2022) would be collected. The reason for collecting recent 5 years articles for corpus compilation is to keep updating of the knowledge development of this discipline, following previous research such as Lu (2018) and Roesler (2020). After the decision of sampling unit,

sources and time span of texts in corpus, it is necessary to consider specific process of sampling method. Following Egbert et al.'s (2022) suggestion, stratified sampling was adopted in this process. This is because collecting all academic journal articles from 10 journals published within 5 years may make the corpus too large to have practical analysis within the PhD frame, and therefore more principles are needed. These will be discussed further in next subsections.

### *Balance*

In the process of stratified sampling, the proportion of texts from each stratum should be considered (Egbert et al., 2022). For the LAWEC, it is about the number of texts selected from each journal. To avoid the potential bias towards a particular journal and issue, initially, it was decided to select the same number of issues and articles from each journal randomly. However, it is common that each journal may have differences on the numbers of issues and articles published every year. As shown in Table 4.1, there were variations in the number of issues published each year (ranging from 4 to 12) and the number of articles in each issue (ranging from 4 to 19).

To take both different numbers of issues and articles published in journals and the principle of balance into consideration, I decided to choose the minimum number of issues published each year and the minimum number of articles published in each issue among 10 journals. Specifically, four issues would be selected from each journal per year, and in every issue, four articles would be selected. If a journal publishes more than four issues each year and/or contain more than four articles in each issue, random selection would be conducted, since randomness in selection can avoid selection bias (Egbert et al., 2022). This means that 80 articles were decided to be randomly chosen from each journal (4 issues \* 4 articles \* 5 years = 80 articles in a journal), and in total 800 articles for the LAWEC (80 articles \* 10 journals = 800 articles in total).

The balance between tokens and texts in corpus also deserves consideration. Due to the difference in the length of texts, it is common that the number of tokens is different if the corpus collects the same number of texts from each stratum (Oakey, 2009). The LAWEC has the same issue: if the same number of tokens from each journal is selected, the number of texts involved is different; if the same number of texts from each journal is chosen, there exists the variation of word count in tokens. The former is isolexical corpus while the latter is isotextual corpus (Oakey, 2009). Such difference between the number of tokens and texts may lead to the differences of linguistic distribution, that is, target linguistic variables across the corpus are different in values (Egbert et al., 2020; Egbert et al., 2022). For example, in the LAWEC, this could have influence on the ranking order of frequent items when selecting specialised vocabulary for polysemy analysis. However, based on his research on 3-word lexical bundles, Oakey (2009) suggests that such risks of imbalance between tokens and texts can be largely addressed through meeting either one of the following criteria. For isolexical corpus, the number of

texts selected could be more than 500; for isotextual corpus, a 5-million-token corpus or that in a larger size is acceptable. For this thesis, since the number of articles selected from each journal is the same and in total 800 articles were decided to collect, the LAWEC meets the criterion of isolexical corpus to reduce the risk of imbalance. Apart from balance, it is also necessary to consider the corpus size when creating a representative corpus, which is related to the principle of size.

Table 4.1 Information about the ten chosen journals

Journal title	Issues Published Each Year	Papers Published in Each Issue (Average)	Papers Collected/Published in 5 Years	Impact Factors	Levels
Transportation Research Part E: Logistics and Transportation Review	12	19	1149	10.047	Q1
International Journal of Systems Science: Operations and Logistics	4	8	131	9.04	Q1
Journal of Business Logistics	4	4/6	90	7.875	Q1
International Journal of Physical Distribution and Logistics Management	10	4	191	7.29	Q1
International Journal of Logistics Research and Applications	6 (2018-21) 12 (2022)	5	162	5.992	Q2
International Journal of Logistics Management	4 (2018-21) 5 (2022)	13	247	5.446	Q2
Asia Pacific Journal of Marketing and Logistics	5 (2018-19) 8 (2020) 10 (2021-22)	11	373	4.643	Q3
Maritime Economics and Logistics	4	7	141	3.841	Q3
Naval Research Logistics	8	6	212	1.806	Q4
International Journal of Shipping and Transport Logistics	4	9	180	1.264	Q4

### *Size*

For the decision on the size of corpus, there exists no absolute criterion, and it depends on the purposes of research and data availability (Koester, 2022; McEnery and Brookes, 2022). Generally, it is

recommended that a corpus should consist of millions of tokens (Sinclair, 1991; Nation, 2016; Hunston, 2022), but the specific size of the corpus can be different, especially when the corpus is of specialised purposes (Koester, 2022). Egbert et al. (2022) suggest that a corpus should allow linguistic variables of interest to occur a certain of times across the corpus for analysis. This is because one of tasks of corpus linguistics is to describe what is typical in language use, instead of everything (Hunston, 2022). Considering my research aims are to have analysis on identified specialised Logistics words to examine word senses and polysemy, the corpus should be large enough to allow lower frequency specialised vocabulary candidates to have a reasonable number of occurrences. The larger the corpus is, the more reliable the results could be (Nation, 2016). The LAWEC should be at least one million tokens in size, since Stefanowitsch (2020) suggests that corpus with millions of tokens is adequate to investigate most linguistic phenomenon.

Also, since the LAWEC focuses on a specific discipline, it is crucial to consider guidelines on size of specialised corpus. O’Keeffe, McCarthy and Carter (2007) recommends that specialised corpora could be around 5 million tokens in size, but Koester (2022) argues that it can be smaller in practice, taking practical issues (e.g., data availability, ethical issues, the length of time for collection, fundings) into consideration. On reviewing about 50 journal articles and doctoral theses on specialised vocabulary identification and list development with corpora (see Chapter 2), only about 10% (about 5-8 texts) generate results from corpora larger than 5 million tokens, 25% beyond 3 million, 30% ranging from 1 to 3 million, and the rest are less than 1 million. Therefore, it was estimated that a corpus larger than 1 million tokens and a little bit more than/around 3 million tokens is acceptable in size for this research.

However, a gap may exist between principles and practice of the corpus compilation. The actual steps that I took for corpus compilation should be described when applying principles discussed above in creating the LAWEC. Modifications were made when in necessary, including the issue of data availability, balance of tokens, specific sampling selection process, and size of corpus, which will be presented in the next subsection.

#### 4.1.2 Actual steps and modifications of corpus compilation in practice

##### *The adjustment on data availability*

In terms of data availability, ideally, 10 top key journals were decided as the potential resources. All except one were available through the library website of the University of Leeds by the time of data collection in 2023 May. The articles of *International Journal of Shipping and Transport Logistics* cannot be accessed through the university library website officially by the time of writing. Therefore, articles in the rest nine journals published from the period of 2018-2022 are to be collected. Each journal was coded based on initial letters of the journal, for example, *Journal of Business Logistics* coded as

BL (see Table 4.2). The exclusion of texts from a journal may not have huge impacts on the representativeness of my corpus. This is because the absent journal has the least Impact Factors and ranking among all target journals, as Table 4.1 shows. It indicates that this journal may not as influential as other nine in terms of spreading Logistic disciplinary knowledge in this field.

### *The adjustment on balance of tokens from each journal*

The balance of tokens from each journal also needs modification. Through a trial practice of data collection, it was realized that the length of articles from nine journals varies a lot. Official journal websites and author submission guidelines were searched for the estimation of suggested maximum word count of these articles (see Table 4.2). For those journals which do not provide clear guidance on word count for submission, I contacted with their current (Co-)editor-in-Chief for further information (marked with \*).

Table 4.2 Suggested maximum word count of journals

Journal title and Code	Maximum word count
Transportation Research Part E: Logistics and Transportation Review (TR)	12,000*
International Journal of Systems Science: Operations and Logistics (OL)	10,000*
Journal of Business Logistics (BL)	12,000*
International Journal of Physical Distribution and Logistics Management (PDLM)	10,000
International Journal of Logistics Research and Applications (LR)	10,000
International Journal of Logistics Management (LM)	10,000
Asia Pacific Journal of Marketing and Logistics (ML)	8,000
Maritime Economics and Logistics (MEL)	8,000
Naval Research Logistics (NRL)	10,000*

Through Table 4.2, it can be observed that the maximum length of articles varies across nine journals. This may have impacts on my results on identifying specialised Logistics words, since there could be linguistic distribution bias towards to those longer articles (Egbert et al., 2022). Therefore, modifications were made to attempt to keep tokens more balanced to minimize the bias caused by the difference in the article length.

The adjustment is based on the maximum word count of articles in each journal. From Table 4.2, there are three levels of maximum word count: 8,000, 10,000, and 12,000. From the shortest to the longest one, they each are around 20% shorter than the longer one. Therefore, the number of articles selected from each journal can follow such difference for keeping the balance on tokens. Specifically,

the number of articles chosen from the journals with maximum 10,000 word count is not changed, namely four articles from each issue, as presented in subsection 4.1.1. The number of articles chosen from journals with maximum 12,000 word count needs to be 20% smaller than that in the original design, due to its longer texts ( $4 - 4*20\% = 3.2 \approx 3$  articles from each issue), and, for the journals with maximum 8,000 word count, 20% more articles should be added, compared to that those in the original plan ( $4 + 4*20\% = 4.8 \approx 5$  articles each issue).

Such decision has three advantages. First, although the number of articles selected from each journal is different now, this modification is based on the length of articles. It attempts to reduce the potential “Matthew Effect” of linguistic distribution in the LAWEC, that is, linguistic patterns may be influenced by and recur more frequently in longer texts. Second, with modifications, the total number of articles selected has not been changed between the original and current plan (see Table 4.3). In total there are 720 articles in the LAWEC, the main difference being the number of articles from journals with different requirement on word count. This indicates that such changes on the number of articles chosen may not have much influence on the whole corpus in terms of total sampling units, but it realises a more balanced corpus in general. Third, it also keeps the balance on tokens of each journal in a better way (see Table 4.4). In the original plan, the estimated number of tokens of each selected journal varies a lot, due to different lengths of journal article publishing requirements. It ranges from 640,000 words to 960,000 words. The number of tokens from journals with shortest text length is 1/3 smaller than that from journals with longest text length. However, this bias is largely reduced in the current plan. Although the difference of estimated number of tokens in each selected journal still exists, such difference is much smaller than that in the original one. In this current plan, the estimated number of tokens from each journal has a range from 720,000 words to 800,000 words, suggesting only a nearly 10% difference of word counts in different journals. In this way, a more balanced corpus can be developed, through attentions to text lengths and number of articles in each journal. This can help with selecting specialised Logistics vocabulary with less bias towards some specific journals due to length of texts.

Table 4.3 Original and current corpus design on number of texts

Corpus Design	Word Count in Texts	Number of Journals	Time Period of Collection (years)	Number of Issues in Collection	Number of Articles in Collection	Total Number of Articles
Original Design	N/A	9	5	4	4	720
Modified Design	8,000 words	2	5	4	5	720
	10,000 words	5	5	4	4	
	12,000 words	2	5	4	3	

Table 4.4 Original and current plan on sample selection

Journal Code	Issues Selected	Maximum Word Count	Papers Selected in Each Journal (Original/Current)	Total Selected (Original/Current)	Papers Maximum Estimated Tokens (Original/Current)
TR	20	12,000	4/3	80/60	960,000/720,000
BL	20	12,000	4/3	80/60	960,000/720,000
OL	20	10,000	4/4	80/80	800,000/800,000
NRL	20	10,000	4/4	80/80	800,000/800,000
PDLM	20	10,000	4/4	80/80	800,000/800,000
LR	20	10,000	4/4	80/80	800,000/800,000
LM	20	10,000	4/4	80/80	800,000/800,000
ML	20	8,000	4/5	80/100	640,000/800,000
MEL	20	8,000	4/5	80/100	640,000/800,000
<b>Total</b>	<b>180</b>	<b>n/a</b>	<b>n/a</b>	<b>720/720</b>	<b>7,200,000/7,040,000</b>

### *The specific sampling selection process*

The third issue is the specific sampling selection process. It involves four steps: material collection, coding, random selection on issues and that on specific articles. For step 1 material collection, based on the principles discussed in section 4.1.1, I collected all research articles published within recent five years from all nine journals available, downloading these texts manually in PDF version. There were 2,728 texts in total. Metadata of all these 2,728 texts was also collected and stored independently in an Excel working sheet. The collection of metadata is an important substep of material collection and corpus compilation, since this helps with finding sources of texts (Weissen, 2015). Metadata of texts in the LAWEC includes information of authorship and institution of authors, which can help with

evaluation of representativeness of corpus (see section 4.1.3). Metadata also contains codes for texts in the LAWEC.

For step 2 (coding the texts in the LAWEC), each article is attached with an individual code, consisting of four components: code of journal, publishing year, issue number and paper number. For example, in the code PDLM180103, PDLM refers to the journal *International Journal of Physical Distribution and Logistics Management* based on initial letters of key information in the journal title; 18 is for the publishing year 2018; 01 is for the first issue published in that year, and 03 for the third paper in that issue (see Table 4.5 for coding example). All codes are stored as a part of metadata for the LAWEC compilation.

Table 4.5 Coding Example

Coding Example: PDLM180103				
Components	Journal Code	Publishing Year	Issue Number	Paper Number
Examples	PDLM	18	01	03
Interpretation	<i>International Journal of Physical Distribution and Logistics Management</i>	2018	The first issue published in 2018	The third paper published in the first issue of 2018

After coding, it is to consider step 3 (random selection of issues and specific articles). For randomly selecting four issues a year and 20 issues a journal, the codes were input on the free available random selection website [random.org/list](http://random.org/list) (Haahr, 1998). In selecting issues, part of the code (publishing year + issue number) was used in this process, as Figure 4.1 shows. For example, in the random selection of issues published by PDLM in 2018, there were 10 issues, so 1801, 1802, 1803...1810 represented each issue of that year respectively, and all 10 codes were input on the website.

[doubt, contact us.](#)

## List Randomizer

This form allows you to arrange the items of a list in random order. The randomness comes from atmospheric noise, which for many purposes is better than the pseudo-random number algorithms typically used in computer programs.

### Part 1: Enter List Items

Enter your items in the field below, each on a separate line. Items can be numbers, names, email addresses, etc. A maximum of 10,000 items are allowed. Please don't enter anything you would consider confidential ([here's why](#)).

```
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
```

(you're viewing this form [securely](#))

### Part 2: Go!

Be patient! It may take a little while to randomize your list...

Figure 4.1 Input codes for random issue selection

The random selection website generates a random list of order, which to some extent reduces the risk of my subjective selection and ensures the reliability of the research. The top four will be the issues selected. For example, as Figure 4.2 shows, 1803, 1804, 1807 and 1808 were the top four. It means that the third, fourth, seventh and eighth issue published by PDLM in 2018 were selected as the resources of samples. This method is only suitable for journals publishing more than four issues every year. If a journal only publishes four issues a year, then there is no need to use random selection website since all four issues were selected. The above process was repeated in each journal and each recent five years respectively. In all, 180 issues were selected (4 issues \* 5 years \* 9 journals = 180 issues). A note was taken for all 180 selected issues, in preparation for article selection among issues.

[Home](#)
[Games](#)
[Numbers](#)
[Lists & More](#)
[Drawings](#)
[Web Tools](#)
[Statistics](#)
[Testimonials](#)
[Learn More](#)

[Login](#)

# RANDOM.ORG

True Random Number Service

**Advisory:** We only operate services from the RANDOM.ORG domain. Other sites that claim to be operated by us are impostors. If in doubt, [contact us](#).

## List Randomizer

There were 10 items in your list. Here they are in random order:

1. 1803
2. 1807
3. 1808
4. 1804
5. 1802
6. 1801
7. 1806
8. 1810
9. 1809
10. 1805

IP: 109.175.189.57  
 Timestamp: 2023-05-15 17:42:27 UTC

**Don't use this service for giveaways! Use [Multi-Round Giveaways](#) instead**

Figure 4.2 Examples of random issue selection

Once the issues in each journal had been selected, the final step was to use the random selection website again to randomly select papers in each issue. The number 1, 2, 3, 4 and so were input into the website box, each number representing a paper in an issue. For example, as Figure 4.3 shows, in the process of selecting papers in an issue of *the Maritime Economics and Logistics* (Issue 4 in 2018), there are 7 papers in this issue, so seven numbers were input (1-7).

**Advisory:** We only operate services from the RANDOM.ORG domain. Other sites that claim to be operated by us are impostors. If in doubt, [contact us](#).

### List Randomizer

This form allows you to arrange the items of a list in random order. The randomness comes from atmospheric noise, which for many purposes is better than the pseudo-random number algorithms typically used in computer programs.

#### Part 1: Enter List Items

Enter your items in the field below, each on a separate line. Items can be numbers, names, email addresses, etc. A maximum of 10,000 items are allowed. Please don't enter anything you would consider confidential ([here's why](#)).

1  
2  
3  
4  
5  
6  
7

(you're viewing this form securely)

#### Part 2: Go!

Be patient! It may take a little while to randomize your list...

Figure 4 3 Input code for random paper selection

With the random selection website, a random list of order is generated for the selection of papers. I selected the number of papers according to the decision made previously for the current plan in Table 4.3, that is, different number of articles were selected based on the journal publishing requirement on word count. Journals with longer texts had fewer articles selected, and journals with shorter texts had more texts selected. For example, in selecting papers of the 2018 Issue 4 (*The Maritime Economics and Logistics*, a journal with shorter texts), five papers should be selected according to the current plan in Table 4.3. This means that the top five number presented in the random selection list should be considered. As Figure 4.4 shows, 1, 6, 3, 2, 5 are the top five numbers, indicating the first, second, third, fifth and sixth papers in this issue were chosen as the samples. Therefore, the papers coded as MEL180401, MEL180402, MEL180403, MEL180405, and MEL180406 were samples selected. Such process was repeated to all 180 issues, and in total 720 papers were chosen. Metadata of these selected 720 articles were stored separately in an Excel working sheet.

Home Games Numbers Lists & More Drawings Web Tools Statistics Testimonials Learn More  
Login

# RANDOM.ORG

Search RANDOM.ORG

True Random Number Service

**Advisory:** We only operate services from the RANDOM.ORG domain. Other sites that claim to be operated by us are impostors. If in doubt, contact us.

## List Randomizer

There were 7 items in your list. Here they are in random order:

1. 1
2. 6
3. 3
4. 2
5. 5
6. 4
7. 7

IP: 109.175.189.57  
Timestamp: 2023-05-16 16:07:43 UTC

**Don't use this service for giveaways! Use Multi-Round Giveaways instead**

---

© 1998-2023 RANDOM.ORG  
Follow us: [Twitter](#) | [Facebook](#)  
[Terms and Conditions](#)  
[About Us](#)

Figure 4.4 Examples of random paper selection

### *The actual size of corpus*

After the selection of all 720 texts as samples in the LAWEC, the actual corpus size needs to be checked, since there is only an estimation of the corpus size presented in the Table 4.3 based on the maximum word count of articles. All texts were converted into txt. format, and only the whole main body texts were included, following the selection criteria in the 4.1.1. The cleaning process was done manually, so was the checking on converted texts. All converted texts were subject to the #LancsBox 6.0 (Brezina et al., 2021) for word count calculation, and the specific information is presented in the Table 4.6. In total, the LAWEC contains 720 texts and 5,357,636 tokens.

Table 4.6 The size of the Logistics Written Academic English Corpus

Journal Code	Issues Selected	Papers Selected	Maximum Word Count	Total Tokens
TR	20	60	12,000	549,966
BL	20	60	12,000	554,876
PDLM	20	80	10,000	630,471
LR	20	80	10,000	590,118
LM	20	80	10,000	630,043
ML	20	80	10,000	547,931
NRL	20	80	10,000	658,215
OL	20	100	8,000	595,391
MEL	20	100	8,000	600,490
<b>Total</b>	<b>180</b>	<b>720</b>	<b>n/a</b>	<b>5,357,636</b>

### 4.1.3 The evaluation of corpus

The final stage for corpus compilation is the evaluation of corpus in terms of its representativeness and quality (Biber, 1993). As discussed at the beginning of this chapter, this section addresses the last one of three stages in the LAWEC compilation: evaluating the extent to which the samples could represent the target population and how my corpus realises the principles.

#### *Representativeness*

The LAWEC realises the principle of representativeness through five criteria: domain consideration (Egbert et al., 2022), register consideration (Biber, 1993), journal selection, random selection of sampling units and authorship. In terms of domain consideration, all samples collected in the LAWEC are written academic journal articles, which directly fits the domain of the LAWEC and the purpose of corpus compilation in this research: to identify specialised vocabulary and its polysemous sense(s) in written academic Logistics English.

Regarding the first criteria register consideration, the LAWEC has a special focus on academic journal articles. Although it only focuses on one register in academic written language in Logistics English, accumulating evidence exists to show that academic journal article may be more suitable to represent academic written language. This is because of its more informative feature in conveying language and discipline-specific knowledge, a wider range of lexis in use, length for practical analysis, and a wider population of potential audience (Conrad, 2001; de Chazal, 2014; Lu, 2018; Roesler, 2020). It should be acknowledged that academic journal article is only one of registers in academic written

language, which indicates that the LAWEC compilation has bias toward a limited register and may have influence on interested linguistic phenomenon involved in analysis. Nevertheless, it is more feasible to focus on one register in a specialised corpus (Koester, 2022) and have detailed analysis fitting my research aims on identifying specialised words and their polysemous sense(s).

For the third criteria journal selection, all texts are from the top journals in the Logistics field recommended by JCR, an annually-updated, multidimensional-assessed and transparent academic organization with good international reputation (Clarivate, 2022). Journal selection based on recommendation in the JCR helps to realise the representativeness because such recommendation includes journals that are likely to be read by academics and professionals in the field of Logistics. The more likeliness the journals may be read, the more probabilities that the words in use may be useful for specialised vocabulary identification and could be the typical linguistic phenomenon in analysis. Thus, it helps generate more accepted results for specialised Logistics vocabulary. The journal selection based on the JCR also ensures the quality of corpus, since all journals selected are peer-reviewed and are with good reputation in terms of Impact Factor. Moreover, a professional in Logistics also shows her familiarity with the recommended journals, which implies that these selected journals are authentically used and widely accepted in Logistics.

Fourth, random issue and paper selection also contributes to the principle of representativeness. As discussed in 4.1.2, all issues and papers are randomly chosen on a random selection website. Such process keeps the objectivity in sample selection, since all selection were automatically conducted by an online tool rather than the subjective criteria which may have bias toward some articles. Relying on a random selection website could largely minimize human involvement and bias in selection, and ensures the reliability and representativeness in sampling method.

Finally, in terms of authorship, the metadata of the texts selected shows that the contributors are with a wide range of demographic information. For all 720 texts in the LAWEC, about 90% are co-authored. There exist more than 800 authors from about 100 institutions in more than a dozen of countries and regions around the world, of which China (including Hong Kong, Macao, and Taiwan), the US, the UK, the Netherlands and Germany are the top five countries. The co-authored feature and a large number of authors involved indicate that the samples included can largely reduce the bias toward the influence of a few authors in academic writing style, and the texts can be considered as the products of teamwork and collaboration (Lu, 2018). Moreover, the demographics of authors implies that the samples are not limited to a certain English variation. Instead, it represents varieties of English worldwide and shows its representativeness in English language use. Therefore, specialised Logistics vocabulary selection and word sense analysis based on the LAWEC can generate more accepted results.

### *Balance*

The LAWEC realises the principle of balance on three aspects: numbers of issues selected from each journal, consideration on balance between the number of papers selected in each issue and the tokens chosen from each journal, and the balance between tokens and texts in terms of isolexical or isotextual corpus.

First, equal number of issues was selected from each journal, namely four issues a year and 20 issues a journal. This keeps the balance on issues between each journal. Indeed, it does not take the difference of the number of issues and papers published in each year and by each journal into consideration (see Table 4.1 on basic information of selected journals), and it may cause the different probabilities for selecting issues and papers. For example, TR publishes 12 issues each year while there are eight in NRL. This means that the probability for selecting an issue in TR is around 8% ( $1/12=0.083\approx 8\%$ ) while that in NRL is around 12.5% ( $1/8=0.125=12.5\%$ ). It may not reflect what living language situation is and may have impacts on linguistic distribution in terms of specialised vocabulary used across topics in Logistics. However, I intentionally selected the equal number of issues from each journal for the clarity and feasibility of sampling method. Moreover, the topics of Logistics is not within the scope of this research, since this research aims to identify specialised Logistics vocabulary in the Logistics discipline.

Second, the current plan of samples selection (Table 4.3) addresses the balance between number of papers in selection and tokens of each journal. Compared to the original plan that selected the same number of papers from each journal, I adjusted on the number of papers selected according to their maximum word count. If the same number of papers were selected, then the journal with larger maximum word count in papers would have more tokens, and thus it may cause great imbalance between journals. As discussed in section 4.1.2, if the original plan was applied, the number of tokens selected from journals with shorter texts was at least 1/3 smaller than that from journals with longer texts. My trial on the original plan proves it. For example, in the trial, TR (20 issues, 80 papers, 12,000 maximum word count) had around 800,000 tokens, while that of OL (20 issues, 80 papers, 8,000 maximum word count) was around 470,000. There existed a difference of around double in size between these two journals in the original plan. However, such difference is eliminated to a great extent after adjustment: TR (60 papers) with around 540,000 tokens and OL (100 papers) around 590,000 tokens (Table 4.6). The gap in size of selected articles from these two journals was largely narrowed to around 10%. For all nine journals, the biggest gap in size of selected articles was narrowed to around 20%, compared to that of 50% in the original plan.

Third, it addresses the issue of balance between tokens and texts. As mentioned by Oakey (2009), it is inevitable to have imbalance between tokens and texts selected within a corpus. According to his suggestion, meeting either of the criteria (a) collecting more than 500 texts or (b) building a corpus with more than 5 million tokens in size can minimize the impact on linguistic distribution caused by

imbalance between tokens and texts. For the LAWEC, it is with 720 texts and around 535,000 tokens, meeting both of the criteria suggested by Oakey (2009). This means that although there exists imbalance between texts and tokens from each journal in the LAWEC, such imbalance may have less influence on the results.

### *Size*

Generally, there is no absolute rule on the size of corpus. For the evaluation on this principle, research aims and the content of corpus need to be considered (Koester, 2022). One of aims is to identify the most frequent specialised vocabulary in Logistic written academic English. This aim belongs to the quantitative analysis of linguistic phenomenon in corpus (Egbert et al., 2022). For such analysis, it is common to check whether the corpus is large enough to permit the reasonable recurrence of relatively low-frequency linguistic types (Egbert et al., 2022; Nation, 2016). This means that the LAWEC should be at least millions of tokens in size to ensure the occurrence of lower frequency words. As discussed earlier, the LAWEC consists of 720 journal articles from nine top journals in the Logistics field with 5.35 million tokens. It first meets the general criteria of being beyond one million tokens in size. Such criterion is generally considered as the threshold for (specialised) vocabulary list development with modern corpus (Dang, 2020; Nation, 2016). It then meets the suggestion by O’Keffee et al. (2007) that the size of a small written specialised corpus could be around 5 million tokens. Moreover, through my own observation on research of word list development and polysemy analysis with corpus, the size of my corpus is larger than that in 95% studies involved (discussed in section 4.1.1). This indicates that the size of my corpus is acceptable and relatively large enough in my field, and it may provide reliable results for my analysis.

## 4.2 Logistics specialised vocabulary selection criteria and list development

### 4.2.1 Decision on unit of counting

For identifying specialised vocabulary and developing a word list, unit of counting should be first decided and then is the specific selection criteria (Webb, 2021). This is because different unit of counting can generate different results for word list development (Dang, 2020; Nation, 2016). The decision depends on the research purposes and the potential users’ proficiency (Dang, 2020; Schmitt, and Schmitt, 2020) (see details in section 2.2). For my research, the specialised vocabulary selected in Logistics academic written English is for further polysemous sense(s) analysis, and aims to be inclusive

for a wide range of audience interested in Logistics, either academic or professionals, or with high or low L2 proficiency.

With research aims considered, word type may be the best suitable one for my research, among four common counting units in word list development (word type, lemma, flemma and 6-level word family). It is decided based on four aspects. First, studies (Chon and Shin, 2013; Sinclair, 1991, 2003; Stubbs, 2001) have shown that the difference of word form could lead to different collocational behaviours and then may influence the word sense(s) in context. Such different collocational behaviours and influence on word sense(s) led by word forms happens to both inflections (Sinclair, 1991) and derivatives (Bauer and Nation, 2020). This suggests that some specialised senses could be ignored or overestimated if the inflections or derivatives are counted as one unit. Second, the blurred boundary between sub-categories of specialised vocabulary (academic and technical vocabulary) indicates that specialised vocabulary may have discipline-related or technical meanings in a specific context (Nation, 2013). If the specialised vocabulary contains technical meanings, it is probable that it forms terms/technical phrases where word forms cannot be changed casually (Coxhead et al., 2020; Coxhead and Drayton, 2023). This suggests that some specific word types in the same lemma or word family may not have specialised senses in a specific context as others have. Third, research has shown that L2 learners' knowledge of affixes may vary according to their L1 influence and L2 proficiency (McLean et al., 2021). Consequently, the larger counting unit (e.g., 6-level word family) may not be suitable for the inclusion of wider population of audience, since there is no guarantee that a learner could know all affixes of a 6-level word family (Schmitt and Schmitt, 2020; Schmitt, 2023). Fourth, as discussed in section 2.2, word type is recognized as a common counting unit for specialised vocabulary research and word list development in specific contexts. This counting unit has been widely used in different specific contexts, for example, Chinese Traditional Medicine (Lu, 2018), Rugby (Coxhead and Benson, 2022), Aviation (Coxhead and Drayton, 2023).

#### 4.2.2 Software selection

Numerous corpus analysis software exists, usually categorized as online, offline and DIY software (Anthony, 2022). For choosing a suitable software in corpus analysis, research aims, software availability, raw data access and ease of practice are to be considered. One aim of my research is to develop a specialised Logistics vocabulary list by identifying the most frequent specialised vocabulary in Logistics written academic context. This means that the software I choose must have the function of word list and I have the access to upload my raw data. Also, the software should be able to practice common objective selection criteria of word list development (e.g., keyword, range, dispersion and frequency). #LancsBox 6.0 (Brezina et al., 2021) was selected as the software for specialised Logistics vocabulary list development. This software has four strengths. First, #LancsBox 6.0 is available on the

website for downloading, and it is free to use (Brezina, 2018). This means that there is no need to have additional license or payment for such software. Second, it has the ability to process large corpus and its raw data in a relative short period of time and generally few worries of computer crash. Third, it provides various statistical indicators for quantitative corpus analysis (Brezina, 2018). For example, it provides several options in calculation of frequency, range and dispersion respectively, and thus it is able to meet the needs of applying different selection criteria. Fourth, it contains around a score of in-built corpora which can be selected as reference corpora without uploading new corpus if not necessary. In sum, #LancsBox 6.0 is free for download and use, adequate in statistical analysis and easy to practice and process corpus quantitative analysis, fitting my aim on identifying specialised vocabulary in Logistics based on objective selection criteria.

### 4.2.3 Specialised vocabulary selection criteria

This section presents my selection criteria for specialised vocabulary list development: keyness, dispersion, range as corpus-driven criteria. Human involvement for focusing on lexical words is also one of the criteria, dealing with exclusion of function words and having individual lists for acronyms, proper names, and foreign words.

#### *Keyword analysis*

Trough keyword analysis, the aboutness of the corpus of interest can be shown (Archer, 2009; Brezina, 2018). Keyword analysis extracts items with significant different frequency than those in the reference corpus (Thomson and Hunston, 2020). With keyword analysis, how language (vocabulary in use in my research) is used differently across corpora can be manifested, instead of relying only on frequency (Brezina, 2018), since frequency itself cannot directly show to what extent the statistics are different (Baker, 2006). Keyword analysis has been applied as a key criterion in specialised vocabulary list development (Paquot, 2007; Piribabadi and Rahmany, 2014; Tongpoon-Patanasorn, 2018).

For keyword analysis, two issues need to be considered: reference corpora selection and statistical index decision. In terms of reference corpora selection, there is no absolute rule, since it depends on the research aims and the size of corpus (Brezina, 2018; Hunston, 2022). Generally, the reference corpora should be as large as possible (Brezina, 2018), and it can be mega general corpora like BNC or COCA (Weisser, 2015). Rayson (2019, p.322) recommends three criteria in selecting reference corpus: “representativeness, homogeneity and comparability”. Representativeness refers to whether the corpus contains main types of texts in general language; homogeneity is for internal consistency of components; comparability addresses for the similarity of sampling methods (Rayson, 2019). For practical selection,

BNC 2014 Baby+ (Brezina, 2019) was selected as the reference corpus in this research based on four reasons. First, BNC 2014 Baby + is already attached to #LancsBox 6.0, without further need to upload new data for reference corpus. This brings convenience in specialised vocabulary list development with keyword analysis. Second, this corpus covers a wide range of registers used in general English (Brezina, 2018). If set as the reference corpus, it can help with identifying typical linguistic phenomenon of vocabulary in use in Logistics English. Third, it is one of the largest corpora attached to #LancsBox 6.0 dominant with written language (about 80% in written English, and 20% in spoken English) (Brezina, 2018), which fits my research aim in identifying specialised vocabulary in Logistics written language. Fourth, it is the latest general corpus (containing data from 2010 to 2017) attached to #LancsBox 6.0, which, to some extent, represents the current authentic English language use rather than those in past times. Indeed, although BNC 2014 Baby+ may not be ideal in terms of homogeneity and comparability, it is a suitable reference corpus in terms of representativeness and practicality, with good availability and accessibility.

For statistics index decision, #LancsBox 6.0 provides five indexes on keyword analysis: simple maths, LogLik, %DIFF, LogRatio and Cohen's d. In my research, Cohen's d was chosen due to its strength of taking dispersion into consideration (Brezina, 2018). This is because corpus research (e.g., Dang, 2017, 2018a, 2018b; Roesler, 2020) notices that some items may distribute unevenly across topics and texts as technical words which are used in a limited range of texts but have high frequency in corpus. These items may not fit my aims to identify the most frequent and evenly distributed specialised vocabulary used across Logistics academic journals. By applying Cohen's d, these items can be excluded, to some extent. The statistical threshold of Cohen's d is beyond 0.2, since those below may present an extreme uneven distribution within the corpus. Cohen's d 0.2 is interpreted as "minimal important difference value" (Pojanapunya and Waston-Todd, 2021). Moreover, it is also the default value in #LancsBox 6.0 (Brezina, 2019).

### *Dispersion*

Dispersion manifests the evenness of the linguistic item distribution within the corpus (Dang, 2017). For specialised word list development, it is one of the criteria to exclude technical words with extreme limited range in use, that is, the items are infrequent terms that are only used rarely in several texts (Lu, 2018). #LancsBox 6.0 provides CV, Juilland's D, DP and  $DP_{norm}$  as main statistical indexes of dispersion (Brezina, 2018). In my research,  $DP_{norm}$  was chosen. This is because other indexes like CV and Juilland's D depend on the number of corpus (Brezina, 2018), while DP and  $DP_{norm}$  are not. Moreover,  $DP_{norm}$  is the normalization of DP and concerns with the issue of range (Gries, 2008; Lijffijt and Gries, 2012). It falls between (including) 0 and 1, where being closer to 0 represents evenly distribution while to 1 for unevenness. Items with  $DP_{norm}$  beyond 0.95 are excluded, following Deignan et al. (2023).

### *Range*

Range is also a common criterion in corpus-derived word list development (Dang, 2020). It addresses how many files the items of interest occur in the corpus (Coxhead, 2000). The cutting-off point of range was set that an item needs to occur in all nine journals involved and at least 50% of texts, namely 360 articles in my corpus. This threshold was chosen for being consistent with previous specialised word list development research (Coxhead, 2000; Wang, Liang and Ge, 2008; Hus, 2013; Khani and Tazik, 2013; Dang, 2017). Such decision ensures all lexical item candidates are used frequently in the field of Logistics and may bring benefits to potential users for specialised vocabulary learning through the by-product word list.

### *Lexical words*

Focusing on lexical words is the final criterion. This involves two issues: to exclude function words and to exclude other non-lexical words. For exclusion of function words, a function word list is developed (Appendix A) to extract specialised words. This is because the function words usually do not convey specific information (McCarthy, 2023), and thus they may not be related to discipline-specific senses. Indeed, there exist scholars arguing that function words may have their own functions in conveying meaning (cf. Sinclair, 1991, 2003; Weisser, 2015). However, considering my research aims are to identify specialised Logistics vocabulary for discipline-specific words and polysemy analysis, I decide to focus on lexical words which explicitly express specific meaning in context. The function word list consists of 306 items, with reference to West (1953), Biber et al., (1999), Brezina and Gablasova (2015) and Nation (2022). Nation's (2022) list was chosen as the blueprint, and each other reference was checked manually, with repeated items excluded. All function words in my corpus meeting previous selection criteria will be presented in an individual word list.

There can be other non-lexical items not included in the aforementioned function word list, for example, acronyms (e.g., "SCM" representing "supply chain management"), proper names (e.g., China), and foreign words (e.g.,  $\beta$ ). The examples presented in the brackets are not all such items in a thorough list but intentionally extracted from the word list for the whole corpus generated on #LancsBox 6.0. These items deserve to be addressed since they may convey discipline-related information implicitly or can be considered as non-words that are less useful in disciplinary learning. Specialised word list developers do not normally cover these items in their lists, for example, Yang (2015), Dang (2017, 2018a, 2018b), Ashrafzadeh (2021) and Khany and Kalantari (2021). However, in my research, I decide to have other individual lists to present these items. Nation and Kobeleva (2016) suggests these items

meeting selection criteria can be presented in an additional individual list, because they may be related to the content of texts and attract interest of readers.

Specifically, first, all acronyms will be presented in an acronyms list. This is because they are usually proper names, having specific unique reference that may be highly related to the context. It is also because “they are clearly distinguishable from other types of words” (Nation, 2016, p.86). For example, SC (supply chain, frequency = 2098;  $DP_{norm} = 0.92$ ; range = 60) may be useful for language users in Logistics field, since “purchasing, materials handling, **logistics**, transportation, inventory control and **supply chain management** have continued to evolve, causing many of these functional areas to intersect with one another” (Michigan State University, 2023, original emphasis). Here below is a concordance extract on SC:

*The proposed SC design model used both cross-functional and logistical SC drivers to build a binary integer programming model.*

(LM180116, emphasis added)

These suggest that the term “supply chain” and its acronyms “SC” are expected to be linked with Logistics. Therefore, I believe that it is worth listing acronyms in another individual list.

Second, proper nouns/names and foreign words needed to be addressed and extracted. For proper nouns/names, Nicklin, Patterson and McLean (2023) empirically argue that the assumption that learners generally have no difficulty in recognizing such words can be a myth. They thus suggest proper nouns/names not be in an off-list. Rather, these words should be taken into consideration of lexical coverage. However, due to the fact that my research aims are not closely related to examining lexical coverage in Logistics field, these non-lexical items were excluded from the final list because of their less relevant to disciplinary knowledge or concepts. Nation and Kobeleva (2016) recommend prototype theory in dealing with proper names with intuition. With this, personnel names, one of the core inevitable prototypical categories of proper names in language use, can be identified. For example, “Xie” (a Chinese surname; frequency = 74;  $DP_{norm} = 0.95$ ; range = 41) will be excluded. Such criteria can also be applied to identifying and excluding foreign words, for example, Roman numeral “iii” (frequency = 548;  $DP_{norm} = 0.68$ ; range = 231) or mathematical symbol “ $\beta$ ” (frequency = 855;  $DP_{norm} = 0.82$ ; range = 138). Here below is a concordance extract for “Xie”, “iii” and “ $\beta$ ” respectively:

*Specifically, some logistics enterprises engaged in NLSPD only provide basic services, such as transportation and warehousing, while others provide additional functions, such as logistics finance services (Chen and Xie, 2009).*

(ML190409, emphasis added)

*Three streams of literature are therefore relevant in positioning this paper – (i) supply chain visibility and transparency, (ii) blockchain basics, and (iii) data capture on and for the blockchain.*

(BL210304, emphasis added)

*The interaction term between financial inclusion and ID is insignificant ( $\beta = -0.3$ ,  $p = .73$ ) and so is the direct effect of financial inclusion.*

(BL210303, emphasis added)

However, it is worth reading concordance lines to have a double check to see whether the items are related to Logistics disciplinary knowledge or academic discourse. For example, although *Likert* (frequency = 161;  $DP_{\text{norm}} = 0.86$ ; range = 100) is a personnel name, it now often collocates with *a* and *scale*, forming the multi-word unit “a Likert scale” and referring to a common research instrument in academic field. Here is a concordance extract of *Likert*:

*Second, because the dependent variables were measured on a **Likert** scale from 1 to 5, each model was also estimated using an ordered logistic regression (Winship and Mare, 1990).*

(BL190301, emphasis added)

This indicates that knowing proper names like “Likert” may be helpful in reading academic journal articles. Therefore, I believe that it is necessary to include and present such items in an individual list.

### *Pilot study of specialised vocabulary selection*

The LAWEC was uploaded to #Lancs6.0 (Brezina et al., 2021) for a pilot study in selecting specialised vocabulary. The aforementioned selection criteria are followed, namely

- (i) Lexical unit: individual word type;
- (ii) Keyword analysis:  $\geq 0.2$  in Cohen’s  $d$  (with BNC2014 Baby+ as the reference corpus);
- (iii) Dispersion:  $\leq 0.95$  in  $DP_{\text{norm}}$ ;
- (iv) Range:  $\geq 360$  texts in range;
- (v) Lexical words;
- (vi) Individual list for function words, acronyms, and discipline-/academic-related proper names and foreign words.

With all criteria applied, the pilot study presents the top 30 specialised items in list, ranking from high to low in Cohen's  $d$  value (see Table 4.7). On checking the items in pilot study manually, it seems that the specialised Logistics vocabulary selection criteria work well in extracting academic vocabulary. Some discipline-specific vocabulary (e.g., "supply", "chain", "logistics") shown in the pilot study were not presented in the table below because of their relatively low ranking in Cohen's  $d$ . The next step will be developing the full specialised Logistics vocabulary list. LoSAWEL (see the full list in Appendix B) and analysing these word senses in the LAWEC.

Table 4.7 Top 30 specialised selected items (ranked by Cohen's d from high to low)

Rank	Items	Cohen's d	DP <sub>norm</sub>	Range	Frequency
1	results	1.71	0.29	698	8485
2	abstract	1.69	0.19	679	768
3	study	1.46	0.33	689	10867
4	introduction	1.45	0.23	706	981
5	based	1.41	0.29	700	6626
6	literature	1.40	0.33	678	6297
7	section	1.31	0.37	605	4570
8	paper	1.31	0.32	667	4407
9	table	1.25	0.33	623	5032
10	model	1.24	0.40	664	14757
11	following	1.23	0.30	692	4438
12	considered	1.21	0.36	661	3695
13	provide	1.18	0.32	642	2936
14	related	1.15	0.36	641	3457
15	provides	1.15	0.36	573	1965
16	consider	1.13	0.35	602	3034
17	due	1.13	0.33	642	3664
18	analysis	1.13	0.34	680	6531
19	increase	1.11	0.37	628	3545
20	result	1.10	0.33	634	2933
21	follows	1.09	0.43	593	2716
22	respectively	1.09	0.45	538	2247
23	example	1.09	0.33	648	4564
24	addition	1.09	0.39	566	2271
25	research	1.06	0.39	687	13112
26	shown	1.05	0.38	599	2700
27	implications	1.04	0.85	457	1667
28	using	1.03	0.32	689	6190
29	finally	1.02	0.35	596	1940
30	according	1.02	0.38	593	2727

### 4.3 Word sense analysis

This section presents procedures of word sense analysis through corpus, including tool choosing (section 4.3.1), collocation analysis, span decision and LogDice threshold (section 4.3.2), concordance

reading (section 4.3.3) and word sense identification and splitting through dictionaries (section 4.3.4). Word sense analysis is the preparatory stage of answering RQ2 on word sense comparison in three corpora (see section 4.4) and RQ3 on polysemy in Logistics specialised vocabulary. Word sense analysis was conducted in three corpora, with the same procedures. In this section, I provide examples of word sense analysis procedures in the LAWEC, the self-built specialised Logistics corpus.

### 4.3.1 Tool choosing for word sense analysis

The Sketch Engine (Kirlgariff et al., 2014) was selected as the analysis tool for polysemous sense identification. It is an online-based corpus analysis website which is connected to more than 100 corpora in different languages (Kirlgariff et al., 2014). This website is chosen for three reasons. Firstly, it provides different options in calculation of collocation strength (e.g., MI, MI2, MI3, T-score, Loglikelihood), which meet different needs in collocation analysis. Since it is usually to analyse the word through its collocations (Firth, 1957), this may be helpful in senses identification with collocation checking.

Secondly, the Sketch Engine has the function of “Random selection” on concordance lines. This function helps with reading a large amount of concordance lines. For example, there are thousands of occurrences of each specialised Logistics word in Table 4.7 generated from the pilot study of specialised word list development. With random selection of concordance lines applied, it is more practical to conduct word sense analysis in concordance reading. Stratified selection or random selection of concordance lines is also a common method in corpus linguistic research (Weisser, 2016). On the Sketch Engine, it has the default value on 200 concordance lines for random selection (Kirlgariff et al., 2014). However, the specific amount of concordance for random selection can be re-set according to research purposes and to a large extent reduces the bias of selection manually, which contributes to the reliability and validity of research. The issues of collocation checking and random selection on concordance lines will be discussed further later in the section 4.3.2.

Thirdly, the Sketch Engine is connected to more than 100 corpora suitable for different contexts or languages. For example, it has the access to the British Academic Written English Corpus (BAWE), BNC1994, English WEB 2021 (ENTENTEN21), covering a wide range of language contexts in use. The wide cross-corpora access fits my research aims to compare word senses of target words in specialised, general academic and general use contexts. With the access to more corpora, I could choose more representative reference corpora in sense comparisons, which helps addressing RQ2. For cross-corpora comparison in word senses and analysis of polysemous words, it will be discussed later in the section 4.4.

### 4.3.2 Collocation checking

Collocation analysis helps with disambiguating senses of words through other near-by words (Hunston, 2022). Therefore, it was taken as a step to analyse word senses in contexts and identify polysemous specialised Logistics words, addressing the RQ2 and RQ3 of this research. Considering that the target word needs to be analysed through the items it companies (Firth, 1957; Deignan et al., 2023; Partington, 1998; Sinclair, 1991), the items presented in the word list are set as nodes and analysed together with their collocations. The collocation analysis consisted of four following steps:

The first step was setting the node words for the analysis. Items from the specialised Logistics word list developed in RQ1 was used as the node words. One word (type) was used for each searching.

The second stage was selecting the collocation measure. LogDice was chosen as a measure of collocation strength in the presented study because it could be compared across corpora in different size and overcome the bias of MI and T-score which are influenced by the frequency rank order of collocates (Gablasova, Brezina and McEnery, 2017). As for the LogDice threshold, Pinto et al. (2023) suggest that  $\text{LogDice} \geq 7$  can be the criterion. However, the threshold depends on the purposes of research and the number of collocations generated. The threshold suggested by Pinto et al. (2023) can be a reference, but mine was finally decided as  $\text{LogDice} \geq 9$  after a pilot study which will be discussed in the section 4.3.5.

The third stage was defining the span of the search. I decided to choose +/-5 as the span. the widest span setting in the Sketch Engine. Adopting +/-5, the widest span setting in Sketch Engine, would be more likely to include both continuous and discontinuous collocations. This may help with providing a better picture on word sense in analysis. If a narrower span is applied, it is then more likely to include adjacent and continuous collocations, with much ignorance of discontinuous collocations which also contribute to word sense. For example, in searching collocates of the node word “model” with +/-5 and LogDice, the collocate developed, with 543 occurrences, is at one of the top positions as typical collocate (presented in Figure 4.5). Then with examining the concordance lines of the collocational behavior “developed – model” (shown in Figure 4.6), it can be seen that “the model developed by” (e.g., concordance 1 and 3) is a longer continuous collocation, while “developed (...) a model” is a discontinuous collocation (e.g., concordance 11, 12, 16). Although there is no clear distinction between these two collocations in terms of senses of “model” in these example concordances, these two kinds of collocations both have their role in the collocational behavior of the node word and its senses in context. If a narrower span, for example, +/-3, is applied, it is likely to miss some concordances. This will probably affect to what extent a sense is used for a specific word type in the LAWEC, for example, the probabilities of a certain word sense. Therefore, a wider span is needed.

CONCORDANCE Logistics Academic Written English Corpus

word model • 14,790  
2,318.4 per million tokens • 0.23%

Collocations [CHANGE CRITERIA](#) [BACK TO CONCORDANCE](#)

	Word	Cooccurrences <sup>?</sup>	Candidates <sup>?</sup>	LogDice ↓		Word	Cooccurrences <sup>?</sup>	Candidates <sup>?</sup>	LogDice ↓
1	<input type="checkbox"/> proposed	900	4,238	10.60 ...	11	<input type="checkbox"/> with	1,253	34,629	9.70 ...
2	<input type="checkbox"/> The	1,851	36,587	10.21 ...	12	<input type="checkbox"/> .	6,059	226,760	9.68 ...
3	<input type="checkbox"/> our	747	7,191	10.12 ...	13	<input type="checkbox"/> in	2,762	98,613	9.64 ...
4	<input type="checkbox"/> a	3,470	88,175	10.11 ...	14	<input type="checkbox"/> is	2,110	74,604	9.60 ...
5	<input type="checkbox"/> developed	543	2,757	9.99 ...	15	<input type="checkbox"/> we	752	17,646	9.57 ...
6	<input type="checkbox"/> model	885	14,109	9.97 ...	16	<input type="checkbox"/> this	833	22,097	9.53 ...
7	<input type="checkbox"/> an	964	18,933	9.87 ...	17	<input type="checkbox"/> In	723	17,790	9.51 ...
8	<input type="checkbox"/> for	1,733	49,307	9.79 ...	18	<input type="checkbox"/> to	2,930	123,187	9.44 ...
9	<input type="checkbox"/> the	9,108	327,929	9.77 ...	19	<input type="checkbox"/> was	504	10,241	9.37 ...
10	<input type="checkbox"/> inventory	516	4,996	9.74 ...	20	<input type="checkbox"/> mathematical	304	692	9.33 ...

Figure 4.5 Screenshot of part of collocates of *model*

CONCORDANCE Logistics Academic Written English Corpus

word model • 14,319  
2,244.57 per million tokens • 0.22%

filter developed -5, 5 • 540  
84.65 per million tokens • 0.0085%

Sample 10 • 10  
1.57 per million tokens • 0.00016%

Details  Left context  KWIC  Right context

1	<input type="checkbox"/> <input type="checkbox"/> wine production process.</s><s>The author <b>developed</b> a decision <b>model</b> that facilitates the selection of which postponement types should be
2	<input type="checkbox"/> <input type="checkbox"/> rfirming the internal consistency, a performance- based hierarchical <b>model</b> has been <b>developed</b> as shown in Figure 1. 2.1.</s><s>Model constr
3	<input type="checkbox"/> <input type="checkbox"/> ne for a given number of emergency facilities.</s><s>The P-median <b>model developed</b> by Pir- kul and Schilling (1988) set out to minimise the to
4	<input type="checkbox"/> <input type="checkbox"/> hat characterise any real-life inventory situation simultaneously, the <b>model</b> has been <b>developed</b> in the mixed fuzzy random environment.</s><s>
5	<input type="checkbox"/> <input type="checkbox"/> ashflow approach.</s><s>Shah, Gor, and Wee (2010) <b>developed</b> the <b>model</b> to study the effect of deterioration when demand is price-sensitive.<
6	<input type="checkbox"/> <input type="checkbox"/> :/s><s>Hemapriya and Uthayakumar (2017) <b>developed</b> an inventory <b>model</b> with uncertain demand and lost sales reduction under service level
7	<input type="checkbox"/> <input type="checkbox"/> ned its effect on EOQ model.</s><s>Chiu (2003) <b>developed</b> an EPQ <b>model</b> and assumed production system could produce defective products.
8	<input type="checkbox"/> <input type="checkbox"/> : last but not least, Cavalcante et al. (2019) <b>developed</b> an integrated <b>model</b> with a simulation of machine learning to study the risk of suppliers t
9	<input type="checkbox"/> <input type="checkbox"/> era leading up to deregulation, Corsi and Martin (1982) <b>developed</b> a <b>model</b> to explain turnover among owner operators.</s><s>Beilock and Cap
10	<input type="checkbox"/> <input type="checkbox"/> :/s><s>Design/methodology/approach – An exploratory conceptual <b>model</b> was first <b>developed</b> from a literature review of the general barriers tr

Figure 4.6 Concordance lines examples of the node *model* and the collocate developed

The final stage was focusing on collocates presented by words, rather than punctuations and numbers. It is also necessary to take the number of concordance lines of collocates for checking in consideration. In general, the higher frequency the collocates have in the LAWEC, the more concordance lines will be

generated, and thus the more information will be gained for senses identification for my research. Moreover, with more occurrences of collocates, the collocates and the node can form significant collocations which typically represent the collocational behavior of the node word (Hunston, 2002), and this may contribute to identifying not only how many senses but also the dominant sense(s) of the specialised words used in contexts. However, the node words in query may have dozens of collocates and each collocate may have hundreds of or thousands of concordance lines generated in the LAWEC. It could not be practical to check each concordance line generated from each collocation searching. Therefore, to have a balance between practicality and validity of results in concordance line checking and word sense analysis, a decision was made to have random selection on concordance lines whose number is beyond 1,000. For those below 1,000, all concordance lines will be checked. Let's take the top 10 collocates of the node *model* as an example. As Figure 4.7 shows, four of top 10 collocates of *model* have more than 1,000 occurrences, namely "The", "a", "for" and "the", with occurrences of 1,851, 3,740, 1,733, and 9,108 respectively. The "get a random sample" will be applied to these four collocates when reading their concordance lines.

word *model* • 14,790
2,318.4 per million tokens • 0.23%
i

⚡

## Collocations

CHANGE CRITERIA
BACK TO CONCORDANCE

	Word	Cooccurrences <sup>?</sup>	Candidates <sup>?</sup>	LogDice ↓
1	<input type="checkbox"/> proposed	900	4,238	10.60 ...
2	<input type="checkbox"/> The	1,851	36,587	10.21 ...
3	<input type="checkbox"/> our	747	7,191	10.12 ...
4	<input type="checkbox"/> a	3,470	88,175	10.11 ...
5	<input type="checkbox"/> developed	543	2,757	9.99 ...
6	<input type="checkbox"/> model	885	14,109	9.97 ...
7	<input type="checkbox"/> an	964	18,933	9.87 ...
8	<input type="checkbox"/> for	1,733	49,307	9.79 ...
9	<input type="checkbox"/> the	9,108	327,929	9.77 ...
10	<input type="checkbox"/> inventory	516	4,996	9.74 ...

Figure 4.7 The top 10 collocates of *model* and their occurrences in concordance lines

Specifically, for those collocates with more than 1,000 concordance lines, I will set different amount of concordance lines in “get a random sample” for four rounds of random selection on each collocate, namely, 100, 200, 300 and 400 random concordance lines in selection. For setting different statistics for the amount of concordance lines, this is because the Sketch Engine can only provide different concordance lines randomly with different statistics set (Kirlgarriff et al., 2014). If the same number, for example, 200 is set for concordances of a collocates, the concordance lines will be the same, regardless of how many times it is searched. To illustrate the random selection and the concordance lines it generates clearly, I selected the collocate “for” of the node *model* as an example, setting the amount of random samples as 200 and 300 respectively. The different sample concordance lines are generated and presented in Figure 4.8 and 4.9. For the four rounds of random selection on concordance lines, this is for the attempt to check more concordance lines to obtain more information for senses identification to a great extent. Each concordance presented will be checked manually, and the steps for checking concordance lines are discussed in the next section.

The screenshot shows a concordance search interface with the following statistics:

- word *model* • 14,790 (2,318.4 per million tokens • 0.23%)
- filter for -5, 5 • 1,733 (271.66 per million tokens • 0.027%)
- Sample 200 • 200 (31.35 per million tokens • 0.0031%)

The interface includes a search bar with 'KWIC' selected, and a list of concordance lines. The concordance lines are numbered 1 to 10 and show the word 'model' in red, with the collocate 'for' in black. The text is truncated on both sides with '</s><s>' markers.

Line	Text
1	doc#0 in was used as a foundation to provide the boundaries and parameters for a model that is testable and likely empirically valid (Eisenhardt 1989; Ketokivi and Ch...
2	doc#0 institution partner [with whom I have a] crucial partnership for my business model .</s><s>There is a business life cycle in these social businesses where certa...
3	doc#0 odern businesses.</s><s>Future researchers can utilize and build upon our model for sustained collaboration with practitioners, as well as our combination of...
4	doc#0 results for testing step (a), model 2 shows the results for testing step (b), and model 3 shows the results for testing both steps (c) and (d).</s><s>Specifically, mc...
5	doc#1 1 dependent variables have theoretical justification for being included in the model .</s><s>RWA assumes the model is properly specified and it should not be...
6	doc#1 is stated that RWA should not be used to statistically select variables for the model .</s><s>If researchers need statistical help in selecting variables for a mode...
7	doc#2 s><s>First, based on costtrade-off insights, we develop a new mathematical model for the last-mile distribution problem, which we call the Driver Helper Dispat...
8	doc#2 ed.</s><s>Finally, in the optimal carton set selection stage, a mathematical model for selecting n3 cartons as an optimal carton set and minimizing the shippin...
9	doc#2 int validity was examined for each pair of constructs appearing in the overall model .</s><s>Average variance extracted for each construct was calculated and c...
10	doc#2 t when an employee is disagreeable, it is an opportunity for the manager to model exceptional self-control and to be courteous and, thereby, have a chance to c...

Figure 4.8 Samples of concordance lines of node *model* and its collocate for (number = 200)

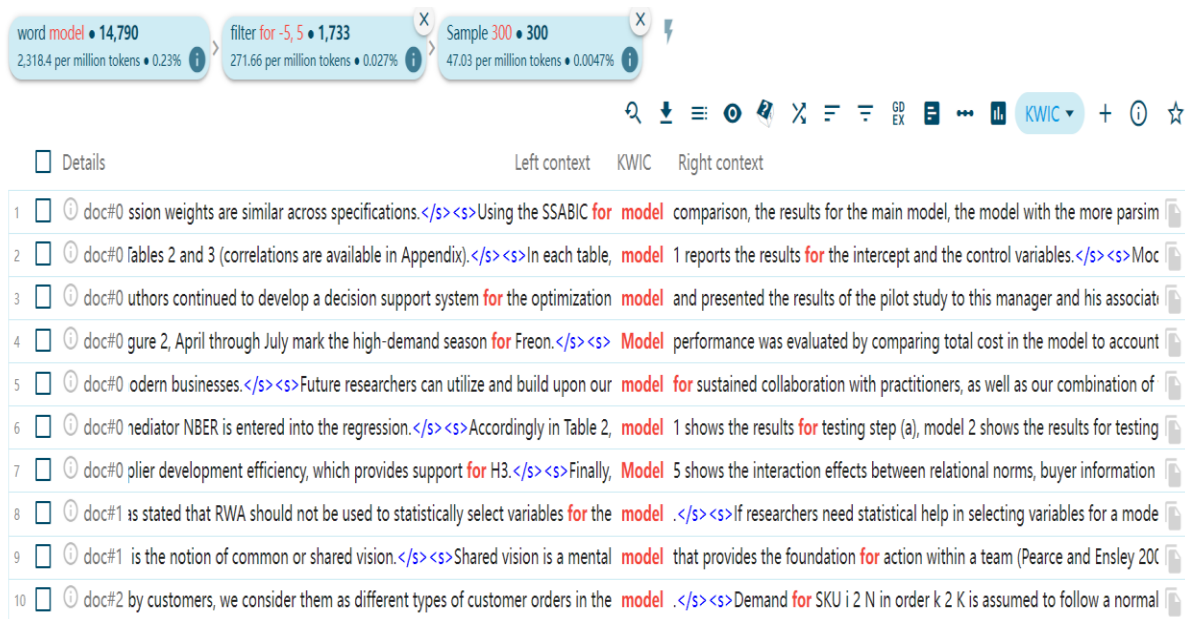


Figure 4.9 Samples of concordance lines of node *model* and its collocate for (number = 300)

### 4.3.3 Reading concordance

The collocation checking mentioned in the previous section helps with contextual clues for word sense analysis. However, collocations only may not be sufficient for word senses identification since they may be incomplete in structure (e.g., the node *model* and its collocate *developed* in longer sequence “the (...) model developed by” discussed above). Without specific contexts such as concordance lines or wider contexts covering several sentences in the original corpus text, it is difficult to define how the nodes are used and the senses the nodes have, because form and meaning are interconnected in contexts (Hunston, 2022a, 2022b; Sinclair, 1991). Concordance lines with node words set in central and texts on both left and right side of node words provide a more complete contexts for word sense analysis in complex lexical-grammatical structures (Viana, 2022), and thus it may be helpful to read concordance lines as a further step in word sense identification through authentic language use.

### 4.3.4 Identifying and splitting senses with dictionaries

Both collocation checking and concordance reading provide contextual clues in identifying senses that the node words have. However, it is difficult to reach an agreement on how many senses an item may have (Moon, 1987) and this should be dependent on the purposes of research. My research aims to identify potential difficulties on polysemous word senses that Logistics language users may have when encountering the specialised items included in lists. This research follows Deignan and Love’s (2021)

study on word sense splitting method, using the narrow splitter in distinguishing everyday senses, general academic senses and those specifically used in specialised academic contexts. For example, there exists a difference between the everyday sense of *values* (the moral principles) and that used in general and specialised academic contexts where it means mathematical statistics. Empirical studies, for example, interviews in Deignan, Semino and Paul (2019) and diagnosis vocabulary test results on polysemy in Mićović and Beko (2022) also show that learners may have difficulty in distinguishing daily and professional senses. Therefore, I believe that such distinction may cause difficulty when language users with less professional experience encounter them.

In practice, senses identification and splitting were assisted with dictionary entries. This is applied by previous scholars such as Skoufaki and Petrić (2021) and Le and Miller (2023). Following Skoufaki and Petrić (2021), *the Collins COBUILD Advanced Learner's Dictionary 10<sup>th</sup> edition* (Collins, 2023), the latest version by the time of writing, will also be used in my research. This dictionary derives from a 4.5-billion-token Collins corpus, which contains authentic examples and could represent language use in real situation. Although Skoufaki and Petrić (2021) acknowledge that the Collins dictionary could underestimate polysemy, this dictionary focuses on senses that are relevant to language learners in learning. A technical dictionary on Logistics was also be used, following Le and Miller's (2023) research where the authors choose a Medical technical dictionary for discipline-specific sense identification. This is because Hyland and Tse (2007) argue that in discipline-specific contexts the words may have specific senses which are quite different from those in other contexts. Collins dictionary, a dictionary for language learning in general use, may not contain the discipline-specific or technical senses, and thus there is a need of technical dictionary. Due to the relative scarcity of recent published discipline-specific Logistic dictionaries, the IWLA Logistics glossary and dictionary [The online Logistic dictionary \(IWLA-Glossary-and-Terminology.pdf\)](#) was selected as reference to distinguish technical senses of items, because of its recent updating and free availability.

For the assistance of dictionary entries in senses identification and splitting, each node word and its collocations were looked up in both dictionaries, for a double check on whether there exists a technical sense. The process will be as follows:

- (1) Reading the concordance lines and focusing on the node word, the word class of the node word, and the collocation in searching;
- (2) Looking up the node word with its corresponding word class presented in the concordance lines in dictionaries, and if necessary, looking up the whole collocation to see whether it is an entry as a phrase;
- (3) Matching the most suitable sense in dictionaries with that presented in the concordance line;
- (4) If the matching fails, reading the wider contexts in original texts for further definition, and if necessary, consulting the concordance lines in question with professionals.

For example, in the LAWEC, node word *model* (frequency = 14757;  $DP_{norm} = 0.40$ ; range = 667) collocates with “mathematical” (304 occurrences). As shown in Figure 4.10, *model* is used as a countable noun and refers to an umbrella for various systems or theoretical frameworks applied to Logistics research. This matches with the sense of “A model of a system or process is a theoretical description that can help you understand how the system or process works, or how it might work” (COBUILD, 2023, p.954).

The screenshot shows a concordance tool interface for the Logistics Academic Written English Corpus. The search term is 'mathematical model'. The interface displays search filters: 'word model' with 14,790 occurrences (2.318.4 per million tokens, 0.23%), 'filter mathematical -5, 5' with 304 occurrences (47.65 per million tokens, 0.0048%), and 'Sample 10' with 10 occurrences (1.57 per million tokens, 0.00016%). The concordance lines are displayed in a table with columns for 'Left context', 'KWIC', and 'Right context'. The table shows 10 lines of text, each with a checkbox and a magnifying glass icon. The KWIC column highlights the search term 'mathematical model' in red. The right context column shows the text following the search term, often starting with 'M for finding the optimal carton set is as follows: In model M, the ot'.

Figure 4.10 Examples of concordance lines for collocation “mathematical model”

### 4.3.5 Pilot study on word senses identification

A pilot study was conducted for identifying word senses to test the feasibility and practicality of my research design. All criteria and steps mentioned above applied in order. The pilot study also has an aim to define the threshold of LogDice value through several rounds of attempts on different cut-off points.

Table 4.8 presents the top 10 specialised items that were selected as node words. They were all searched through the function Concordance>collocations on the Sketch Engine, with LogDice and +/- 5 set as criteria. Table 4.8 lists the number of collocates of each node word has by applying different cut-off points of LogDice value. For example, if the LogDice was set as 10 or above, *results* has 6 collocates. However, if the LogDice cut-off point gets lower, the number of collocates increases to 21 ( $\text{LogDice} \geq 9$ ), 82 ( $\text{LogDice} \geq 8$ ), and 1,719 ( $\text{LogDice} \geq 4$ ).

Table 4.8 The number of collocations generated with different LogDice thresholds

Items	LogDice Thresholds and Number of Collocations Generated						
	$\geq 10$	$\geq 9$	$\geq 8$	$\geq 7$	$\geq 6$	$\geq 5$	$\geq 4$
results	6	21	82	206	483	960	1,719
abstract	2	8	13	44	168	310	385
study	5	26	84	233	521	971	1,720
introduction	4	7	15	59	219	372	452
based	1	5	41	169	521	1,116	1,987
literature	3	19	78	204	468	989	1,596
section	12	33	91	181	376	717	1,315
paper	9	18	58	152	343	806	1,485
table	14	31	66	160	349	771	1,539
model	4	33	93	256	567	1,121	1,930

For the threshold of LogDice value, different cut-off points have been used in previous studies:  $\text{LogDice} \geq 7$  (Pinto et al., 2023),  $\text{LogDice} \geq 5$  (Rychlý, 2008) and  $\text{LogDice} \geq 4$  (Cao and Deignan, 2019; Cao, Pho and Chi, 2022). If these suggested thresholds were adopted, the total number of collocates for 10 items in total would be very large, ranging from 44 to 1,987. In contrast, if  $\text{LogDice} \geq 10$  was used, it would generate less than 15 collocations for 10 node words, which may not be sufficient to provide much information in polysemous word sense identification. For the threshold  $\text{LogDice} \geq 8$ , the number of collocations generated is not balanced across all 10 nodes when compared to  $\text{LogDice} \geq 9$ . Such imbalance may have influence on concordance lines reading for senses identification, for example, much collocation evidence for word sense analysis on some items and insufficient collocational and contextual clues for other items. Therefore, on the aspects of practically and sufficient data generated for word sense analysis,  $\text{LogDice} \geq 9$  as the cut-off point may be more suitable in my research.

For word sense analysis, one of the items in Table 4.7 *paper* was intentionally selected as the node word in search. By applying  $\text{LogDice} \geq 9$  and +/-5 span for collocation analysis, the collocate list of *paper* was generated, where 18 collocates meet the criteria, as Figure 4.11 shows.

	Word	Cooccurrences <sup>?</sup>	Candidates <sup>?</sup>	LogDice ↓		Word	Cooccurrences <sup>?</sup>	Candidates <sup>?</sup>	LogDice ↓
1	<input type="checkbox"/> this	2,147	22,097	11.37 ***	11	<input type="checkbox"/> organized	133	262	9.86 ***
2	<input type="checkbox"/> Research	423	1,000	11.32 ***	12	<input type="checkbox"/> In	597	17,790	9.78 ***
3	<input type="checkbox"/> type	464	2,121	11.18 ***	13	<input type="checkbox"/> The	963	36,587	9.59 ***
4	<input type="checkbox"/> Introduction	321	587	11.04 ***	14	<input type="checkbox"/> organised	103	181	9.52 ***
5	<input type="checkbox"/> follows	367	2,714	10.72 ***	15	<input type="checkbox"/> rest	106	349	9.51 ***
6	<input type="checkbox"/> Paper	215	259	10.56 ***	16	<input type="checkbox"/> we	447	17,646	9.38 ***
7	<input type="checkbox"/> This	696	10,928	10.54 ***	17	<input type="checkbox"/> contributions	85	752	9.08 ***
8	<input type="checkbox"/> remainder	168	257	10.20 ***	18	<input type="checkbox"/> concludes	71	144	9.00 ***
9	<input type="checkbox"/> purpose	181	842	10.14 ***	19	<input type="checkbox"/> -	197	8,552	8.96 ***
10	<input type="checkbox"/> paper	245	4,153	9.87 ***	20	<input type="checkbox"/> structured	72	361	8.95 ***

Figure 4.11 Collocate list of *paper* in +/-5 span and with LogDice  $\geq 9$

Concordances of collocate “this” having more than 1,000 occurrences were randomly sampled for concordance reading. Four rounds of random sample selection were conducted on it, that is, 100, 200, 300 and 400 concordances in random sample selection respectively. Thus, 1,000 concordances of “this” were checked rather than 2,147. Also, all concordances of the rest 17 collocates were checked thoroughly. Through concordance reading and dictionary referencing, two senses were identified: *paper* as “a kind of material” and *paper* as “academic research”. Here below is a concordance extract for each sense identified respectively:

*paper* as a kind of material:

..., concluding that technologies have their place in onboarding (e.g., when it comes to turning a *paper* checklist into an automated workflow), ... (BL200405, emphasis added)

*paper* as academic research:

The Wasserstein ambiguity set is good for dealing with the scarcity of data. Thus, this *paper* introduces the DRO with Wasserstein ambiguity set to the joint optimum of robust supply chain network design, ... (TR220708, emphasis added)

Once the concordance lines were checked, the occurrence of each sense in citations was calculated. In this way, how many senses and what sense is typically frequently used in the LAWEC can be noted. This provides data for further word sense comparison with senses in general academic and general contexts later (see section 4.4). A simple frequency-based calculation on percentage of each sense in occurrence was taken, adapted from West (1953). The formula is presented below:

$$\text{Occurrence of Sense A in citations} = \frac{\text{Number of concordances with Sense A identified}}{\text{Number of all concordances checked}} \times 100\%$$

For *paper*, in total 6,585 concordances were checked, where only 8 occurrences exist with the sense of *paper* as “a kind of material” and all the rest have the sense of *paper* as “academic research”. Therefore, the occurrence probability of the sense of *paper* is  $8/6,585 \times 100\%$ . The occurrence probability of two

senses of the polysemous word *paper* in the LAWEC are presented in Table 4.9 below. Word senses for other specialised Logistics word will be analysed in the same way (see section 5.3 and the word sense list in Appendix B).

Table 4.9 Probabilities of two senses of paper occurring in the LAWEC

Item in search	Sense	Probability of Occurrence in the LAWEC
<i>paper</i>	A kind of material	0.12%
	Academic research	99.88%

#### 4.4 Comparison of word senses between contexts

The third research aim of my study is to compare the senses of specialised vocabulary items in the Logistics academic written context and those in general academic context and general use context respectively. It is to see the difference of senses of specialized Logistics words used in different specific contexts. Having word sense comparisons between contexts can show what senses of a specific word type has in a certain context and what sense is typically frequently used in that context. Previous empirical studies like Deignan, Semino and Paul (2019) and Mićović and Beko (2022) have shown that learners may not be proficiently able to recognize and distinguish daily senses and professional senses in a specific context. Word sense comparisons between contexts may be helpful for learners in bridging the “gap” of senses when they step into more specialized contexts. To realize the word sense comparison, word sense analysis will be conducted in general academic context and general use context respectively, which will have steps discussed later:

- (1) Choose a corpus representing general academic and general use contexts respectively;
- (2) Set the items in the LoSAWEL as node word in search;
- (3) Collocational analysis on LoSAWEL items in the general academic and general use corpus respectively;
- (4) Identify senses of LoSAWEL items in the general academic and general use corpus respectively, through concordance line reading and dictionary checking;
- (5) Compare word senses in the LAWEC corpus and those in the general academic corpus;
- (6) Compare word senses in the LAWEC corpus and those in the general use corpus.

Section 4.4.1 and 4.4.2 will describe the first step in detail. Section 4.4.3 will focus on steps (2) – (4). Section 4.4.4 is on the last two steps for word sense comparisons.

#### 4.4.1 Choosing corpora for general academic contexts

The British Academic Written English Corpus (BAWE) was selected for general academic use context for three reasons. First, this corpus shows its representativeness of disciplines in general academic written contexts. It focuses on the written academic contexts and collects more than 20 disciplines covering four main research areas (Arts and Humanities, Life Sciences, Physical Sciences and Social Sciences) (Heuboeck, Holmes and Nesi, 2010). This fits my desire in choosing a written-material dominant corpus and its wide coverage of disciplines suits the characteristics of general academic context. Second, this corpus is fairly balanced in data collection. The BAWE generally has an evenly distribution in data of four research areas (Heuboeck, Holmes and Nesi, 2010). This indicates that the senses of specialised items identified in this corpus may not confront much imbalanced-data-based influence. Third, this corpus is also available on the Sketch Engine, with accessibility to its data online. However, it should be acknowledged that the data collected by this corpus are mostly essays written by undergraduates or postgraduates in the UK with the merit grade or above. This suggests that the BAWE is less representative on other general academic written registers such as textbooks, research journal articles or research proposals. The authorship does not fit well with that of my corpus – the LAWEC where the texts are mostly by established scholars or the researchers who already have obtained the PhD degree. The discrepancy of authorship may have its influence on the word use and its senses, as empirical studies such as Coxhead et al. (2020) show that there exists much difference on vocabulary in use between learners and experts. Moreover, the discrepancy also exists in registers. The texts in the LAWEC are academic journal articles while those in the BAWE are mostly essays or other registers frequently used in university. Registers may also be a factor affecting word in use (de Chazal, 2014), and possibly, on word sense. Another factor is the quality of texts: academic journal articles being peer-reviewed and thus with wider acceptance in academics, while the students' writing may contain misuse of language. Nevertheless, although there exist limitations for choosing the BAWE, these limitations cannot overweight the benefits it brings to my research, considering it being one of the most popular, representative and available general academic written corpora.

Recognizing the limitations of the BAWE, a measure was taken to reduce the bias of the BAWE texts towards the results of word sense comparison between a specialized context and the general academic context. Originally, I planned to use COCA academic subcorpus to cross check the word sense comparison between the specialized context and the general academic context. COCA (Davies, 2008) is one of the largest corpora of the world, and its academic subcorpus contains more than 120 million tokens. However, due to non-availability of the license, this measure cannot be realized as planned. Therefore, an alternative measure was planned based on real situation - the written academic part of BNC1994 was extracted through the Sketch Engine for a double check on a group of randomly selected LoSAWEL items in comparison. A written academic subcorpus of the BNC was compiled on the Sketch

Engine. It has similar procedures of the compilation of the BNC non-academic writing subcorpus mentioned above, the difference being the options of text selected. This subcorpus contains 17,625,626 tokens, covering a wide range of disciplines (e.g., Humanities, Arts, Medicine, Engineering and Technology). A group of items in the LoSAWEL was randomly selected (10%) through the random selection website (<https://www.random.org/>) (Haahr, 1998) for word sense identification and comparison.

#### 4.4.2 Choosing corpora for general use contexts

The British National Corpus (BNC1994) was chosen to represent general use context. It is because its written part has its dominance in this corpus (about 90%) (Burnard, 2007), which generally fits my research aims in terms of discourse. Moreover, this corpus includes a wide range of domain of written materials, ranging from natural and pure sciences in informative texts to leisure in imaginative texts in different proportions (Burnard, 2007), which is generally considered as a good representation for written language use in general contexts (Friginal, 2018). Additionally, its size (100 million tokens) indicates that it would be large enough to allow some specialised items which are less frequently used in general contexts to have adequate occurrences and sufficient concordance lines for word senses identification. This corpus is also available on the Sketch Engine and brings convenience to access its data. One of the limitations is that the materials collected in this corpus were mostly published during the period 1975-1994 (Burnard, 2007), and it is not a monitoring corpus which updates or has further collection for its data regularly. Thus, it may not fully reflect the recency of word senses in use for the 21<sup>st</sup> century, since the word senses and use may change from time to time (Stubbs, 2001; Hunston, 2022a). However, considering the BNC2014 is not available on the Sketch Engine by the time of writing, BNC1994 can be a suitable corpus representing general use contexts, based on availability and practicality. Another potential bias of the BNC1994 is that it also contains written academic texts. These texts may have influence on my word sense comparison. Since the register of the LAWEC, the self-built specialized corpus, is written academic journal article, it has similarities to that part of the BNC1994. This, thus, may overestimate the similarity of word sense in two corpora. To reduce this bias, not all BNC1994 was used. Instead, the academic part of BNC1994 was extracted and excluded through the Sketch Engine. The Sketch Engine has a function of building subcorpus of the BNC1994, based on classification of text register. Specifically, it is through Dashboard > Manage Corpus > Subcorpora > Create subcorpus > David Lee's classification. In this classification, written academic texts are marked, for example, "W\_ac\_medicine". All options except those marked with "W\_ac\_" were selected to create a BNC non-academic subcorpus (hereafter BNC1994 NAW) on the Sketch Engine. This subcorpus contains 94,473,730 tokens, accounting for 84.1% of the original BNC1994. This size is still large enough to represent a general use corpus for my word sense comparison.

#### 4.4.3 Word sense identification in the BAWE and the BNC1994 NAW

After the selection of corpora representing general academic and general use contexts respectively, word sense identification will be conducted in each corpus. The procedures are almost the same as those presented in section 4.3. The main difference is on the contexts and corpora. Specifically, the specialised items generated and used for searching word senses in the LAWEC are set as node words again in the BAWE and the BNC1994 respectively. All node words will be analysed through collocational checking with the same span (+/-5) and indicator of collocation (LogDice), concordance lines reading (reading all concordances whose number is below 1,000 and random samples on those beyond 1,000), and dictionary checking.

However, the threshold of LogDice in collocational analysis will be different. Here the top 10 specialised items presented in section 4.3 are selected as examples again. The Table 4.10 and Table 4.11 present the LogDice values of each of 10 specialised items and the number of collocates generated in the BAWE and the BNC1994 respectively.

Table 4.10 The amount of collocates generated with different LogDice threshold in the BAWE

Items	LogDice Thresholds and Amount of Collocates Generated						
	$\geq 10$	$\geq 9$	$\geq 8$	$\geq 7$	$\geq 6$	$\geq 5$	$\geq 4$
results	2	5	33	154	419	959	1,000
abstract	1	3	11	39	78	122	173
study	n/a	3	18	94	352	972	1,000
introduction	2	4	21	6	99	321	1,000
based	2	4	19	68	350	953	1,000
literature	1	3	12	61	254	496	583
section	n/a	4	27	92	310	878	1,000
paper	1	4	12	78	260	520	613
table	7	15	35	96	307	876	1,000
model	n/a	3	20	115	316	1,000	n/a

Table 4.11 The amount of collocates generated with different LogDice threshold in the BNC1994 NAW

Items	LogDice Thresholds and Amount of Collocates Generated						
	$\geq 10$	$\geq 9$	$\geq 8$	$\geq 7$	$\geq 6$	$\geq 5$	$\geq 4$
results	n/a	a/a	9	38	506	1,000	n/a
abstract	n/a	1	4	11	63	246	448
study	n/a	n/a	3	40	183	658	1,000
introduction	n/a	n/a	n/a	6	50	239	1,000
based	n/a	2	2	22	132	571	1,000
literature	n/a	n/a	4	17	75	258	949
section	n/a	4	17	44	146	504	1,000
paper	1	3	9	33	105	448	1,000
table	n/a	2	13	52	162	493	1,000
model	n/a	n/a	2	15	92	453	1,000

It can be seen that the number of collocates for different LogDice thresholds ranging from 4 to 10 in two corpora are quite imbalanced, when compared to those presented in the Table 4.7 in section 4.3. This means that setting a definite value for LogDice threshold in collocational analysis as what was done in the LAWEC corpus (LogDice  $\geq 9$ ) across these two corpora (BAWE and BNC1994 NAW) may not be as feasible as being expected. Therefore, the top seven collocates were selected for each LoSAWEL items on collocational analysis in the BAWE and the BNC1994 NAW. This threshold was set based on the reason that the LoSAWEL items each have seven collocates on average in the LAWEC corpus. Focusing on the top seven collocates of the LoSAWEL items in the BAWE and the BNC1994 NAW respectively is a way to generally balance the number of collocates of the LoSAWEL items across three corpora in my research.

Another difference is the dictionaries for dictionary checking. As discussed in the section 4.3, the latest version of the COBUILD dictionary (Collins, 2023) and the discipline-specific glossary and dictionary of Logistics – IWAL (IWAL, 2018) were selected for word sense analysis in the LAWEC. Considering that the BAWE and the BNC1994 have less relation to the Logistics disciplinary concept, there is no need to use IWAL when analyzing senses in these two corpora. The COBUILD dictionary was kept because of its scope in general use English. For the word sense analysis in the BAWE, disciplinary dictionaries in the *Oxford Quick Reference* series were chosen as references to distinguish disciplinary senses, together with the COBUILD dictionary. This is because this series of disciplinary dictionaries cover a wide range of disciplines and are best-sellers consisting of hundreds of thousands

of concise entries (<https://www.oxfordreference.com/>). Before referencing *the Oxford Quick Reference* dictionary series, two criteria should be met in word sense analysis:

- (1) Through collocational analysis and concordance line reading, all concordances of a certain collocate in the BAWE were categorized in the same discipline **only**, when in word sense analysis;
- (2) The COBUILD may (not) provide a clear definition on such word sense.

For example, “judicial” is one of the top seven collocates of the LoSAWEL item *review* in the BAWE, and all its concordance were categorized in the Law discipline only (see Figure 4.12 and 4.13). The COBUILD dictionary marks the entry of “judicial review” as a multi-word unit in Law, and this unit is also included as an entry in *the Oxford Quick Reference – A Dictionary of Law* (Law, 2022). Therefore, “judicial review” is annotated as a register-specific collocation of *review* in the Law field.

The screenshot displays the 'CONCORDANCE' interface for the British Academic Written English Corpus (BAWE). The search term is 'judicial review', with 129 results. The interface includes a search bar, filters for 'word review' (2,157) and 'filter judicial -5, 5' (129), and a list of concordance lines. The concordance lines show the phrase 'judicial review' in red text within various sentences, such as 'Before considering the substantive arguments for judicial review, there are a number of procedural issues to address.' and 'The Exclusivity Principle [1] requires that claims for judicial review be brought by judicial review procedure [2] rather than ordinary civil procedure.'

Figure. 4.12 Extract concordances of “judicial review” in the BAWE

**Display and count metadata**

Select the metadata to be displayed in the concordance. Click to calculate statistics.

Display above lines?  Shorten to 15 characters

Before considering the substantive arguments for **judicial review**, there are a number of procedural issues to address.

type to search (1)

<input type="checkbox"/> Token number	831877	
<input type="checkbox"/> Document number	258	
<input type="checkbox"/> Course title	Law	
<input type="checkbox"/> Date of writing	2005-03	
<input type="checkbox"/> Discipline	Law	
<input type="checkbox"/> Discipline group (area)	CC	

CLOSE **SAVE**

Figure 4.13 The metadata of an extract of concordance of “judicial review” in the BAWE

#### 4.4.4 Pilot study on cross-corpus comparison of word senses

This section reports a pilot study on examples of word senses across three corpora: the LAWEC for Logistics academic written context, the BAWE for general academic context, and the BNC1994 NAW for general use context. The specialised item *model* is selected intentionally as an example for this pilot study.

The pilot study consists of two stages: identifying word senses of the example word *model* in the LAWEC, the BAWE and the BNC1994 NAW as the first one, and comparing senses across three corpora in different contexts as the second. For the first stage, the example word *model* is set as the node word, being searched and analysed in three corpora respectively (see the examined collocates in the three corpora in Figure 4.14, 4.15 and 4.16). The main procedures are generally the same as presented above: collocational checking, concordance lines reading (with random sample selection if necessary) and dictionary checking.

	Word	Cooccurrences <sup>?</sup>	Candidates <sup>?</sup>	LogDice ↓		Word	Cooccurrences <sup>?</sup>	Candidates <sup>?</sup>	LogDice ↓		
1	<input type="checkbox"/> proposed	900	4,238	10.60	...	26	<input type="checkbox"/> We	388	7,857	9.13	...
2	<input type="checkbox"/> The	1,851	36,587	10.21	...	27	<input type="checkbox"/> used	374	7,284	9.12	...
3	<input type="checkbox"/> our	747	7,191	10.12	...	28	<input type="checkbox"/> structural	264	931	9.10	...
4	<input type="checkbox"/> a	3,470	88,175	10.11	...	29	<input type="checkbox"/> fit	263	1,039	9.09	...
5	<input type="checkbox"/> developed	543	2,757	9.99	...	30	<input type="checkbox"/> by	724	29,218	9.07	...
6	<input type="checkbox"/> model	885	14,109	9.97	...	31	<input type="checkbox"/> programming	254	987	9.04	...
7	<input type="checkbox"/> an	964	18,933	9.87	...	32	<input type="checkbox"/> can	557	20,551	9.01	...
8	<input type="checkbox"/> for	1,733	49,307	9.79	...	33	<input type="checkbox"/> as	824	37,534	9.01	...
9	<input type="checkbox"/> the	9,108	327,929	9.77	...	34	<input type="checkbox"/> 3	336	7,335	8.96	...
10	<input type="checkbox"/> inventory	516	4,996	9.74	...	35	<input type="checkbox"/> 1	550	21,654	8.95	...
11	<input type="checkbox"/> with	1,253	34,629	9.70	...	36	<input type="checkbox"/> based	300	5,418	8.93	...
12	<input type="checkbox"/> .	6,059	226,760	9.68	...	37	<input type="checkbox"/> and	2,932	183,851	8.92	...
13	<input type="checkbox"/> in	2,762	98,613	9.64	...	38	<input type="checkbox"/> under	277	4,178	8.90	...
14	<input type="checkbox"/> is	2,110	74,604	9.60	...	39	<input type="checkbox"/> ,	5,049	339,230	8.87	...
15	<input type="checkbox"/> we	752	17,646	9.57	...	40	<input type="checkbox"/> presented	239	2,072	8.86	...
16	<input type="checkbox"/> this	833	22,097	9.53	...	41	<input type="checkbox"/> be	609	28,597	8.85	...
17	<input type="checkbox"/> In	723	17,790	9.51	...	42	<input type="checkbox"/> (	1,984	127,674	8.83	...
18	<input type="checkbox"/> to	2,930	123,187	9.44	...	43	<input type="checkbox"/> 2	376	12,295	8.83	...
19	<input type="checkbox"/> was	504	10,241	9.37	...	44	<input type="checkbox"/> measurement	214	917	8.80	...
20	<input type="checkbox"/> mathematical	304	692	9.33	...	45	<input type="checkbox"/> which	431	16,989	8.80	...
21	<input type="checkbox"/> that	1,285	52,619	9.29	...	46	<input type="checkbox"/> stochastic	215	1,079	8.79	...
22	<input type="checkbox"/> results	426	7,936	9.26	...	47	<input type="checkbox"/> conceptual	207	727	8.77	...
23	<input type="checkbox"/> using	374	5,653	9.23	...	48	<input type="checkbox"/> :	387	14,546	8.76	...
24	<input type="checkbox"/> of	3,452	175,596	9.21	...	49	<input type="checkbox"/> on	659	36,070	8.73	...
25	<input type="checkbox"/> )	2,478	127,765	9.15	...	50	<input type="checkbox"/> This	319	10,928	8.67	...

Figure 4.14 Examined collocates of *model* in the LAWEC

	Word	Cooccurrences <sup>?</sup>	Candidates <sup>?</sup>	LogDice ↓
1	<input type="checkbox"/> model	230	4,048	9.79 ...
2	<input type="checkbox"/> Bohr	117	236	9.67 ...
3	<input type="checkbox"/> atom	78	406	9.03 ...
4	<input type="checkbox"/> Figure	101	3,829	8.64 ...
5	<input type="checkbox"/> based	91	3,103	8.62 ...
6	<input type="checkbox"/> The	762	62,638	8.54 ...
7	<input type="checkbox"/> Model	54	410	8.50 ...

Figure 4.15 Examined collocates of *model* in the BAWE

	Word	Cooccurrences <sup>?</sup>	Candidates <sup>?</sup>	LogDice ↓
1	<input type="checkbox"/> Serial	112	248	8.62 ...
2	<input type="checkbox"/> model	202	7,981	8.60 ...
3	<input type="checkbox"/> Printer	87	344	8.24 ...
4	<input type="checkbox"/> Computer	93	2,633	8.02 ...
5	<input type="checkbox"/> Asset	73	254	8.00 ...
6	<input type="checkbox"/> Location	71	268	7.96 ...
7	<input type="checkbox"/> IBM	78	3,976	7.61 ...

Figure 4.16 Examined collocates of *model* in the BNC1994 NAW

Through collocational analysis, concordance lines reading and dictionary checking, the senses of *model* in three corpora is presented in the Table 4.12 below. Two senses of *model* are identified in the three corpora, suggesting that *model* is a polysemous word in the LoSAWEL, especially in the LAWEC context.

The second stage of the pilot study is to have a comparison of senses across three corpora. For *model* in the LAWEC, it can be a noun for “theoretical framework”, or as a verb for “an action to make a model”. The occurrence in citations of senses of *model* in the LAWEC are also calculated, with *model* as “a theoretical framework” for the dominant sense (98.81%) and *model* as “an action to make a model” (1.19%). However, *model* is a monosemous word in both the BAWE and the BNC1994 NAW context, only having the sense of “a theoretical framework”.

In comparison, the sense of “make a model” only occurred in the LAWEC. In later LAWEC-BAWE and LAWEC-BNC1994 NAW comparisons, word senses like this were named as LAWEC unique senses, representing senses that were only found in the LAWEC. The sense of “a theoretical framework” occurred in three corpora, and the percentage of occurrences did not have big differences. Word senses like this were named as LAWEC-BAWE shared senses (S) or LAWEC-BNC1994 NAW shared senses (S) (“S” means the sense of percentages of occurrence in citations are similar in two corpora in comparison). However, if a word sense was identified in both the LAWEC and BAWE, or in the LAWEC and the BNC1994 NAW, and the difference between their percentages in citations is a larger than 1.5, this sense is considered as a LAWEC-BAWE shared sense (D) or LAWEC-BNC 1994 NAW shared sense (D). I decided the 1.5-time threshold by checking data of all senses, by comparing with those of 2-time threshold and 2.5-time threshold, ensuring discipline-specific senses can be extracted.

Table 4.12 Word senses of *model* in three corpora (emphasis added)

	LAWEC (%)	Example Concordance in the LAWEC	BAWE (%)	Example Concordance in the BAWE	BNC1994 (%)	Example Concordance in the BNC1994
Word Senses	a theoretical framework ( <i>n.</i> ) (98.81%)	Second is that a Poisson <b>model</b> exhibited poor fit to our data,... (BL180304)	a theoretical framework ( <i>n.</i> ) (100%)	In this <b>model</b> , the metropolis may be seen as the expression, or content, of the various forms-of-sociation ... (13)	a theoretical framework ( <i>n.</i> ) (100%)	Copeland (1976) developed a simple <b>model</b> of the effects of the arrival of a single piece of information on price and volume. (3980)
	make a model ( <i>v.</i> ) (1.19%)	We <b>model</b> a retailer ordering perfect substitutes (Zinn and Liu 2001; McKinnon et al. 2007) procured from two suppliers... (BL180103)	N/A	N/A	N/A	N/A

## 4.5 Summary

In this chapter, data collection and data analysis methods for each research questions are discussed. Section 4.1 is on data collection for this research, where a Logistics academic written corpus the LAWEC was compiled. The principles of compiling a corpus (representativeness, balance and size) and the practical design of the LAWEC based on those principles were discussed. The LAWEC was also evaluated based on three principles. In section 4.2, criteria on Logistics specialized academic written English word list (LoSAWEL) development were presented. They are based on keyword analysis, dispersion, range, and lexical words in quantitative selection via #LancsBox 6.0., addressing data analysis for RQ1. Methods for RQ2 and RQ3 were discussed in section 4.3, where word sense identification in the LAWEC through the Sketch Engine was focused. Steps of word sense identification (collocational analysis, concordance line reading and dictionary checking) were discussed one by one, with examples, screenshots and extract of concordances provided. Specifically,  $\text{LogDice} \geq 9$  was set as the cut-off point for the collocational analysis. All concordances of collocates whose number is below 1,000 were read, and for those beyond 1,000, four rounds of random samples were conducted. COBUILD dictionary and the IWLA, a Logistics discipline-specific glossary and dictionary, were used as key references. Calculation method on word senses was also presented with examples. For RQ2 on word sense comparison between the LAWEC and general academic context, and between the LAWEC and general use context, they are presented in section 4.4. All steps of analysis were generally similar to those in the section 4.3. The BAWE and the BNC1994 NAW were selected to represent general academic and general use context respectively. A slight change was applied to the cut-off point of LogDice and dictionary checking, based on the nature of the BAWE and the BNC1994 NAW. A pilot study was conducted for each research questions respectively.

## Chapter 5 Results

### 5.0 Introduction

This chapter presents results of my study, with the attempt to answer my research questions. Section 5.1 addresses RQ1 and reports the results of word list development, lexical coverage of the word list in corpus and the comparison with other corpus-derived word lists. Section 5.2 focuses on RQ2 in identifying sense(s) of words in three different corpora: the LAWEC, the BAWE and the BNC1994 NAW. It also addresses the word sense comparison between the LAWEC and the BAWE, and that between the LAWEC and the BNC1994 NAW respectively. Section 5.3 focuses on RQ3, discussing polysemy of the Logistics specialized vocabulary list in the LAWEC.

### 5.1 Development of the Logistics Academic Written English Word List (LoSAWEL)

#### 5.1.1 The specialised words in Logistics academic written English

Following the procedures mentioned in section 4.1, the criteria of Keywords (Cohen's  $d \geq 0.2$ ), Dispersion ( $DP_{\text{norm}} \leq 0.95$ ) and Range (Range  $\geq 360$ ) were applied in sequence, and these finally led to 382 items left (see Table 5.1). The Lexical word criterion was set as a filter (see Table 5.2), with the reference of the Function Word List (Appendix A) for the exclusion of function words. This step resulted 45 function words identified. Moreover, other 22 non-lexical items which do not belong to either function words or foreign words, acronyms, abbreviations or discipline-related proper nouns were also identified by my scanning the items. These non-lexical items, for example, were mostly words of years such as 2002, 2014, 2019. This could be the reason that all in-text citations were kept when cleaning the texts, and these words of years were frequent in in-text citations. A quick check on concordance lines proves this. However, since all these non-lexical items do not convey any linguistic or disciplinary meaning, they were excluded and no need to be in an individual list. Then, I checked the remaining 315 items manually ( $382-45-22 = 315$ ) to extract foreign words, acronyms, abbreviations, discipline-related proper names and developed an individual word list for them, separated from the LoSAWEL. Interestingly, there were no acronyms and discipline-related proper nouns were extracted in the rest of words, since none of them met previous statistical selection criteria. Only two abbreviations were found (*i.e.* and *e.g.*) and they are both frequently used in academic contexts, which is in accordance with the genre of texts collected in my corpus - the academic journal articles. Besides, two items (*et* and *al*) were identified as foreign words in type. However, they are actually considered to be meaningful when they appear as *et al.*, and thus I listed this meaningful foreign word together with two general academic

abbreviations in an individual list. Finally, a list containing 311 word types was generated ( $315 - 2 - 2 = 311$ ) as the LoSAWEL for the specialised words in Logistics academic written English.

Table 5.1 Keywords, Dispersion, and Range criteria applied for word selection

Criteria	Word Type Left
Keywords (Cohen's $d \geq 0.2$ )	2802
Dispersion ( $DP_{norm} \leq 0.95$ )	2361
Range (Range $\geq 360$ )	382

Table 5.2 Criteria of lexical words applied for word selection

Excluded Words	Amount	Examples	Methods
Function Words	45	the, thus, of	Function Word List as a reference
Non-lexical Words	22	2002, 2014, 2019	
Foreign Words in type	2	et, al	
Abbreviations	2	i.e., e.g.	Checking manually
Acronyms	0	n/a	
Discipline-related Proper Nouns	0	n/a	

There was no specific cutting-off point on frequency to select specialised words in the procedures. However, given that the criteria of keyword analysis, dispersion and range all have relations with word frequency in calculation (Brezina, 2018), applying these three criteria can ensure the words selected do not have a relatively low frequency in the corpus. The frequency of all selected words was beyond 500 and its average was 2,826, with only around 9% items ( $n = 28$ ) being less than 1,000. The word with the least frequency is *conclusion* (frequency = 559; range = 379;  $DP_{norm} = 0.465$ ); while the highest one is *supply* (frequency = 15770; range = 513;  $DP_{norm} = 0.579$ ). All words had the value of Range ranging from 360 to 706, *introduction* being the highest one. For dispersion ( $DP_{norm}$ ), *abstract* (frequency = 768; range = 679;  $DP_{norm} = 0.186$ ) is considered the most evenly distributed word across texts; *discussion* (frequency = 1,012; range = 465;  $DP_{norm} = 0.867$ ) being the most unevenly distributed one. The statistics will be discussed in detail later. Table 5.3, 5.4 and 5.5 list the top 10 specialised words of the word list based on the ranking of frequency, range and dispersion (evenly distributed words) respectively.

Table 5.3 The top 10 frequent specialised words in the LoSAWEL

Ranking.	Item	Frequency	Range	Dispersion ( $DP_{norm}$ )
1	supply	15,770	513	0.58
2	model	14,747	664	0.40
3	research	13,112	687	0.39
4	chain	12,233	495	0.59
5	study	10,867	689	0.33
6	cost	10,427	570	0.43
7	logistics	9,314	405	0.69
8	data	8,827	630	0.43
9	product	8,535	498	0.61
10	results	8,485	698	0.29

Table 5.4 The top 10 specialised words in the LoSAWEL based on Range

Ranking	Item	Range	Frequency	Dispersion ( $DP_{norm}$ )
1	introduction	706	981	0.23
2	used	704	7,285	0.30
3	based	700	6,628	0.29
4	results	698	8,485	0.29
5	following	692	4,438	0.30
6	study	689	10,867	0.33
7	using	689	6,190	0.32
8	research	687	13,112	0.39
9	analysis	680	6,531	0.34
10	abstract	679	768	0.19

Table 5.5 The top 10 evenly distributed specialised words in the LoSAWEL

Ranking	Item	Dispersion (DP <sub>norm</sub> )	Frequency	Range
1	abstract	0.19	768	679
2	introduction	0.23	981	706
3	based	0.29	6,628	700
4	results	0.29	8,485	698
5	following	0.30	4,438	692
6	used	0.30	7,285	704
7	different	0.32	7,922	654
8	provide	0.32	2,936	642
9	further	0.32	3,794	659
10	using	0.32	6,190	689

Two lexical features are shown: it consists of both general academic words and discipline-specific words. The academic feature was demonstrated through general academic words. For example, *results* and *study*, which also appeared in the AVL, are among the top 10 frequent, wide in range and evenly distributed words in the Logistics word lists (appearing in all three tables). The disciplinary feature was shown by discipline-specific words. For example, *supply*, *chain* and *logistics*, in Table 5.3 among the top 10 frequent words, are specifically used in the discipline Logistics with high frequency. The specific analysis on the statistics is presented in the next sections.

### 5.1.2 Statistical analysis on frequency, range and dispersion on LoSAWEL words

#### *Frequency of the LoSAWEL words*

Table 5.6 presents an overview of the frequency of the LoSAWEL words. It contains levels of word frequency, the amount of word types and its percentage at each level. All word types had a frequency higher than 500 (559 as the minimum). Although the frequency minimum point of my list is 500 times in the whole corpus, as shown in Table 5.6, 90.7% of the items in the list have the frequency of at least 1000. More than half (52.3%) of the items occurred more than 2,000 times in the corpus.

Table 5.6 Frequency of the LoSAWEL words

Frequency in the LAWEC	Number of Types	Percentage
500-1,000	28	9.3%
1,001-2,000	117	38.4%
2,001-3,000	69	22.5%
3,001-4,000	33	10.3%
4,001-5,000	20	6.3%
5,001-6,000	12	3.6%
6,001-7,000	13	4%
7,001-8,000	9	2.3%
8,001-9,000	3	1%
9,001-10,000	1	0.3%
>10,001	6	2%
Total	311	100%

#### *Range of the LoSAWEL words*

For the criterion of Range, it was set that all words should appear in at least 360 of 720 texts or more. The table 5.7 shows that a large number of words occurred in a wide range of texts. About 37.8% of items ( $n = 118$ ) appeared in 401-500 texts, which occupied the largest proportion. Also, more than 75.2% of words in the list ( $n = 231$ ) were presented in at least 400 texts. Apart from these, all 311 words in the list were in all nine journals of Logistics, with a manual checking on these items.

Table 5.7 Range of the LoSAWEL words

Range in the LAWEC	Number of Types	Percentage
360-400	80	24.8%
401-500	118	37.8%
501-600	69	22.8%
601-700	41	13.6%
> 701	3	1%
Total	311	100%

#### *Dispersion of the LoSAWEL words*

The criterion of dispersion set that each item should have the  $DP_{norm}$  lower than 0.95 for extracting words that are more evenly distributed across texts in the corpus. Table 5.8 presents that, although the  $DP_{norm} \cong 0.95$  was set as the cutting-off point, all words had a dispersion lower than 0.90. Most of

words (about 75%;  $n = 235$ ) had a dispersion between 0.4 and 0.6. Moreover, the statistics of dispersion of more than 90% of words ( $n = 288$ ) was smaller than 0.6.

Table 5.8 Dispersion of the LoSAWEL words

Dispersion ( $DP_{norm}$ ) in the LAWEC	Number of Types	Percentage
0.81-0.90	2	0.7%
0.71-0.80	2	0.7%
0.61-0.70	19	6.3%
0.51-0.60	116	37.1%
0.41-0.50	119	38.7%
0.31-0.40	48	14.9%
0.21-0.30	4	1.3%
$\leq 0.20$	1	0.3%
Total	311	100%

### 5.1.3 Lexical coverage of the LoSAWEL

It is necessary to have an analysis of the lexical coverage of the items in the generated word list. To calculate the lexical coverage of the word list in the corpus, it is needed to firstly find out the frequency of all items in the word list, and then to have this statistics divide the total tokens of the corpus. The formula is as follows:

$$\text{Lexical Coverage} = \text{Total Frequency of All Items in the List} \div \text{Total Tokens in the Corpus} * 100\%$$

Table 5.9 provides the results of lexical coverage of the LoSAWEL words in the whole corpus. The lexical coverage was conducted on the Sketch Engine, by searching each item for word frequency.

Table 5.9 Lexical coverage of the LoSAWEL

Items in the LoSAWEL and Its Related Individual Lists	Lexical Words	Function Words	Foreign Word	Abbreviations	Total
Number of Items	311	45	1	2	359
Item Frequency in the Corpus	853,585	1,299,470	54,530	8,547	2,216,132
Tokens in the Whole Corpus	5,357,636 (102,576 types)				
Lexical Coverage	15.93%	24.25%	1.02%	0.16%	41.36%

As section 5.1.1 mentions, there were 311 lexical words extracted as the specialised vocabulary in the LoSAWEL. These 311 lexical words totally occurred 853,585 times in the whole corpus (5,357,636 tokens), and they provided 15.93% coverage in this corpus (nearly 1/6). It should be mentioned that these 311 types only occupied about 0.3% of all word types in the corpus ( $311/102,576 * 100\% \approx 0.3\%$ ). It manifests that a relatively small amount of specialised words provided a high lexical coverage in the Logistics academic written context.

Furthermore, although function words do not convey any disciplinary meaning and are not subject to the analysis for polysemous words in my study, they are related to the content of the corpus. Therefore, it is also necessary to calculate, as a whole, the lexical coverage of items in the LoSAWEL and that of these function words and other non-lexical words (including foreign words and abbreviations) meeting all statistical criteria mentioned in previous sections (keywords analysis, dispersion and range). It should be stated that the #LancsBox 6.0 automatically recognized the foreign word *et al* as two word types, generating frequency for each type. However, it should be actually considered as one. Thus, these two “types” were combined as one in an individual list, and their frequency should be added together in calculation. Table 5.9 shows the lexical coverage of the function words, the foreign word, abbreviations and that of total related words respectively. It can be seen that in total 359 words provided a coverage of 41.36% in the whole corpus. The next section focuses on comparing the LoSAWEL and other corpus-derived word lists.

#### 5.1.4 Comparison between the LoSAWEL and other corpus-derived word lists

##### *General use words and the LoSAWEL*

To analyse the nature of specialised words in the LoSAWEL in depth, a further analysis was conducted to see the lexical profile of words in the LoSAWEL. The Nation’s (2012) BNC/COCA 25 base word lists were chosen as the reference word lists, since this series of word lists was acknowledged as the one with good reputation to represent general English contexts (Nation, 2016). All 311 word types in the LoSAWEL were categorized into the BNC/COCA general use frequency levels. It should be clear here that the Nation’s (2012) BNC/COCA 25 base word lists choose word families as the unit of counting which is different from that used in the LoSAWEL (word type). This means that the words in the LoSAWEL cannot be directly compared with those in the BNC/COCA 25 base word lists. A checking should be conducted to see which member(s) of a word family in the BNC/COCA 25 base word lists also appeared in the LoSAWEL. For example, the word family PRACTICE (in the 2<sup>nd</sup> 1,000 level) has two members *practice* and *practical* included in the LoSAWEL. Table 5.10 presents the results of number of items of the LoSAWEL in each BNC/COCA general use frequency level. It also shows the proportions of overlapping items at each frequency level, lexical coverage it provides and item examples.

Table 5.10 The LoSAWEL words in the Nation's (2012) BNC/COCA 25 base word lists

BNC/COCA Frequency Level	Number of Items in the LoSAWEL	Percentage in the LoSAWEL	LoSAWEL Lexical Coverage in the LAWEC corpus	Examples of Items
1 <sup>st</sup> 1,000	106	35.8%	6.32%	ability, addition
2 <sup>nd</sup> 1,000	127	40.4%	6.35%	benefits, common
3 <sup>rd</sup> 1,000	71	22.9%	2.96%	discussion, effects
4 <sup>th</sup> 1,000	5	1.5%	0.03%	empirical
5 <sup>th</sup> 1,000	1	0.3%	0.10%	optimal
6 <sup>th</sup> 1,000	1	0.3%	0.17%	logistics
7 <sup>th</sup> - 25 <sup>th</sup> 1,000	n/a	n/a	n/a	n/a
Off List	n/a	n/a	n/a	n/a
Total	311	100%	15.93%	n/a

It can be observed that all LoSAWEL items appeared in the first six BNC/COCA word lists. Specifically, the LoSAWEL items at the 1<sup>st</sup> and 2<sup>nd</sup> 1,000 BNC/COCA frequency levels were more than 3/4 among all specialised vocabulary in the list. In addition, 97.75% items in the LoSAWEL were among the most frequent 3,000 BNC/COCA word lists, with only seven in later lists. It also demonstrates a fact that the LoSAWEL items are mostly general high frequency words.

#### *General academic written words in the LoSAWEL*

Another analysis was conducted to find out the percentage of the general academic words had in the LoSAWEL. Two general academic word lists, Coxhead's (2000) AWL and Gardner and Davies' (2014) AVL, were chosen as the reference lists. They were selected because they are widely accepted to be representative in general academic written English contexts and generated with the assistance of corpora. Moreover, these two corpus-derived word lists represent two main methods widely used in word list development, with the AWL relying on ready-made list and the AVL built on scratch. Choosing these lists as reference may provide a fuller picture on the general academic words in the LoSAWEL.

Table 5.11 presents the overlap between the items in the LoSAWEL and those in the AWL and the AVL respectively. As the AWL and the AVL choose different counting unit (word family in the AWL and lemma in the AVL). The members of word families and those of lemmas in the AWL and the AVL were checked to see which of them overlap with items in the LoSAWEL.

Table 5.11 General academic written words in the LoSAWEL

Word List	Items in the LoSAWEL	Percentage in the LoSAWEL	Lexical Coverage in the LAWEC corpus	Examples of Items
AWL	94	31.13%	4.03%	areas, create
AVL (1-500)	196	64.90%	8.80%	study, management
AVL (all)	205	67.88%	9.30%	further, optimal

As Table 5.11 shows, 94 word types are in both the LoSAWEL and the AWL. These word types provided are about a third of all items in the LoSAWEL and provide a lexical coverage of 4.03% in the whole corpus. For the AVL, two comparisons were made: one between the LoSAWEL and the first 500 lemmas of the AVL and the other on the all lemmas ( $n = 3015$ ) in the AVL. Nearly 2/3 of the LoSAWEL word types occurred in the AVL (1-500), with a lexical coverage of 8.80% in the corpus. Whilst only 9 more types were added if the whole AVL was checked. Such results manifest that there was a high proportion of general academic words in the LoSAWEL, reflecting the academic genre of the collected texts. This also indicates that there exists a certain group of general academic words which is frequently used across a wide range of disciplines.

The statistics also show that generally the LoSAWEL had a much larger overlap with the AVL, either for the first 500 lemmas or the whole list, when compared to that with the AWL. Such difference in the overlap was almost twice between the statistics of the AWL and those of the AVL. This could be because the AWL excluded 2,000 high frequent words when developing the list, while the AVL developers did not rely on any previous word lists in development. If the lexical coverage of the 2,000 high frequent words was considered and added to that of the AWL, it would provide higher lexical coverage than that of the AVL.

Moreover, nearly 1/3 of LoSAWEL items were not included in the AVL, and about 2/3 of LoSAWEL items were not in the AWL. The differences of three word lists involved comparison here suggest the discipline-specific scope of my vocabulary list.

## 5.2 Comparison of LoSAWEL item word senses in Logistics academic written context, general academic written context and general use context

### 5.2.1 Word senses of LoSAWEL items in Logistics academic written context

The LoSAWEL consisted of 311 individual word types. Word sense analysis was conducted for all 311 items via the Sketch Engine in the LAWEC, the corpus representing the Logistics academic written context. It followed procedures presented in section 4.3, namely, collocation analysis, concordance line reading, and general and discipline-specific dictionary checking. Each item was annotated with sense definition, word sense occurrence (presented in percentage) and corresponding note on part of speech. All LoSAWEL items are *in italics* in examples, and in ***bold italics*** in concordance extracts. Table 5.12 presents an example of word sense identification in the LAWEC. The full list of LoSAWEL item word senses in the LAWEC is in Appendix B.

Table 5.12 An example of word sense identification in the LAWEC

Item (Word Type)	Sense Definition	Occurrence (Percentage)	Extracted Concordance	Number of Senses
<i>Figure</i>	diagram ( <i>n.</i> )	99.99%	Testing H1–H4 involves analysis of the paths shown in <b><i>Figure 1.</i></b> (LM200208)	2
	number/statistics ( <i>n.</i> )	0.01%	... this <b><i>Figure</i></b> has a coefficient of variation ... (BL180103)	

This example shows word sense information of the case *Figure*. Two senses were identified for this individual word type, and both senses were marked as noun. These two senses presented a big difference in occurrence: the sense of “diagram” accounts for 99.99% in concordance lines and that of “number/statistics” only accounts for 0.01%.

#### *The number of word senses in the LAWEC*

Word sense analysis in the LAWEC generated 472 identified senses from all 311 LoSAWEL items. Items were categorized according to the number of senses identified, with percentage of each category provided (see Table 5.13).

Table 5.13 The number of word senses and items

Number of Senses	1 sense	2 senses	3 senses	4 senses	5 senses	6 senses	7 senses	Total
Number of Items	206 (66.24%)	72 (23.15%)	18 (5.79%)	10 (3.22%)	3 (0.96%)	1 (0.32%)	1 (0.32%)	311

Table 5.13 shows that 206 of 311 word types had 1 sense, the largest category among seven. It suggests that, in the LAWEC, 66.24% LoSAWEL word types were monosemous words, and more than a third of items (33.76%;  $n = 105$ ) were polysemous words. Findings of polysemous words and senses in the LAWEC will be presented in detail in section 5.3. This table also shows that there is a generally downward trend of the number of senses in categories and their corresponding percentages among categories. Most of items have one sense, and fewer items have several senses (also see Figure 5.1).

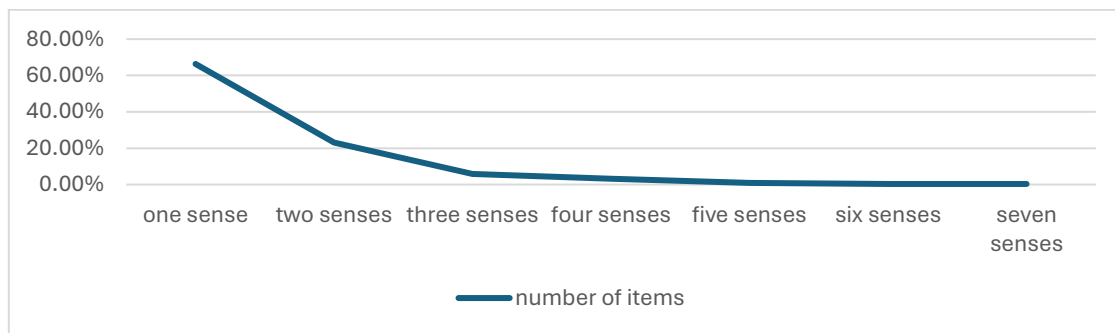


Figure 5.1 Correlation between the number of word senses and the number of items (with corresponding percentages)

### *Word senses and corresponding part of speech in the LAWEC*

By referring to the COBUILD dictionary (2023) and the Logistics discipline-specific dictionary and glossary IWRA (2018), word senses were categorized according to their part of speech. Table 5.14 summarizes the number of word senses in each part of speech category and its corresponding percentage. A selected example of word sense in each part of speech category is also presented.

Table 5.14 Part of speech categories of word senses in the LAWEC

Part of Speech Category	Number of Senses (Percentage)	Example of Item and Sense	Extract Concordance Line
noun	198 (41.95%)	<i>Figure</i> : diagram	... , as was done by Carter et al. (2015) and as is also shown in <b>Figure 2</b> . (BL180102)
verb	114 (24.15%)	<i>study</i> : to explore something	We <b>study</b> 21 extended (five-tier) supply chain networks of firms .... (BL210203)
adjective	82 (17.37%)	<i>additional</i> : extra things	<b>Additional</b> research can provide insights into omnichannel distribution, .... (TR220601)
adverb	15 (3.18%)	<i>further</i> : to introduce other issues of the same topic	<b>Further</b> , this research adds a second dimension, competition, to study the dynamics of retail-switching behaviour of ... (LR200303).
preposition	4 (0.85%)	<i>per</i> : to describe rates or ratios	Moreover, the fatal accident rates have risen from a low of 1.29 <b>per</b> 100 million vehicle ... (BL180202)
Total	413 (87.50%)	n/a	n/a

In word sense analysis, 87.50% identified senses ( $n = 413$ ) were classified into five part of speech categories. The remaining 12.50% senses were expressed through register-specific collocations, which will be discussed in the next sub-section. Of all five categories of part of speech, noun is the largest group, accounting for more than 40% of all senses identified. This was to be expected, since noun is also the most frequent part of speech in English. Also, the LAWEC consisted of written academic texts, which is considered to contain more nominalisation. Compared to that of nouns, percentages of other categories are much smaller. Verbs accounted for 24.15% of senses, and adjectives for 17.37%. Much fewer adverb senses were identified, only for 3.18%. There were 4 senses (0.85%) categorized as prepositions. This is beyond expectation, because function words were excluded in LoSAWEL item selection based on a function word list (see Appendix A). It may result from a blurred boundary between lexical and function words, or polysemy of prepositions. Polysemy of prepositions will be discussed further in section 5.3 and 6.3. Table 5.15 summarises all four word types with preposition senses in the LAWEC.

Table 5.15 Four preposition word senses and their word types in the LAWEC

Item	Sense	Occurrence (Percentage)	Concordance Line Extracted	Number of Senses
<i>following</i>	after a particular event ( <i>prep.</i> )	2.27%	... realist synthesis is used to summarize the findings <b>following</b> pre-defined research questions. (LM200105)	3
<i>regarding</i>	indicate the subject which is being discussed ( <i>prep.</i> )	100%	Accordingly, questions <b>regarding</b> how to redesign supply chains to manage risk and improve sustainability have been moving up ... (BL180102)	1
<i>including</i>	to introduce examples that are part of something being discussed ( <i>prep.</i> )	100%	However, many reselling platforms, <b>including</b> Amazon.com and JD.com, ... (TR220713)	1
<i>per</i>	to express rates and ratios ( <i>prep.</i> )	79.43%	Moreover, the fatal accident rates have risen from a low of 1.29 <b>per</b> 100 million vehicle miles in 2010 to 1.40 in 2014 (FMCSA 2016). (BL180202)	3

### *Word senses and idiomatic register-specific collocations in the LAWEC*

Meanings may also be expressed through register-specific collocations, where at least two words co-occur together to express a meaning holistically. In these collocations, at least one LoSAWEL item was the component. Some collocations identified may not convey disciplinary knowledge, but they present different functions in context, when compared to their component words. For example, the LoSAWEL item *terms* collocates with “in” and “of”, forming a register-specific collocation “in terms of”. This collocation has nouns or noun phrases occurring after it, and is for “specifying which aspect of it you are discussing or from what point of view you are considering it” (COBUILD, 2023, p.1535), for example:

..., but they also need to be flexible **in terms of** time, content and resources to incorporate new requirements and open up for innovative packaging solutions ... (LM200302)

This phenomenon also applies to other collocations like ‘for example’, ‘according to’ and ‘in general’ (see Table 5.16). Specifically, ‘for example’ is frequently used at the beginning of or in the middle of sentences, proceeding nouns, noun phrases, or a main clause. Differently, a main clause was not observed after the individual word ‘example’. ‘according’ was observed always to collocate with ‘to’ at the beginning of or in the middle of sentences. ‘in general’ was frequently observed at the beginning of sentences to indicate something is true in most cases, which modifies the sentence or verbs, but ‘general’ works as an adjective, and only occurs before nouns or noun phrases.

Of all 472 senses in the LAWEC, 59 senses (12.50%) were expressed through register-specific collocations. Two categories of register-specific collocations were observed in the LAWEC. They are marked with (*ph.*) and (*tm.*) in the full list of word sense in the Appendix B, based on whether they convey disciplinary knowledge or not. Collocations marked with (*ph.*) do not convey disciplinary knowledge but have different functions when compared to their component words. I used (*ph.*) to represent the [PHRASE] marker used in the COBUILD dictionary (2023). For those conveying disciplinary knowledge and listed in the IWAL discipline-specific dictionary and glossary (2018), I used (*tm.*) to represent these collocations are used as terms in the corpora. Hereafter, I shall call them general list collocations and disciplinary collocations respectively. For example, all occurrences of *addition* were found in the fixed collocation “in addition (to)” in the LAWEC. It means “to mention another item connected with the subject that you are discussing”, and this collocation is marked as [PHRASE] in the COBUILD dictionary (p.19). General list collocations like “in addition (to)”, therefore, were marked with (*ph.*). I used the mark (*ph.*) in Appendix B because it does not appear to specialise to convey disciplinary knowledge. There were 22 general list collocations identified, accounting for 4.66% of all senses. Table 5.16 below presents other five examples and extracts of general list collocations.

Table 5.16 Examples of general list collocations in the LAWEC

Item	General list collocations	Percentage in citations	Concordance Line Extracted
<i>general</i>	in general	69.61%	<b>In <i>general</i></b> , lean manufacturing as a management approach seeks to minimise waste and wastage. (OL220401)
<i>terms</i>	in practical terms	0.75%	<b>In <i>practical terms</i></b> , our results should assist shipowners and related companies to understand changes in demand ... (MEL190305)
<i>example</i>	for example	42.80%	<b>For <i>example</i></b> , Tita (2017) quotes the president of a six-truck motor carrier as stating, ... (BL200104)
<i>order</i>	in order to	37.68%	<b>In <i>order to</i></b> answer the above-mentioned two research questions, the current research adopts a multiple case study methodology ... (LR180403)
<i>according</i>	according to	100%	<b>According to</b> Hair et al. (2011) and Sarstedt et al. (2014), convergent validity is used to measure and examine the extent that ... (ML200707)

Disciplinary collocations are related to disciplinary knowledge. They are not listed in the COBUILD dictionary but are listed as entries in the Logistics discipline-specific dictionary and glossary IWRA (2018). For example, *lead* collocates with “times” (LogDice = 10.88) and “time” (LogDice = 10.14) in “lead time(s)”. An extract is presented below:

Customers expect shorter **lead times**, which puts pressure on reducing throughput times in warehouses, that is, the total time required from order placement to it being picked, sorted, packed and shipped (Hübner et al., 2015). (PDLM180901)

This collocation refers to “the total time between an order’s placement and its receipt” (IWRA, 2018, p.69), and is used as a term. Therefore, tm-collocations like “lead time(s)” were marked with (*tm.*) in this research, to distinguish with those of (*ph.*) in Appendix B. There were 37 disciplinary collocations identified in the LAWEC, accounting for 7.84% of all identified senses. Table 5.17 below presents other five examples and extracts of disciplinary collocations.

Apart from conveying disciplinary knowledge, checking the concordance lines showed that disciplinary collocations are an indication of its relations with Business, Transportation or Mathematics. For example, in the extracts of “lead time(s)” above and those in Table 5.17 below, “customers”, “customer satisfaction”, “pricing” and “e-retailers” are related to business events. “shipped” and “containers” are an indication of transportation. “quantity” and specific numbers listed show their relations with mathematics. These may suggest an interdisciplinary feature of Logistics.

Table 5.17 Examples of disciplinary collocations in the LAWEC

Item	Tm-collocations	Percentage in citations	Concordance Line Extracted
<i>case</i>	case pack (quantity)	0.08%	However, the <b>case pack quantity</b> can increase by 12 units (going from one dozen items in a <b>case pack</b> to two dozen items in a case pack) ... (BL180301)
<i>means</i>	means objectives	20.62%	..., by prioritising increasing competitiveness among identified <b>means objectives</b> in the APP. (MEL200303)
<i>order</i>	order fulfillment	6.24%	Lei, Jasin, and Sinha (2018) consider a similar joint pricing and <b>order fulfillment</b> problem for e-retailers. (NRL200801)
<i>size</i>	lot size	15.76%	..., and the <b>lot size</b> of containers handled by customers. (MEL200104)

<i>chain</i>	supply chain (management)	94.56%	For example, in the framework proposed by Lummus et al. (2003), a flexible <b>supply chain</b> leads to customer satisfaction and inventory minimization. (LM180115)
--------------	------------------------------	--------	--

### 5.2.2 Word sense of LoSAWEL items in the general academic written context

Word sense analysis for 311 LoSAWEL items was then conducted in the BAWE using Sketch Engine. It followed the same procedures as those in the LAWEC, the only differences being the choice of dictionary. If a fixed collocation was observed and only occurred in texts from a particular subject in the BAWE, a discipline-specific dictionary of that subject was then used for double checking. The rationale and discipline-specific dictionary selection was presented in section 4.3.

503 senses were identified in the BAWE. The following information of word senses was identified: meaning, part of speech and percentage in citations. The list of word sense in the BAWE was also in Appendix B. Table 5.18 presents an example of word sense identification in the BAWE.

Table 5.18 An example of word sense identification in the BAWE

Item	Sense Definition	percentage of citations	Extracted Concordance	Number of Senses
<i>role</i>	to have a particular position (n.)	99.81%	As previously stated, antibodies play a crucial <b>role</b> in the immune system, ... (182)	2
	in a fixed collocation “role play” (n.)	0.19%	... intend to outline the development of a child's knowledge of social roles using evidence from <b>role</b> play. (2106)	

In the BAWE, two noun senses were identified in the concordance of *role*. Both senses typically collocated with “play” (LogDice = 11.32). The sense “to have a particular position” accounted for most of occurrences (99.81%). It was frequently observed in multi-word expressions like “play an important role (in)”, “play a crucial role (in)”, or “play a vital role (in)” to show a person or an object has a particular function in a situation or society. The other sense was observed in a fixed collocation “role play”, which means to “imitate the character and behaviour of someone” (0.19%). It had only six occurrences in concordances, from four different texts in Linguistics. Due to all its occurrence in texts of a particular discipline in the BAWE, a double check was needed for deciding whether it is a tm-collocation. *The Oxford Concise Dictionary of Linguistics 3rd edition* (Matthews, 2014) was used for

double checking, but “role play” was not listed as terminology. Therefore, “role play” was not marked with (*tm.*).

### *Word sense and corresponding part of speech in the BAWE*

The 503 senses in the BAWE were also categorized according to their part of speech. Table 5.19 summarizes the number of senses and corresponding percentages in each part of speech category. An example of sense and an extract is given for each category respectively.

Table 5.19 Part of speech categories of word senses in the BAWE

Part of Speech Category	Number of Senses (Percentage)	Example of Item and Sense	Extract Concordance Line
noun	213 (42.35%)	<i>difference</i> : the way in which two things are unlike each other	... there seemed to be something of a <b><i>difference</i></b> between the types of customers. (8)
verb	122 (24.25%)	<i>found</i> : usually in passive voice, to show something exists	Amazingly almost all adult Neanderthal fossils <b><i>found</i></b> to date show an astounding array ... (2760)
adjective	87 (17.30%)	<i>further</i> : something additional	..., an occupation which had previously been female, a result which had many <b><i>further</i></b> implications for women in both gender and class terms. (2)
adverb	17 (3.38%)	<i>further</i> : to introduce some related topics	<b><i>Further</i></b> , Blascovich, Mendes, Hunter and Saloman (1999) found ... (96)
preposition	3 (0.60%)	<i>including</i> : containing	..., the organization had all the strongpoint and weakness of traditional hierarchical organizations, <b><i>including</i></b> formal process, ... (633)
Total	442 (87.87%)	n/a	n/a

Table 5.19 shows that 442 of 503 senses (87.87%) were classified into five part of speech categories. The remaining 61 senses (12.13%) were expressed through register-specific collocations which will be discussed in the next sub-section. Generally, the percentages of senses of five categories in the BAWE are quite similar to those in the LAWEC. Specifically, the noun category is the largest group, accounting for more than 40% of senses identified in the BAWE. Verbs and adjectives accounted 24.25% and 17.30%

respectively. A few adverb and preposition senses were identified, both less than 5%, with the preposition category as the smallest.

### *Word sense and idiomatic register-specific collocations in the BAWE*

Register-specific collocations were also observed in the BAWE. There were 61 register-specific collocations identified, with 29 general list collocations and 32 disciplinary collocations. They each account for around 6% of all senses. Five selected examples of two register-specific collocation categories were provided in Table 5.20 and 5.21 respectively.

For disciplinary collocations, information of specific disciplines they occur was also provided in []. Disciplinary collocations from 13 disciplines were observed: Biology, Business, Economics, Engineering, Food Science, Hospitality, Law, Linguistics, Management, Medicine, Physics, Politics, and Sociology. For example, *review* collocated with “judicial” (LogDice = 10.76), forming a disciplinary collocation “judicial review”. It only occurred in texts from Law, and was listed as an entry in the *Oxford Concise Dictionary of Law 10 edition* (Law, 2022). This collocation was marked with (*tm.*) [Law], as shown in Table 5.21 and Appendix B.

Table 5.20 Examples of general list collocations in the BAWE

Item	General list collocations	Percentage in citations	Concordance Line Extracted
<i>addition</i>	in addition (to)	100%	<b>In addition</b> they also described the increasing importance of echocardiography, ... (196)
<i>case</i>	in case	1.14%	Participants were provided with the experimenters' contact details <b>in case</b> they later wanted to withdraw their information. (67)
<i>order</i>	in order to	96.47%	..., <b>in order to</b> create an institution that through its comprehensiveness, ... (296)
<i>terms</i>	in terms of	72.06%	By expressing the monopolist's marginal revenue curve <b>in terms of</b> elasticity, we can show that the monopolist's price mark-up, ... (744)
<i>account</i>	take (...) into account	99.51%	A jury may <b>take into account</b> the degree of harm risk and the defendant's 'state of mind' ... (270)

Table 5.21 Examples of disciplinary collocations in the BAWE

Item	Disciplinary collocations	Percentage in citations	Concordance Line Extracted
<i>abstract</i>	abstract labour [Sociology]	61.40%	<b>Abstract labour</b> is also social labour as they confer social quality on commodities. (1230)
<i>review</i>	judicial review [Law]	62.50%	The Exclusivity Principle [1] requires that claims for judicial review be brought by <b>judicial review</b> procedure [2] rather than ordinary civil procedure. (258)
<i>inventory</i>	phonetic inventory [Linguistics]	5.19%	... a descriptive profile of the child's speech and language abilities is made, for example by collecting a <b>phonetic inventory</b> as well as conducting the Phonological Assessment of Child Speech (PACS). (2226)
<i>number</i>	Reynolds number [Engineering]	30.72%	The pipe work is set up so that the <b>Reynolds number</b> is at a transition stage somewhere between laminar and turbulent flow, ... (101)
<i>state</i>	state intervention [Politics]	23.78%	So the <b>state intervention</b> is undermined, and the state policies have three kinds of change: ... (982)

### 5.2.3 Word senses of LoSAWEL items in the general use context

Word sense analysis for 311 LoSAWEL items was then conducted in the BNC1994 NAW using Sketch Engine. The procedures were the same as mentioned in previous sections, namely, collocation analysis, concordance line checking and COBUILD dictionary checking.

562 senses were identified in the BNC1994 NAW. For each sense, part of speech and percentage of citations were also identified (see Appendix B). Table 5.22 presents an example of word sense identification in this corpus.

In the BNC1994 NAW, three noun senses were identified in the concordances of *Figure*. The sense of “diagram to show information” had the largest percentage in citations (98.00%). Senses of “a particular number or statistics” and “people” were also observed, with much smaller percentage of occurrence in citations.

Table 5.22 An example of word sense identification in the BNC1994 NAW

Item	Sense Definition	Percentage of Citations	Extracted Concordance	Number of Senses
<i>Figure</i>	diagram to show information ( <i>n.</i> )	98.00%	... they are composed largely of space, as shown in <b>Figure 4</b> . (3217)	3
	a particular number or statistics ( <i>n.</i> )	1.10%	If the exact gross pay is not in the table, use the nearest lower <b>Figure</b> ... (1577)	
	referring to “people” ( <i>n.</i> )	0.90%	... a medieval sculpture of a <b>Figure</b> broken on a wheel shown above a corpse twisted around a swastika. (2488)	

#### *Word sense and corresponding part of speech in the BNC1994 NAW*

These 562 senses in the BNC1994 NAW were categorized according to their part of speech. Table 5.23 summarizes the number of senses and corresponding percentages in each part of speech category. An example of sense and an extract is given for each category respectively.

Table 5.23 shows that 519 of 562 senses (92.35%) were classified into five part of speech categories. The remaining 43 senses (7.65%) were expressed through register-specific collocations which will be discussed in the next sub-section. The noun category is the largest group, accounting for more than 45% of senses identified in this context. Verbs and adjectives accounted for 23.13% and 20.28% respectively. A small number of senses were classified into adverbs or prepositions, in total less than 4%, with the preposition category the smallest.

Table 5.23 Part of speech category of word senses in the BNC1994 NAW

Part of Speech Category	Number of Senses (Percentage)	Example of Item and Sense	Extract Concordance Line
noun	253 (45.02%)	<i>address</i> : referring to place ( <i>n.</i> )	the <b>address</b> of the first store location in the pair. (2452)
verb	130 (23.13%)	<i>found</i> : to notice somebody in a particular situation ( <i>v.</i> )	she <b>found</b> herself, back in England, designing her own kitchen. (1)
adjective	114 (20.28%)	<i>general</i> : to involve a lot of people ( <i>adj.</i> )	But wait, on the political horizon there comes a <b>general</b> election. (90)
adverb	18 (3.20%)	<i>mainly</i> : to describe most of something ( <i>adv.</i> )	The woods beyond the rectory garden consisted <b>mainly</b> of beech trees,. (648)
preposition	4 (0.71%)	<i>including</i> : containing ( <i>prep.</i> )	several endemic species, <b>including</b> a red musk parrot, banded iguana and insular flying fox. (245)
Total	519 (92.35%)	n/a	n/a

### *Word sense and idiomatic register-specific collocations in the BNC1994 NAW*

Register-specific collocations were also observed in the BNC1994 NAW. There were 43 register-specific collocations, accounting for 7.65% of all senses identified in the BNC1994 NAW. 42 were general list collocations and only one for disciplinary collocations. Five selected examples were presented in Table 5.24 below.

The only disciplinary collocation in the BNC1994 NAW is “market share”. The COBUILD dictionary classified it as a noun, but in this research I classified it as a disciplinary collocation based on the following reasons. First, “market share” is an entry listed in the Logistics disciplinary specific dictionary IWLA (2018, p.76), which meets the criteria of being a disciplinary collocation in Logistics (see section 4.4 for criteria). Second, “market share” is the only one that was identified in both the LAWEC and the BNC1994 NAW, with all other disciplinary collocations being in the LAWEC only. Therefore, for consistency of classification in two corpora and cross-corpus comparison, it was also classified as a disciplinary collocation in the BNC1994 NAW. two extracts of “market share” in the BNC1994 NAW are as follows:

And as long as he was prepared to give the spending money, ie in the form of a voucher, er we were taking away the **market share** away from etcetera of this world. (78)

The acquisition increased GA's Canadian premium income by just over a third, lifting its **market share** to just over 8%. (487)

Table 5.24 Five examples of general list collocations in the BNC1994 NAW

Item	General List Collocations	Percentage of Citations	Concordance Line Extracted
<i>addition</i>	in addition (to)	100%	<b>In addition</b> , an agreement between an accepting shareholder and the bank ... (205)
<i>case</i>	in case	1.78%	<b>In case</b> anybody doesn't know me, I think I've had dealings with most people here, ... (567)
<i>means</i>	(this is) by no means	9.51%	Nigel is <b>by no means</b> the fool she took him for. (284)
<i>particular</i>	in particular	10.59%	<b>In particular</b> , we have re-designed the back page of the newsletter ... (550)
<i>set</i>	set aside	23.05%	In another drawer, <b>set aside</b> from everything else, he found other letters that were full of love. (668)

The next two sections are about the word sense comparison between different contexts: the comparison between the LAWEC and the BAWE is in section 5.2.4, and that between the LAWEC and BNC1994 NAW is in section 5.2.5.

## 5.2.4 Word sense comparison between the LAWEC and the BAWE

This section aims to answer the first sub-question of RQ2 *To what extent are the words identified used in ways specific to Logistics? a) is their use different from the general academic context?* The comparison was made between word senses identified in the LAWEC (in section 5.2.1) and those in the BAWE (in section 5.2.2). It will be discussed on three aspects:

- (1) an overview of word sense comparison of number of word senses identified in each part of speech and register-specific collocation category;
- (2) word senses/register-specific collocations that are identified in the LAWEC but not meeting selection criteria in the BAWE;

- (3) word senses/register-specific collocations shared between the LAWEC and the BAWE, but with different percentage in citations (hereafter LAWEC-BAWE shared sense/register-specific collocation (D), D representing “different percentage in citations”).

At the end of this sub-section, some examples of word sense comparison between the LAWEC and the BAWE are listed for analysis.

*An overview of word sense comparison between the LAWEC and the BAWE*

Word sense analysis shows that 472 and 503 senses were identified in the LAWEC and the BAWE respectively. Table 5.25 shows the number of word senses identified at each LoSAWEL level in two corpora. The comparison shows that, in the BAWE, more word senses were identified at each LoSAWEL level, when compared to those in the LAWEC. Although more senses were identified in the BAWE, there is no big difference on percentage of senses at levels between corpora. For example, word senses identified at Level 2 accounted for around 40% of all senses in both corpora. The percentages of word senses at other levels in the LAWEC were similar to their counterparts in the BAWE.

Table 5.25 Comparison of word senses identified in the LAWEC and the BAWE at each LoSAWEL level

Level	Number of Items	Number of Senses in the LAWEC	Percentage of Senses in the LAWEC	Number of Senses in the BAWE	Percentage of Senses in the BAWE
Level 1	106	179	37.92%	196	38.97%
Level 2	127	194	41.10%	200	39.76%
Level 3	71	91	19.28%	97	19.28%
Level 4	7	8	1.69%	10	1.99%
Total	311	472	≈ 100.00%	503	100.00%

The comparison of number of word senses in each part of speech category between the LAWEC and the BAWE is presented in Table 5.26, and that on each register-specific collocation category is in Table 5.27. Table 5.26 shows that, in both corpora, the largest part of speech category is noun, and the smallest one is preposition. It also shows that in the BAWE, all categories of part of speech have more senses identified when compared to those in the LAWEC, except for prepositions. This is because the BAWE covers a wider range of disciplines than the LAWEC does. In contrast, 3 preposition senses were identified in the BAWE but 4 in the LAWEC, which will be discussed further in next sub-sections. Percentage of senses in each part of speech category in the LAWEC is similar to its counterpart in the BAWE. For example, in both corpora, noun senses accounted for more than 40%, verbs for around 24%

and adjectives for around 17%. A small number of adverb senses and prepositions was identified in both corpora, with adverbs accounting for nearly 3% and prepositions for less than 1%.

Table 5.26 Comparison of word senses in part of speech categories in the LAWEC and the BAWE

Part of Speech	Number of Senses in the LAWEC	Percentage of Senses in the LAWEC	Number of Senses in the BAWE	Percentage of Senses in the BAWE
noun	198	41.95%	213	42.35%
verb	114	24.15%	122	24.25%
adjective	82	17.37%	87	17.30%
adverb	15	3.18%	17	3.38%
preposition	4	0.85%	3	0.60%
Total	413	87.50%	442	87.87%

Table 5.27 Comparison of word senses in register-specific collocation categories in the LAWEC and the BAWE

Part of Speech	Number in the LAWEC	Percentage of Senses in the LAWEC	Number in the BAWE	Percentage of Senses in the BAWE
General list collocation	22	4.66%	29	5.77%
Disciplinary collocation	37	7.84%	32	6.36%
Total	59	12.50%	61	12.13%

Table 5.27 shows that two categories of register-specific collocations accounted for around 12% of senses in both corpora. More general list collocations were identified in the BAWE ( $n = 29$ ; 5.77%), than in the LAWEC ( $n = 22$ ; 4.66%). However, there were more disciplinary collocations identified in the LAWEC ( $n = 37$ ; 7.84%) than in the BAWE ( $n = 32$ ; 6.36%).

Based on whether senses were shared between two academic corpora, three kinds of senses and register-specific collocations were found in comparison: a) only in the LAWEC but not meeting selection criteria in the BAWE, b) shared senses (D): shared between the LAWEC and the BAWE but with different percentage in citations, and c) shared senses (S): shared between the academic corpora with similar percentage in citations. The Classification method I used was presented in section 4.3. Specifically, if the occurrence of a word sense is 1.5 times higher or lower than that of its counterpart

in another corpora, this sense is considered as a LAWEC-BAWE shared sense (D); if not, it is a LAWEC-BAWE shared sense (S). Table 5.28 presents the number and percentage of three kinds of senses. This shows that more than half of senses identified in the LAWEC (52.97%, n=250) have similar uses in the two academic corpora. The rest are different from general academic contexts: 107 senses were unique to the LAWEC, and 115 were shared senses with different percentage in citations. Different uses of senses in the LAWEC will be discussed below.

Table 5.28 Number of senses shared or not shared between the LAWEC and the BAWE

	Senses only in the LAWEC not meeting selection criteria in the BAWE	LAWEC-BAWE shared sense (D)	LAWEC-BAWE shared sense (S)
Number of Senses	107	115	250
Percentage	22.67%	24.36%	52.97%

*Word senses found only in the LAWEC but not meeting selection criteria in the BAWE*

107 senses were identified only in the LAWEC, but not meeting selection criteria in the BAWE. They accounted for 22.67% of all 472 senses identified in the LAWEC. 71 of 107 senses were classified according to their part of speech categories and LoSAWEL levels in Table 5.29. The remaining 36 were identified in register-specific collocations, which will be discussed later in this sub-section.

Table 5.29 Part of speech and LoSAWEL levels of senses found only in the LAWEC but not meeting selection criteria in the BAWE

Part of Speech	noun	verb	adjective	adverb	preposition	total
Level 1	13	7	10	0	1	31
Level 2	8	14	4	0	0	26
Level 3	10	2	2	0	0	14
Level 4	0	0	0	0	0	0
total	31	23	16	0	1	71

Table 5.29 shows that senses that were only identified in the LAWEC but not meeting selection criteria in the BAWE were all from Level 1-3. In contrast, at Level 4, all senses identified in the LAWEC were also identified in the BAWE. As shown in section 5.1, items at Level 1-3 are general high

frequency words, and those at Level 4 are general mid frequency words. These suggests that general high frequency words may have different senses between academic contexts. If general word frequency goes lower (e.g., general mid-/low-frequency words), word senses are used consistently between academic contexts. Table 5.29 also shows that almost all these senses were in noun, verb and adjective categories. There was one preposition sense in the LAWEC which was not identified in the BAWE because of selection criteria. All adverb senses in the LAWEC were observed in the BAWE, considered as shared senses.

Table 5.30 presents 10 intentional selected examples of these LAWEC senses that were not identified in the BAWE with current selection criteria. At each level, one example was chosen for noun, verb and adjective senses respectively, and the remaining one was for preposition. Here I focus on some examples in Table 5.30: *industry*, *review*, *limited* and *following*, one example for each part of speech. *industry* means “the work and process of making products in factories” which accounted for 14.45% concordance lines checked. This sense was not identified through collocation analysis and concordance reading in the BAWE, where *industry* was used to refer to a specific trade (e.g., “hotel industry”). In the LAWEC, *review* had two senses identified, but its verb sense “to check or consider something carefully” was not observed with current selection criteria in the BAWE where 17 collocates (LogDice  $\geq 9$ ) and 1,495 concordance lines were checked. For the adjective sense “not in a large number, size, degree or range” of *limited*, it was identified in the LAWEC where *limited* collocated with “capacity” (LogDice = 9.40) and “resources” (LogDice = 9.33). However, in the BAWE, this item collocated with “Pearson” (LogDice = 10.28) and “Education” (LogDice = 10.06), having the sense of “limited company”. Word senses of *industry*, *review* and *limited* in the LAWEC but not identified in the BAWE because of not meeting selection criteria suggest that different senses are generated from different collocations in contexts. The preposition sense of *following* was only identified in the LAWEC, because, in the BAWE, this item was used as an adjective to modify nouns (e.g., “the following equation” or “the following table”) with current selection criteria applied.

Table 5.30 Examples of senses in the LAWEC but not meeting selection criteria in the BAWE

Level	Item	Sense/Use	Occurrence (Percentage)	Concordance Line Extracted
Level 1	<i>area</i>	a particular topic or subject ( <i>n.</i> )	2.86%	Portfolios of customer and supplier relationships are a critical <b>area</b> of research for several reasons. (BL180204)
Level 1	<i>understanding</i>	to know why or how something happens ( <i>v.</i> )	15.21%	R-O theory furthers the resource-based view of the firm by <b>understanding</b> what a firm does ... (PDL190603)
Level 1	<i>high</i>	in priority or importance ( <i>adj.</i> )	24.45%	"we play a <b>high</b> priority on investment on AEO certification," ... (LM180103)
Level 1	<i>following</i>	after that event ( <i>prep.</i> )	2.27%	<b>Following</b> this analysis, we calculated the inverse mills ratio ... (BL220105)
Level 2	<i>industry</i>	the work and process to make products in factories ( <i>n.</i> )	14.55%	While data capture and Auto-ID technologies have existed in the <b>industry</b> for quite some time, ... (BL210304)
Level 2	<i>supply</i>	to provide something that is required ( <i>v.</i> )	14.55%	where the supplier promises to <b>supply</b> all specified goods that the buyer ... (TR190307)
Level 2	<i>limited</i>	not in a large number, size, degree or range ( <i>adj.</i> )	100%	... where a small geographical area or a very <b>limited</b> number of individuals as the subjects of study is selected. (LR180501)
Level 3	<i>distribution</i>	giving or delivering thing to many people and things ( <i>n.</i> )	57.91%	In this regard, when considering an entire <b>distribution</b> system in a country, ... (LR210202)
Level 3	<i>review</i>	to check or consider something carefully ( <i>v.</i> )	2.66%	The regulator then needs to <b>review</b> them. (BL210404)
Level 3	<i>negative</i>	a negative response/reply/decision sometimes means "no" response ( <i>adj.</i> )	9.78%	. This might produce a <b>negative</b> response, ... (LR200504)

*Register-specific collocations found only in the LAWEC but not meeting selection criteria in the BAWE*

Word sense comparison between the LAWEC and the BAWE also showed that 36 register-specific collocations were identified in the LAWEC but not meeting selection criteria in the BAWE. Table 5.28 presents the number and the LoSAWEL levels of these register-specific collocations.

Table 5.31 Number and LoSAWEL levels of register-specific collocations identified in the LAWEC but not meeting selection criteria in the BAWE

Level	General list Collocations	Disciplinary Collocations	total
Level 1	2	15	17
Level 2	2	17	19
Level 3	0	0	0
Level 4	0	0	0
total	4	32	36

Table 5.31 shows that register-specific collocations identified in the LAWEC but not meeting selection criteria in the BAWE were all from Level 1 and 2. This suggests that general high frequency words may have more different register-specific collocations between contexts. Table 5.31 shows that only 4 general list collocations were identified in the LAWEC but not meeting selection criteria in the BAWE. Table 5.32 presents those four different collocations.

For disciplinary collocations, in section 5.2.1 showed that there were 37 collocations in such category identified in the LAWEC. Table 5.31 shows that 32 of 37 were not observed in the BAWE due to not meeting selection criteria. It was to be expected, since the LAWEC focuses on a specific discipline which is not included in the BAWE, and this discipline may contain many register-specific collocations that are not frequently used in other disciplines. Possible reasons for such big differences will be discussed further in section 6.2. Table 5.33 presents five examples on these disciplinary collocations (*tm.*) only occurring in the LAWEC.

Table 5.32 LAWEC general list collocations not identified in the BAWE with current selection criteria

Item	General list Collocations	Occurrence (Percentage)	Concordance Line Extracted
<i>case</i>	a case in point	0.03%	Figure 4 demonstrates <b>a case in point</b> . Figure shows a CMCC instance encountered in iteration 5 of the Voronoi heuristic solving the 17-center problem,... (NRL220403)
<i>general</i>	in general	69.61%	The logit model allows classifying accurately 92.11% of cases <b>in general</b> ; the sensitivity of the model (i.e., the probability of classifying accurately the efficient containerships) is ... (MEL180202)
<i>knowledge</i>	to the best of our knowledge	48.77%	Specifically, <b>to the best of our knowledge</b> , only two studies have investigated new trade-in strategies, ... (TR200210)
<i>respect</i>	with respect to	100%	...to discuss and collect their opinions in regard to every alternative importance <b>with respect to</b> each sub-criteria. (LR180202)

Table 5.32 presents all 4 general list collocations that were only identified in the LAWEC but not in the BAWE because of not meeting selection criteria. For example, *case* had seven senses identified in the LAWEC and five of them were also identified in the BAWE (see Appendix B). The collocation “a case in point” was only identified in the LAWEC, although it did not frequently occur in concordance lines (0.03%). This collocation occurred in 5 different texts in the LAWEC, which can be considered as a different but infrequent use of *case*, compared to that in the BAWE. The collocation “in general” was observed in concordance lines of *general* because of the high logDice value of its collocate “in” (LogDice = 9.06) in the LAWEC. In the BAWE, although “in” was also one of collocates of *general*, its LogDice value (8.48) did not meet the threshold of collocation analysis (LogDice  $\geq$  9). This may be the reason that “in general” was not identified in the checked concordance lines in the BAWE, because of not checking every collocate. The collocation “with respect to” also had similar situation in the BAWE: “with” was a collocate but its LogDice value did not meet the threshold for further analysis. This is one of the limitations of this research, which will be discussed further in section 7.3. For “to the best of our knowledge”, this collocation was frequently observed in LAWEC; however, it was not

identified in collocation analysis and concordance checking in the BAWE. This may be the reason of authors' academic identity, which will be discussed in section 6.2 later.

Table 5.33 presents examples of LAWEC disciplinary collocations not identified in the BAWE because of current selection criteria. For example, in the LAWEC, *size* (a LoSAWEL item) collocated with "lot" (a non-LoSAWEL word) (LogDice = 10.73). However, in the BAWE, "lot" did not collocate with *size*, which showed a different list of collocates. This situation was also observed in other examples: "picking" collocating with *order*, "pack" collocating with *case*, and "cold" collocating with *chain* were only identified in the LAWEC, instead of in the BAWE. These LoSAWEL items had different collocates between corpora meeting selection criteria. This shows that word sense is affected by the context where it occurs, of which collocations are an indication.

Table 5.33 LAWEC disciplinary collocations but not meeting selection criteria in the BAWE

Item	Disciplinary Collocations	Occurrence (Percentage)	Concordance Line Extracted
<i>case</i>	case pack (quantity)	0.08%	In most cases, picking is completed in case packs (De Koster et al., 1999; Gu et al., 2007), where a picker takes an entire <b>case pack</b> from a storage area and puts it on a transportation unit (e.g. pallet or roll cage). (PDL180403)
<i>order</i>	order picking	5.42%	<b>Order picking</b> is the process that retrieves the SKUs and assembles the orders as defined by the pick lists. (NRL200101)
<i>size</i>	lot size	15.76%	From Table 3, we notice that the optimal <b>lot size</b> and shortage period are increasing highly with the higher values of set-up cost. (OL190103)
<i>chain</i>	cold chain (logistics)	0.56%	Although resilience of <b>cold chain</b> is a poorly understood concept among managers in the citrus industry, ... (LM180307)
<i>risk</i>	risk mitigation	11.51%	..., which are crucial in <b>risk mitigation</b> and prevention planning. (PDL210202)

*LAWEC-BAWE shared senses and register-specific collocations (D)*

Word sense comparison between two academic corpora found shared senses (D) which have different percentages in occurrence between these academic contexts. For example, the sense of *lead* “to cause something to happen” is a LAWEC-BAWE shared sense, accounting for 27.45% occurrence in the LAWEC. In the BAWE, it accounted for 95.44% occurrence. I had decided, in section 4.3, that if the occurrence of a sense in one corpus is 1.5 times higher or lower than that in the other corpus, it will be considered as a LAWEC-BAWE shared sense (D). The sense of *lead* above is such a case, due to its large difference of occurrence in two academic corpora.

I identified 115 LAWEC-BAWE shared senses (D). They accounted for 24.36% of all 472 senses identified in the LAWEC. Of these, 102 were categorized according to part of speech, 10 were general list collocations and 3 were disciplinary collocations. Table 5.34 and 5.35 presents the number of each category and LoSAWEL levels. On the aspect of LoSAWEL levels, all shared senses (D) were from Level 1-3, except one disciplinary collocation was from Level 4. These show that general high frequency words have more different use in word senses across academic contexts, and that words with lower general frequency may be used consistently between academic contexts. On the aspect of part of speech categories, the shared senses (D) were mostly nouns ( $n = 61$ ), verbs ( $n = 22$ ) and adjectives ( $n = 16$ ), but few in adverb ( $n = 2$ ) and preposition senses ( $n = 1$ ). This suggests that adverbs and prepositions were mostly used in the same way between academic corpora, which can also be seen in Table 5.34 where most of adverbs and preposition senses were LAWEC-BAWE shared senses (S). For register-specific collocations, Table 5.36 shows that although general list collocations were mostly shared between corpora, nearly half of them (10 in 22) had different use in occurrence. However, the disciplinary collocations were mostly specific to the LAWEC, and few were shared between two academic corpora. These suggest different ways of expressing meaning through collocations in academic contexts. An example will be provided for each part of speech and register-specific collocation category respectively to show different use of word senses in two academic corpora.

Table 5.34 Part of speech of LAWEC-BAWE shared senses (D)

Level	noun	verb	adjective	adverb	preposition	total
Level 1	18	10	4	2	0	34
Level 2	28	12	7	0	1	48
Level 3	15	0	5	0	0	20
Level 4	0	0	0	0	0	0
total	61	22	16	2	1	102

Table 5.35 LAWEC-BAWE shared register-specific collocations (D)

Level	Register-specific Collocations ( <i>ph.</i> )	Register-specific Collocations ( <i>tm.</i> )	total
Level 1	10	0	10
Level 2	0	2	2
Level 3	0	0	0
Level 4	0	1	1
total	10	3	13

Table 5.36 Word sense classification in part of speech and register-specific collocation categories in LAWEC-BAWE word sense comparison

	Number of Senses only Identified in the LAWEC	Number of LAWEC-BAWE Shared Senses with Different Percentage in citations	Number of LAWEC-BAWE Shared Senses with Similar Percentage in citations	Total
noun	31	61	106	198
verb	23	22	69	114
adjective	16	16	50	82
adverb	0	2	13	15
preposition	1	1	2	4
general list collocations	4	10	8	22
disciplinary collocations	32	3	2	37
total	107	115	250	472

### *knowledge*

A noun sense of *knowledge* “information or facts of a particular subject” was identified as a LAWEC-BAWE shared sense (D). In the LAWEC, this sense accounted for 51.23% of occurrence. However, this is the only sense of *knowledge* identified in the BAWE in the current selection criteria. This is because, in the LAWEC, *knowledge* was more likely to be used in a general list collocation “to the best of our knowledge”. For example,

In summary, to the best of our *knowledge*, our paper is the first to study how the manufacturer can use the online channel entrance decision as a strategy ... (TR181208)

It can be the reason that LAWEC collected experienced writers' articles where they were more confident to make claims. The BAWE mainly collected students' writing whose authors were not as experienced as those in the LAWEC on academic writing expressions.

### *lead*

*lead* had two senses identified in the LAWEC, and four in the BAWE. Its verb sense "to cause something to happen" (usually used in "lead to") was a LAWEC-BAWE shared sense (D). This sense accounted for 27.45% and 95.44% of occurrence in the LAWEC and the BAWE respectively, showing a big difference of use in occurrence. This is because, in the LAWEC, *lead* was more likely used in a disciplinary collocation "lead time(s)". In the BAWE, this verb sense was in dominance in occurrence, although *lead* had three other senses identified. The difference of senses of *lead* can also be observed through collocations (+/-5) in two corpora, which is presented in Table 5.37 below. In the LAWEC, *lead* typically collocated with "times", "time", "commitment" and frequently modified by "shorter", "higher" or "short".

... to optimize the total cost by considering the trade-off of having more/less inventory or longer/shorter commitment *lead time*. (NRL190301)

In the BAWE, *lead* frequently collocated with modal verbs like "could", "may", "can" and "might", which can be traced in lines following:

Mintzberg points out 'too much planning may *lead* us to chaos, but so too would too little, more directly' (Boddy 2002:187). (750)

If interest rates increase it will directly affect a households disposable income, the increase could *lead* to lower consumer spending in the short term. (1868)

Table 5.37 Collocates of *lead* in the LAWEC and the BAWE, ordered by LogDice

Rank	LAWEC				BAWE			
	Collocate	LogDice	Freq as Collocate	Freq in the LAWE	Collocate	LogDice	Freq as Collocate	Freq in the BAWE
1	times	10.88	235	2007	could	8.68	136	9254
2	time	10.14	497	12402	may	8.61	149	10818
3	may	9.81	277	8024	can	8.45	287	25202
4	commitment	9.59	78	1253	us	8.20	37	2470
5	shorter	9.24	43	4029	might	8.14	33	2197
6	higher	9.06	99	682	conclusion	8.12	23	1052
7	short	9.00	43	3091	increase	8.03	42	3617

### *particular*

In the LAWEC, *particular* was frequently observed in a general list collocation “in particular”. 99.54% of its occurrences had this sense. *particular* also had the other sense “to describe something that is special” identified in the LAWEC, which only accounted for 0.46% in concordance lines. For example,

In this *particular* case, economies of scale lead to lower transport costs, but lower transport costs help ports become more competitive, ... (MEL200402)

This adjective sense was also identified in the BAWE, but with much higher occurrences (37.90%). Two collocates “group” and “references” played a role for this sense in the BAWE. The LogDice value of these two collocates did not meet the threshold for word sense analysis in the LAWEC, which can be the reason that this adjective sense had low occurrences in the LAWEC.

### *further*

All three senses of *further* were LAWEC-BAWE shared senses, of which one showed different percentage in citations. The sense “to describe greater extent or degree” (*adv.*) is a LAWEC-BAWE shared sense (D), and was more likely to occur in the LAWEC (6.75%), compared to its occurrence in the BAWE (1.32%). In the LAWEC, this sense was observed in patterns like:

- a) “this (research) further found/revealed/explored/extends/investigates/” to show the significance of research,
- b) “to further explore/investigate..., this research...” to show the aim of research or analysis process,
- c) “research can further investigate/explore” for the aim of call for further studies.

These can be traced from following lines.

This line of research could also be *further* pursued. (PDLM200402)

To *further* validate the research hypotheses, data could be collected from other countries. Finally, additional research (LM210202)

Future research can investigate *further* how buying firms might benefit from collaborating with their suppliers in CMSC. (LM210301)

However, in the BAWE, *further* was more likely used to describe the extent or degree of actions. This shows different functions of use in context, although the sense was the same.

..., claiming they were more fragile and in need of extra help, which *further* excluded working-class women from medical treatment. (2)

Braudel did much to help *further* propagate Annales research, establishing the Centre de recherches historiques, and another dozen research teams. (588)

### *per*

The sense of *per* “to express rates and ratios” and was the only LAWEC-BAWE shared sense (D) in preposition. In the LAWEC, it accounted for 79.43% in occurrence, and in the BAWE it was for 52.74%. Collocates play a role here. In the LAWEC, *per* collocated with, for example, “unit”, “day”, “year”, “time”, which are usually in context of describing specific numbers. For example,

... BP is \$600 per tonne of fuel, F is 100 tonnes of fuel *per* day and TC is \$36,000 per day. (MEL210202)

So, an increase of one unit in the average case pack quantity would yield an additional decrease of 1.17% in profit *per* unit time. (BL180301)

In the BAWE, although those mentioned collocates still existed, they had lower LogDice value and less frequency. For example, “unit” was a collocate of *per* in both academic corpora. It was the top collocate of *per* in the LAWEC (LogDice = 12.27; Freq as collocate = 801), but ranked as the fifth in the BAWE (LogDice = 10.40; Freq as collocate = 120). Instead, in the BAWE, “cent” and “capita” had more occurrences. Table 5.38 presents the collocate list of *per* in two academic corpora (LogDice ≥ 9).

Table 5.38 Collocates of *per* in the LAWEC and the BAWE, ordered by LogDice

Rank	Collocate	LAWEC			BAWE			
		LogDice	Freq as Collocate	Freq in the LAWE	Collocate	LogDice	Freq as Collocate	Freq in the BAWE
1	unit	12.27	801	2193	cent	11.92	281	283
2	cent	11.78	376	376	capita	11.37	183	182
3	day	10.48	167	717	week	10.46	118	661
4	cost	10.37	539	10275	unit	10.40	120	831
5	year	10.34	160	931	year	10.29	169	2327
6	time	10.24	573	12402	day	9.91	97	1210
7	average	10.07	199	2947	average	9.67	84	1290
8	item	9.85	115	972	se	9.32	42	67
9	capita	9.73	83	95	share	9.26	61	1161
10	profit	9.70	154	2967	million	9.19	51	766
11	cycle	9.69	112	1327	annum	9.12	36	33
12	total	9.45	177	5205				
13	holding	9.34	77	780				
14	hour	9.05	54	218				

*in order to*

The general list collocation “in order to” was identified in both academic corpora, but with different percentage in citations. In the LAWEC, this collocation accounted for 37.68% in occurrence, but in the BAWE it was for 96.47%. This is because the collocates of *order* in the BAWE (LogDice  $\geq 9$ ) formed two patterns “in order to make” and “in order to achieve”, as shown in Table 5.39 and following lines below. Compared to the collocates in the BAWE, those in the LAWEC were more related to business activities, as shown in Table 5.40.

**In order to** make significant improvement in future, the following points should be considered to get more accurate result: ... (1340)

Williams explains that he has to be 'the social man', always concerned with the good of his people **in order to** achieve the destiny of Rome. (2707)

Table 5.39 Collocates of *order* in the BAWE (LogDice  $\geq 9$ )

Rank	Collocate	LogDice	Freq as Collocate
1	In	10.47	948
2	in	9.83	4,011
3	to	9.66	4,808
4	make	9.04	168
5	achieve	9.02	122

Table 5.40 Collocates of *order* in the LAWEC (LogDice  $\geq 9$ )

Rank	Collocate	LogDice	Freq as Collocate
1	quantity	11.03	516
2	In	9.96	728
3	fulfilment	9.80	182
4	picking	9.72	179
5	in	9.52	2,347
6	to	9.50	2,848
7	quantities	9.44	142
8	an	9.17	440
9	customer	9.13	190
10	optimal	9.12	198
11	online	9.06	146

### *chain*

*chain* was frequently observed in a disciplinary collocation in the LAWEC, for example, “supply chain” (and its variation, such as “supply chain management”, “supply chain network”, “supply chain system”, “supply chain members”, “supply chain partners”, “supply chain performance”). It refers to interconnected journey led by organizations to covert raw materials to products and then to its consumers. This sense accounted for 94.56% concordance lines checked. It was to be expected because supply chain management is a key topic in Logistics, which can be seen from sections of suggested topics of journal websites. This collocation was also identified in the BAWE, accounting for 25.91% in occurrence. It occurred in a wide range of texts from Agriculture, Business, Engineering (Electronic Business Management), Food Sciences, and Management. For example,

... that IKEA is heavily focussed on operations management and back office processes in terms of manufacture, supply *chain* management and distribution. (1098) [Business]

..., the data can be delivered directly from customers to business processing sections through supply *chain* management automatically, not from traditional manual information input with high operational risks ... (1282) [Engineering – Electronic Business Management]

This suggests that Logistics may have some relations with different disciplines, especially Business.

### 5.2.5 Word sense comparison between the LAWEC and the BNC1994 NAW

This section aims to answer the second sub-question of RQ2 *To what extent are the words identified used in ways specific to Logistics? b) is their use different from the general use context?* The comparison was made between word senses identified in the LAWEC (in section 5.2.1) and those in the BNC1994 NAW (in section 5.2.3). Similar to the comparison between in the LAWEC and the BAWE), it covers three aspects:

- (1) an overview of word sense comparison on number of word senses identified in each part of speech and register-specific collocation category;
- (2) word senses and register-specific collocations that are identified in the LAWEC but not meeting selection criteria in the BNC1994 NAW;
- (3) word senses and register-specific collocations shared between the LAWEC and the BNC1994 NAW, but with much different percentages in citations (hereafter LAWEC-BNC shared senses/register-specific collocation (D); D represents “much different percentages in citations”).

#### *An overview of word sense comparison between the LAWEC and the BNC1994 NAW*

Word sense analysis shows that 472 and 562 senses were identified in the LAWEC and the BNC1994 NAW respectively. Table 5.41 shows the number of word senses identified at each LoSAWEL level in two corpora.

Table 5.41 Comparison on word senses identified in the LAWEC and the BNC1994 NAW at each LoSAWEL level

Level	Number of Items	Number of Senses in the LAWEC	Percentage of Senses in the LAWEC	Number of Senses in the BNC1994 NAW	Percentage of Senses in the BNC1994 NAW
Level 1	106	179	37.92%	256	45.55%
Level 2	127	194	41.10%	198	35.23%
Level 3	71	91	19.28%	99	17.62%
Level 4	7	8	1.69%	9	1.60%
Total	311	472	≈ 100.00%	562	100.00%

The comparison shows that, in the BNC1994 NAW, more word senses were identified at each LoSAWEL level, when compared to those in the LAWEC. This is expected because the BNC1994 NAW is a general use corpus, covering a wider range of contexts, when compared to the LAWEC, a discipline-specific corpus. The percentages of word senses at each level in two corpora show a difference. In the LAWEC, word senses at Level 2 had the largest proportion. However, in the BNC1994 NAW, word senses at Level 1 were the largest group, and there was a downward trend of percentage of word senses at Level 1-4. More senses were identified at levels with higher general use word frequency. This is reasonable, because the higher the general use frequency of an item, the more likely it could be a polysemous word.

Table 5.42 shows a comparison of number of word senses in each part of speech category between the LAWEC and the BNC1994 NAW. It shows that, in both corpora, the largest part of speech category is noun, and the smallest one is preposition. Except for prepositions whose number is the same in two corpora, more senses were identified in the other four categories in the BNC1994 NAW, compared to those in the LAWEC. Again, this is to be expected because the BNC1994 NAW is a general use corpus, covering a wider range of contexts.

Table 5.42 Comparison of word senses in part of speech categories in the LAWEC and the BNC1994 NAW

Part of Speech	Number of Senses in the LAWEC	Percentage of Senses in the LAWEC	Number of Senses in the BNC1994 NAW	Percentage of Senses in the BNC1994 NAW
noun	199	42.16%	253	45.02%
verb	114	24.15%	130	23.13%
adjective	82	17.37%	114	20.28%
adverb	15	3.18%	18	3.20%
preposition	4	0.85%	4	0.71%
Total	414	87.71%	519	92.35%

Table 5.43 Comparison of word senses in register-specific collocation categories in the LAWEC and the BNC1994 NAW

Part of Speech	Number in the LAWEC	Percentage of Senses in the LAWEC	Number in the BNC1994 NAW	Percentage of Senses in the BNC1994 NAW
General List Collocations	21	4.45%	42	7.47%
Disciplinary Collocations	37	7.84%	1	0.18%
Total	58	12.29%	43	7.65%

Table 5.43 shows a comparison of each register-specific collocation. It shows that fewer general list collocations (e.g., “in order to”) were identified in the LAWEC, compared to its counterpart in the BNC1994 NAW. Twice as many general list collocation were identified in the general use corpus. In contrast, more disciplinary collocations were identified in the LAWEC ( $n = 37$ ), and only one in the BNC1994 NAW (“market share”). This is to be expected, since disciplinary collocations, as a particular way of expression in discipline-specific contexts, seem unlikely to occur in general use contexts.

Three kinds of senses and collocations were found in comparison, with similar procedures in section 5.2.4: a) only in the LAWEC but not meeting selection criteria in the BNC1994 NAW, b) LAWEC-BNC shared senses (D): shared between the LAWEC and the BNC1994 NAW but with quite different percentages in citations, c) LAWEC-BNC shared senses (S): shared between two corpora with similar percentage in citations. The classification method I used was presented in section 4.3. Table 5.44 presents the number and percentage of three kinds of senses. This shows that less than half of senses identified in the LAWEC (49.79%,  $n = 235$ ) are LAWEC-BNC shared senses (S). The rest (50.21%,  $n = 237$ ) are different from general use contexts: 123 senses were unique in the LAWEC, and 114 were

LAWEC-BNC shared senses (D) (difference in percentages > 1.5 times). These 238 senses in the LAWEC will be discussed in the following subsections. Table 5.45 presents word sense classification in part of speech and register-specific collocation categories in LAWEC-BNC word sense comparison. For example, in the LAWEC, there were 198 noun senses identified. In the LAWEC-BNC comparison, 46 of 198 noun senses were only identified in the LAWEC but not meeting selection criteria in the BNC1994NAW, 59 were LAWEC-BNC shared senses (D), and the remaining 93 were LAWEC-BNC shared senses (S). This table also shows that two thirds of adverbs (n = 10) and three quarters of prepositions (n = 3) were LAWEC-BNC shared senses (S). This suggests that adverbs and prepositions were generally used consistently in both the LAWEC and the BNC1994 NAW. For other categories, less than half of senses were in this category, suggesting the different use of senses.

Table 5.44 Number of senses shared or not shared between the LAWEC and the BNC1994 NAW

	Senses only in the LAWEC but not meeting selection criteria in the BNC1994 NAW	LAWEC-BNC shared senses (D)	LAWEC-BNC shared senses (S)
Number of Senses	123	114	235
Percentage	26.06%	24.15%	49.78%

Table 5.45 Word sense classification in part of speech and register-specific collocation categories in LAWEC-BNC word sense comparison

	Number of Senses only Identified in the LAWEC	Number of LAWEC-BNC Shared Senses (D)	Number of LAWEC-BNC Shared Senses (S)	Total
noun	46	59	93	198
verb	26	22	66	114
adjective	9	20	53	82
adverb	0	5	10	15
preposition	0	1	3	4
general list collocations	6	7	8	21
disciplinary collocations	36	0	1	37
total	123	114	235	472

*Word senses found only in the LAWEC but not meeting selection criteria in the BNC1994 NAW*

123 senses were unique in the LAWEC, when compared to the BNC1994 NAW with current selection criteria applied. They accounted for 26.06% of all 472 senses identified in the LAWEC. 81 of 123 senses were classified according to their part of speech categories and LoSAWEL levels in Table 5.46. The remaining 42 were classified as register-specific collocations, which will be discussed in the later subsection.

Table 5.46 shows that all these senses were in noun, verb and adjective categories. There were no unique adverb and preposition senses identified in the LAWEC, suggesting these two categories were all shared senses between the LAWEC and the BNC1994 NAW. Moreover, regarding the LoSAWEL levels, Level 4 has only one unique sense, and most of unique senses were from Level 1-3. It indicates that general high frequency words have more different senses used in discipline-specific context, when compared to lower frequency words, because items at Level 1-3 are general high frequency words and Level 4 are mid frequency words. Table 5.47 presents 10 selected examples of these senses found only in the LAWEC. At each level, one example was chosen for noun, verb and adjective senses respectively.

Table 5.46 Part of speech and LoSAWEL levels of senses found only in the LAWEC but not meeting selection criteria in the BNC1994 NAW

Part of Speech	noun	verb	adjective	adverb	preposition	total
Level 1	19	8	4	0	0	31
Level 2	15	15	4	0	0	34
Level 3	11	3	1	0	0	15
Level 4	1	0	0	0	0	1
total	46	26	9	0	0	81

Table 5.47 Examples of senses in the LAWEC but not meeting selection criteria in the BNC1994 NAW

Level	Item	Sense/Use	Percentage in Citations	Concordance Line Extracted
Level 1	<i>area</i>	a piece of land ( <i>n.</i> )	97.14%	... a number of sources and receivers are to be deployed randomly with the goal of providing coverage for an <i>area</i> . (NRL200104)
Level 1	<i>address</i>	try to understand a problem or deal with it ( <i>v.</i> )	100%	Therefore, in this study, we <i>address</i> the following research questions (TR220713)
Level 1	<i>general</i>	to talk about something without much detail ( <i>adj.</i> )	30.13%	Many <i>general</i> theories are borrowed from other disciplines such as economics, marketing, and strategic management (Stock 1997). (BL190101)
Level 2	<i>environment</i>	human-made conditions ( <i>n.</i> ); in a collocation “built environment”	16.28%	... to interpret the relationship between the built <i>environment</i> and travel behavior more precisely. (TR191205)
Level 2	<i>measures</i>	to discover or judge how great the amount is ( <i>v.</i> )	2.28%	... includes the transportation cost performance, which <i>measures</i> the costs for purchasing transportation services for one year, including contract and on-demand sourcing. (OL220402)
Level 2	<i>per</i>	in a fixed collocation “per capita” ( <i>adj.</i> )	3.64%	... we grouped countries based on estimates of gross national income per capita. (MEL190301)
Level 3	<i>review</i>	formal explanation made by people in authority ( <i>n.</i> )	97.34%	According to Denyer and Tranfield (2009) and Rousseau et al. (2008), a systematic <i>review</i> requires pre-specified relevance and quality selection criteria.
Level 3	<i>defined</i>	to show, describe and state clearly what it is ( <i>v.</i> ); in collocations like “defined as”	100%	... this paper focuses on gearing, liquidity, profit, activity, cash flow and market ratios as <i>defined</i> in Table 1. (MEL190307)
Level 3	<i>negative</i>	to mean “no” ( <i>adj.</i> )	9.78%	This might produce a <i>negative</i> response, ... (LR200504)
Level 4	<i>remainder</i>	rest of writing ( <i>n.</i> )	100%	The <i>remainder</i> of this paper is organized as follows. (TR200928)

I shall discuss four examples from Table 5.47, one at each level., at Level 1, *address* had only one sense “try to understand a problem or deal with it” used in the LAWEC, through collocation analysis and concordance line checking. This is different from that in the general use context, where *address* is used to refer to “place” (percentage in citations: 100%). Here below are two extracts of *address* in the general use corpus BNC1994 NAW:

Would give your full name and *address*. Er, Michael David er currently stationed at. (320; spoken text)

At Level 2, *per* was found as a component of “per capita” which is used as an adjective in the LAWEC. It was not found in the BNC1994 NAW through collocation analysis with current selection criteria, since the collocate “capita” did not meet the threshold of LogDice ( $\geq 9$ ) in collocation analysis. Other examples are the nouns *review* and *remainder*, at Level 3 and 4. In the LAWEC, *review* had a sense “formal explanation written by the people in authority” (percentage in citations: 97.34%). This was found in collocations like “systematic review”, “research review” or “literature review”. In contrast, with current selection criteria, this item was found in collocations like “judicial review” referring to “a procedure which allows a civil court of law to review the validity of certain legislation” or “book review” which means “descriptions or analysis of new books” in the BNC1994 NAW for general use. The sense of “rest of writing” of *remainder* is the only sense at Level 4 not meeting selection criteria in the BNC1994 NAW. In the BNC1994 NAW, *remainder* was found as “something left of a group of objects” or “a kind of number used in mathematics”. Such differences show the focus of two corpora in comparison: the LAWEC as a discipline-specific corpus has unique senses more related to academic writing and abstract concepts. In contrast, the same items have different senses observed in the BNC1994 NAW, whose senses are close to real life.

*Register-specific collocations found only in the LAWEC but not meeting selection criteria in the BNC1994 NAW*

In section 5.2.1, 58 register-specific collocations were identified in the LAWEC, of which 21 were general list collocations and 37 were disciplinary collocations. Word sense comparison between the LAWEC and the BNC1994 NAW showed that 42 of 58 were found in the LAWEC only (see Table 5.48). This indicates that more than 70% register-specific collocations in the LAWEC were unique, when compared to those in the general use context. Of these 42 collocations, 6 were general list collocations, and 36 were disciplinary collocations. Nearly a third of general list collocations in the LAWEC were not found in the BNC1994 NAW with current selection criteria. Also, except for one (“market share”), all disciplinary collocations were unique in the LAWEC, when compared to those in the general use

context. Table 5.48 presents the number of register-specific collocations in the LAWEC, in the BNC1994 NAW, and those only in the LAWEC.

Table 5.48 Number of register-specific collocations in the LAWEC, BNC1994, and those only in the LAWEC but not meeting selection criteria in the BNC1994 NAW

Register-specific Collocations	LAWEC	BNC1994 NAW	Comparison: Only in the LAWEC but not meeting selection criteria in the BNC1994 NAW
Number of General List Collocations	21	42	6
Number of Disciplinary Collocations	37	1	36
Total	58	43	42

Table 5.49 presents 6 general list collocations found in the LAWEC but not meeting selection criteria in the BNC1994 NAW. They involved 5 items: *case*, *general*, *order*, *terms* and *knowledge*. *case* had two unique general list collocations not meeting selection criteria in the BNC1994 NAW, and the other four items had one respectively. Regarding the LoSAWEL levels of items, except for *knowledge* (“to the best of our knowledge”) at Level 2, the rest 5 were all from Level 1, and none was from Level 3 and 4. This indicates that, for general list collocations, difference exists at general high frequency levels, especially for the most frequent 2,000 words as components. Moreover, these collocations were close to general academic context, since 5 of them were also found in the BAWE, except for “a case in point”.

Table 5.49 Examples of general list collocations found in the LAWEC but not meeting selection criteria in the BNC1994 NAW

Item	Level	General List Collocations	Percentage in Citations	Concordance Line Extracted	Found in the BAWE
<i>case</i>	Level 1	a case in point	0.03%	Figure 4 demonstrates a <b>case in point</b> . Figure shows a CMCC instance encountered in iteration 5 of the Voronoi heuristic solving the 17-center problem,... (NRL220403)	No
<i>case</i>	Level 1	if this is the case	2.41%	<b>If this is the case</b> , the vehicle manufacturer may consider continuing vehicle production ... (NRL220101)	Yes
<i>general</i>	Level 1	in general	69.61%	The logit model allows classifying accurately 92.11% of cases <b>in general</b> ; the sensitivity of the model (i.e., the probability of classifying accurately the efficient containerships) is ... (MEL180202)	Yes
<i>order</i>	Level 1	in order to	37.68%	We analyzed multiple cases <b>in order to</b> provide a more robust and generalizable consensus (Yin, 2014). (LM200202)	Yes
<i>terms</i>	Level 1	in practical/general terms	0.75%	<b>In practical terms</b> , the results of this study should help shipowners and related companies to understand changes... (MEL190305)	Yes
<i>knowledge</i>	Level 2	to the best of our knowledge	48.77%	Specifically, <b>to the best of our knowledge</b> , only two studies have investigated new trade-in strategies, ... (TR200210)	Yes

Table 5.50 Disciplinary collocations in the LAWEC but not meeting selection criteria in the BNC1994 NAW

Item	Disciplinary Collocation	Percentage in Citations	Item	Disciplinary Collocation (Percentage in Citations)	Percentage in Citations
<i>case</i>	case pack	0.08%	<i>common</i>	common agency	9.64%
<i>control</i>	inventory control	30.54%	<i>common</i>	common method variance	7.52%
<i>degree</i>	degree centrality	16.40%	<i>decision</i>	decision variable	30.21%
<i>information</i>	asymmetric information	3.50%	<i>decision</i>	Markov decision chain/process	5.30%
<i>lead</i>	lead time	72.55%	<i>decision</i>	decision tree	4.62%
<i>market</i>	spot market	11.79%	<i>effect</i>	moderating effect	7.86%
<i>market</i>	market power	9.21%	<i>effect</i>	mediating effect	5.60%
<i>means</i>	means objectives	20.62%	<i>effect</i>	interaction effect	3.57%
<i>order</i>	order fulfilment	6.24%	<i>effect</i>	ripple effect	2.73%
<i>order</i>	order picking	5.42%	<i>effects</i>	moderating effects	3.84%
<i>problem</i>	problem instance	4.07%	<i>effects</i>	mediating effects	3.78%
<i>problem</i>	routing problem	3.27%	<i>range</i>	middle range theory	5.95%
<i>problems</i>	routing problems	11.81%	<i>risk</i>	risk management	51.61%
<i>lot</i>	lot size	15.76%	<i>risk</i>	risk mitigation	11.51%
<i>state</i>	state space	24.52%	<i>supply</i>	supply chain management	85.46%
<i>chain</i>	supply chain (management)	94.56%	<i>value</i>	value added	8.24%
<i>chain</i>	cold chain logistics	0.56%	<i>resources</i>	human resources	30.23%
<i>common</i>	common method bias	23.53%	<i>inventory</i>	holding inventory	6.56%

Regarding the disciplinary collocations, all were identified only in the LAWEC rather than in the BNC1994 NAW with current selection criteria, except for “market share” (n = 36). This is to be expected, since Logistics disciplinary collocations are frequent in a discipline-specific context, and less likely to occur in general use context. Table 5.50 above presents those 36 disciplinary collocations. These 36 disciplinary collocations involved 23 items. They show two points. First, an item may have more than

two disciplinary collocations identified. For example, *chain* had two disciplinary collocations found in the LAWEC but not in the BNC1994: “supply chain (management)” and “cold chain logistics”. Second, all these 23 items were polysemous words in the LAWEC, since none of disciplinary collocations had 100% percentage in citations. This further suggests that, in Logistics context, polysemous words may be as components of disciplinary collocations to express discipline-specific concepts. Polysemy is related to the RQ3 and will be discussed in section 5.3.

#### *LAWEC-BNC shared senses and register-specific collocations (D)*

Word sense comparison between the LAWEC and the BNC1994 NAW also found shared senses and collocations with much different percentages in citations: LAWEC-BNC shared senses and register-specific collocations (D). The criterion is consistent with that in section 5.2.4: if a word sense was identified in both the LAWEC and the BNC1994 NAW, and the difference between their percentages in citations is a larger than 1.5, this sense is considered as a LAWEC-BNC shared sense with much different percentages in citations.

The results show that 114 senses belong to such category. They accounted for 24.15% of all senses identified in the LAWEC (also see Table 5.44 above). Of these, 107 senses can be classified according to their part of speech, and 7 were related to register-specific collocations. I shall discuss senses in part of speech categories firstly. Table 5.51 present classification of the number of 107 senses in each part of speech category and each LoSAWEL level.

Table 5.51 Number of LAWEC-BNC shared senses (D) in each part of speech category and each LoSAWEL level

Part of Speech	noun	verb	adjective	adverb	preposition	total
Level 1	23	7	11	5	1	47
Level 2	23	14	7	0	0	44
Level 3	12	1	2	0	0	15
Level 4	1	0	0	0	0	1
total	59	22	20	5	1	107

Regarding the LoSAWEL levels, the table shows that only one shared sense was observed at Level 4 in comparison, and all others were from Level 1-3. It suggests that, the LAWEC-BNC shared senses (D) were mostly observed in general high frequency words, especially the most frequent 2,000 words. Regarding part of speech, the table shows that 101 of 107 senses were nouns, verbs or adjectives, and

the remaining 6 were into adverbs and prepositions. These further indicate that adverbs and prepositions were mostly used consistently across contexts. In contrast, nouns, verbs, and adjectives have more LAWEC-BNC shared senses (D). An example of senses was chosen from each part of speech at each level ( $n = 12$ ), namely, four nouns, three verbs, three adjectives, an adverb and a preposition. Table 5.52 presents these 12 examples, and their corresponding percentages in citations in two corpora.

Table 5.52 Examples of LAWEC-BNC shared senses (D)

Item	Sense Definition	Percentage in Citations in the LAWEC	Percentage in Citations in the BNC1994 NAW
<i>means</i>	to describe methods to do something ( <i>n.</i> )	0.34%	11.28%
<i>application</i>	software ( <i>n.</i> )	63.04%	99.67%
<i>abstract</i>	a section in research paper or writings ( <i>n.</i> )	100%	7.59%
<i>inventory</i>	to supply or stock ( <i>n.</i> )	93.44%	33.33%
<i>found</i>	see or learn something ( <i>v.</i> )	98.96%	1.17%
<i>applied</i>	to use rules/skills in an activity ( <i>v.</i> )	100%	34.30%
<i>decreases</i>	get smaller in amount ( <i>v.</i> )	99.53%	64.44%
<i>major</i>	to describe something important or serious ( <i>adj.</i> )	100%	1.25%
<i>positive</i>	to show something is evident ( <i>adj.</i> )	27.37%	8.12%
<i>negative</i>	to show something is not evident ( <i>adj.</i> )	0.86%	8.78%
<i>further</i>	to introduce related topics ( <i>adv.</i> )	6.75%	0.17%
<i>following</i>	after that event ( <i>prep.</i> )	2.27%	19.23%

The comparison also showed seven LAWEC-BNC shared register-specific collocations (D). They are all general list collocations, from Level 1 ( $n = 5$ ) and Level 2 ( $n = 2$ ). This indicates that, in the LAWEC-BNC comparison, general high frequency words, especially the most frequent 2,000 words, have more LAWEC-BNC shared senses (D), when compared to those at lower frequency levels. No disciplinary collocation was observed in this comparison, supporting the results in previous sub-section that

disciplinary collocations were less likely to frequently occur in general use context. These seven collocations were listed in Table 5.53.

Table 5.53 Examples of LAWEC-BNC1994 NAW shared register-specific collocations (D) – general list collocations

Item	General List Collocations	Percentage in Citations in the LAWEC	Percentage in Citations in the BNC1994 NAW
<i>case</i>	in case	0.27%	1.78%
<i>means</i>	(this) is by no means	0.34%	9.51%
<i>particular</i>	in particular	99.54%	10.59%
<i>terms</i>	in terms of	97.01%	28.26%
<i>used</i>	used to be	0.03%	23.13%
<i>example</i>	for example	42.80%	27.68%
<i>respect</i>	with respect to	100%	31.44%

I chose four examples from Table 5.52 and 5.53 intentionally for analysis. An example from nouns (*abstract*), verbs (*applied*), adjectives (*major*) and general list collocations (*in terms of*) was selected respectively. No specific examples were from adverbs and prepositions, because, in previous sections, the results showed that adverb and preposition senses were generally used consistently across two corpora.

### *abstract*

*abstract* was identified as a monosemous word in the LAWEC, since it had only one noun sense “a section of research papers or writings”. This sense was to be expected because the LAWEC is a disciplinary academic written corpus which collected academic journal articles. In contrast, in the BNC1994 NAW, *abstract* had two senses identified: the noun sense “a section of research papers or writings” had less percentage in citation (7.59%), and the adjective sense “about art” accounted for 92.41% in citations. This is because, in the BNC1994 NAW, *abstract* typically collocated with words related to art, for example, “Expressionist”, “Expressionism”, “ballets”, “paintings”, “Figureurative”, and “canvases”. Also, the BNC1994 NAW excluded academic written texts (marked with W\_ac\_), which can be the reason that the noun sense of *abstract* had much less percentage in citations. Here below are three extracts of *abstract*, of which the first was from the LAWEC for the noun sense, and the other two were from the BNC1994 for the noun and adjective sense respectively.

We thus excluded irrelevant articles by reading the title, *abstract* and full paper. (LM220212)  
[LAWEC, *abstract as a noun*]

In July, the Urgent Issues Task Force released a consensus clarifying the treatment of transfers from current assets to fixed assets. 'Of particular concern,' said the UITF in its *abstract*, 'is the possibility that companies could avoid charging the p&l account with write-downs to net realisable value arising on unsold trading assets.' (3467) [BNC1994 NAW, *abstract as a noun*]

It was preceded by 'The Interpretive Link: Abstract Surrealism into *Abstract* Expressionism 1938–1948', which he presented at the Newport Harbor Art Museum, where he was formerly chief curator, ... (3237) [BNC1994 NAW, *abstract as an adjective*]

### *applied*

*applied* had a verb sense “to use rules/skills in an activity” identified as a shared sense in two corpora. This sense had much different percentages in citations in two corpora: 100% in the LAWEC as a monosemous word and 34.30% in the BNC1994 NAW as a polysemous word. The difference also can be observed from collocations. In the LAWEC, it collocated with “widely”, “been”, and “method” (LogDice  $\geq 9$ ). Here below is an extract from the LAWEC:

In this paper, Logit Model, which has been widely *applied* in various disciplines including transportation, finance and manufacturing, is used. (MEL190307)

In the BNC1994 NAW, collocations of *applied* in analysis formed different senses. For example, *applied* collocated with “Linguistics”, having a sense of “the subject or study has a practical use (*adj.*)” (26.09% in citations), or it collocated with “permission” with the sense of “to ask formally” (21.26% in citations). It also collocated with “principles” to have the shared verb sense “to use rules/skills in an activity” (34.30% in citations). Here below is an extract of *applied* from the BNC1994 NAW for each sense:

(1) *applied* (“to use rules/skills in an activity”, verb, 34.30% in citations)

If the 'polluter pays' and preventative principles were genuinely *applied*, there would, as the memo fears, be 'major agricultural and related consequences'. (1373)

(2) *applied* (“the subject or study has a practical use”, adjective, 26.09% in citations)

... in areas such as Environmental, Labour, Industrial, Development, International and Monetary Economics and *Applied* Econometrics. (405)

(3) *applied* (“to ask formally”, verb, 21.26% in citations)

That is true, but if somebody *applied* for planning permission tomorrow or a or a renewal of planning permission, ... (921)

(4) *applied* (“to be relevant or suitable to someone or something”, especially in “apply to”, verb, 18.36%)

The fact that the rational expectations hypothesis is a free-standing set of criteria which can be *applied* to a wide variety of macroeconomic models was rarely grasped by early Keynesian critics.

(45)

### *major*

*major* was also identified as a monosemous word in the LAWEC, with an adjective sense “to describe something important”. It collocated with and was used to modify “incident” (LogDice: 9.17) and “concern” (LogDice: 9.05). In contrast, this sense was much less frequent in the BNC1994 NAW, only accounting for 1.25% in citations. Also, around 0.60% in citations in the BNC1994 NAW showed the noun sense of “an officer in the British Army”, for example, “Major Francis John William Harvey” or “Major John MacBride”, two military leaders in 19<sup>th</sup> – 20<sup>th</sup> in British history. Most of concordances of *major* in the BNC1994 NAW (98.15% in citations) were about “John Major”, the former Prime Minister of the UK, when the BNC1994 was built. It could be argued that the distinguish of *major* as an adjective, a noun or in a surname would not cause much difficulty when encountering this item. However, in this research, I insisted to split these senses to show the real use of *major* in the BNC1994 NAW. Such difference also indicates the importance of recency of texts in corpus, which affects word sense in use. This is one of the limitations for using the BNC1994 to represent the general use context, which will be discussed further in Chapter 7.

### *in terms of*

The general list collocation “in terms of” accounted for 97.01% concordances in the LAWEC. However, this collocation had 28.26% in citations in the BNC1994 NAW, with 3.43 times of difference of percentage in citations > 1.5 times as threshold. Such difference can be the reason that “in terms of” is more related to academic use. This collocation was double checked in the BAWE, where it accounted for 72.06% in citations (also see Appendix B). The percentage in citations of “in terms of” in the LAWEC and the BAWE showed that this collocation is more frequent in the academic context, when compared to that in the BNC1994 NAW. This also shows that academic contexts may have different preference on general list collocations when compared to general use context.

### 5.2.6 Summary for word sense identification and comparison in the LAWEC, the BAWE, and the BNC1994 NAW

This section aims to answer the RQ2 *To what extent are the words identified used in ways specific to Logistics? a) is their use different from the general academic context? and b) is their use different from the general use context?* The LAWEC, the BAWE, and the BNC1994 NAW were selected to represent Logistics disciplinary academic context, general academic context, and general use context respectively. Based on procedures of word sense identification, word senses of 311 LoSAWEL items were identified in three corpora, presented in sections 5.2.1-3. Among three corpora, the LAWEC had the smallest number of word sense identified ( $n = 472$ ), the BNC had the largest number ( $n = 562$ ), and the BAWE had 503 identified. Two rounds of word sense comparison were conducted: LAWEC-BAWE comparison (section 5.2.4) and LAWEC-BNC comparison (section 5.2.5). I classified three categories of senses in word sense comparison: 1) senses found only in the LAWEC but not meeting criteria in two other corpora, 2) shared senses (D) referring to senses identified in two corpora in comparison but have different percentages in citations, and 3) shared senses (S) referring to senses identified in two corpora in comparison and have similar percentage in citations. The number of senses were presented in Table 5.54 below.

Table 5.54 Two rounds of word sense comparison

Word Sense Comparison	Senses Found Only in the LAWEC but not meeting selection criteria in other corpora (Unique Senses)	Shared Senses (D)	Shared Senses (S)
LAWEC-BAWE Comparison	107 (22.67%)	115 (24.36%)	250 (52.97%)
LAWEC-BNC Comparison	123 (26.06%)	114 (24.15%)	235 (49.78%)

The table shows that more than half of senses were shared senses (S) in the LAWEC-BAWE comparison (52.97%), and less than half of senses were in this category in the LAWEC-BNC comparison (49.78%). This suggests that word senses in the LAWEC had more differences in the BNC 1994 NAW when compared to those in the BAWE. However, in general, around half of senses in the LAWEC had different use in two corpora for comparison.

The results also showed three main findings. First, LAWEC unique senses and shared senses (D) were mainly observed at Level 1-3 in the LAWEC-BAWE comparison but at Level 1 and 2 in the

LAWEC-BNC comparison. This suggests that general high frequency words have more different use in different context, when compared to lower frequency words. Items at Level 4 (general mid frequency words) were mostly used consistently across three corpora. Second, regarding part of speech of senses, adverbs and prepositions were mainly used consistently across corpora, and the main differences of word senses were observed in nouns, verbs and adjectives. Third, regarding register-specific collocations, more than half of general list collocations and more than 85% of disciplinary collocations identified in the LAWEC were unique to the LAWEC, or had different percentages in occurrence, when compared to those in the BAWE and the BNC1994 NAW. This was especially true for disciplinary collocations, as 32 of 37 were not observed in the BAWE, and 36 were not identified in the BNC1994 NAW. This indicates the differences of register-specific collocations in specific contexts.

## 5.3 Polysemy in the Logistics Academic Written Context

### 5.3.1 polysemous LoSAWEL words in the LAWEC

The section 5.3 aims to answer RQ3 *What is the extent and nature of polysemy in the words identified?* In section 5.2.1, Table 5.13 presents the number of word senses identified for each LoSAWEL items in the LAWEC. It shows that 311 LoSAWEL items had 472 senses identified, and, of these, there were 206 monosemous words and 105 polysemous words. Findings of previous section suggest:

- (1) polysemous words account for 35.76% (n = 105) of the LoSAWEL items, more than a third in proportion;
- (2) LoSAWEL polysemous words generated more than half of senses identified in the LAWEC (56.36%; n = 266).

This section has a special focus on these 105 polysemous items, and their senses identified in the LAWEC.

Table 5.55 presents the number of polysemous words of the LoSAWEL, classified by the number of senses identified. It shows that most polysemous words had two senses identified in the LAWEC (n = 72). Fewer polysemous words had many senses, for example, only five polysemous words had more than five senses identified in the LAWEC.

Table 5.55 polysemous words of the LoSAWEL classified by the number of senses identified.

Number of Senses	2 senses	3 senses	4 senses	5 senses	6 senses	7 senses	Total
Number of Items	72	18	10	3	1	1	105

Table 5.56 presents an example of LoSAWEL polysemous word in the LAWEC with examples. This table shows that *means* was identified as a polysemous word in the LAWEC, having four senses. Two of senses were expressed through register-specific collocations: a general list collocation “this (is) by no means” and a disciplinary collocation “means objectives”. The other two senses included a verb sense of “indicate” and a noun sense of “ways”. The percentage in citations show that the verb sense of “indicate” is the most frequent one among four, accounting for 78.69% in citations. The sense expressed through general list collocation and that of “ways” were the least frequent, only for 0.34% in citations.

Table 5.56 An example of LoSAWEL polysemous word in the LAWEC

Item	Sense Definition	Percentage in Citations	example	Number of Senses
<i>means</i>	indicate (v.)	78.69%	This <b>means</b> ICOMP simultaneously measures model fit ... (BL180104)	4
	means objectives (tm.)	20.62%	Materials and ordering lists of the <b>means objectives</b> from the meetings then formed the basis .... (MEL200303)	
	this (is) by no means (ph.)	0.34%	<b>This is by no means</b> surprising given that the role of executives .... (BL200405)	
	way (n.)	0.34%	... take fuller account of government regulation as a <b>means</b> of enhancing their overall efficacy. (LM200103)	

### 5.3.2 LoSAWEL polysemous words and general use word frequency

105 polysemous words were then checked with their levels of the LoSAWEL. Table 5.57 presents the number of polysemous words at each LoSAWEL level. The table shows that, for example, Level 1 had the largest number of polysemous words identified, accounting for 41.51% of all words at this level. The table also shows two points. First, there is a downward trend on percentage of polysemous words at each level. At Level 1, polysemous words accounted for more than 40% of items. The percentage decreases to 33.86% at Level 2, 23.94% at Level 3, and less than 15% at Level 4. This suggests that the higher general use word frequency a word has, the more likely it can be a polysemous word. When the general use word frequency goes lower, it is more likely for an item to be a monosemous word. Figure 5.2 presents the percentage of monosemous words and polysemous words at each LoSAWEL level, and shows the downward trend of polysemous words. Second, most of polysemous words (n = 104) were identified at Level 1-3 which belong to general high frequency words, and more than 80% of

polysemous words ( $n = 87$ ) were identified at Level 1 and 2, related to the most frequent 2,000 words. This indicates that general high frequency words are the main source of polysemous words in the Logistics academic written context, especially for the most frequent 2,000 words.

Table 5.57 Number of polysemous words at each LoSAWEL level

LoSAWEL Level	Number of LoSAWEL Polysemous Words	Number of LoSAWEL Words	Percentage of Polysemous Words at this Level
Level 1	44	106	41.51%
Level 2	43	127	33.86%
Level 3	17	71	23.94%
Level 4	1	7	14.29%
Total	105	311	n/a

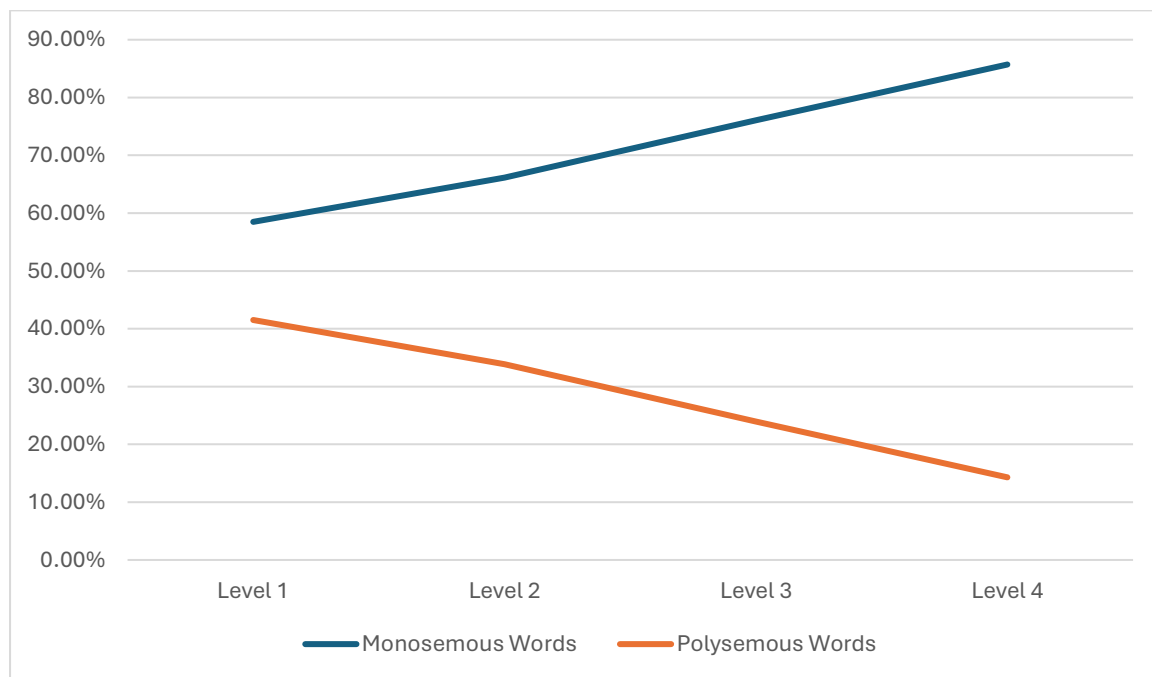


Figure 5.2 Percentage of monosemous and polysemous words at each LoSAWEL level

Table 5.58 below presents a polysemous word example for each LoSAWEL level respectively. Items, word sense definitions, percentage in citations were given in the table, with an example for each sense. For example, at Level 1 *lead* was observed to have two senses, one expressed through a disciplinary collocation “lead time(s)” and the other as a verb sense “to cause something to happen”. The sense expressed through disciplinary collocation “lead time(s)” is more frequent in citations, accounting for

72.55% citations in the LAWEC, while the verb sense is less frequent in the LAWEC (27.45% in citations). Polysemous word examples at other three levels are:

Level 2 *account* (sense 1: a general list collocation “take ... into account”; sense 2: bank account),

Level 3 *function* (sense 1: formula; sense 2: a particular use of something), and

Level 4 *inventory* (sense 1: supply or stock; sense 2: a disciplinary collocation “holding inventory”).

Table 5.58 examples of polysemous words at each LoSAWEL level

Level	Item	Word Sense	Percentage in Citations	Extract of Concordance
Level 1	<i>lead</i>	in a fixed collocation “lead time(s)” ( <i>tm.</i> )	72.55%	The firms in the garment manufacturing industry confront tight product schedules and short <i>lead times</i> . (ML200211)
		to cause something to happen (v.) (lead to)	27.45%	Hence, over-commitment may <i>lead</i> to difficulties in managing expectations. (LM220307)
Level 2	<i>account</i>	in a fixed collocation “take ... into account” ( <i>ph.</i> )	99.91%	Although the decision model for product consolidation developed in this paper <i>takes into account</i> the practical concerns as much as possible, ... (LM180305)
		a type of bank account ( <i>n.</i> ) (interest bearing account)	0.09%	... , then the retailer deposits the sales revenue into an interest bearing <i>account</i> during the time interval ... (OL180402)
Level 3	<i>function</i>	formula ( <i>n.</i> )	99.32%	The retailer's expected cost <i>function</i> is ... (BL180103)
		a particular use of something ( <i>n.</i> )	0.68%	Does resilience imply retaining the predisturbance structure and <i>function</i> , or does it refer to the ability of a system to change structure and function in response to a shock? (BL210301)
Level 4	<i>inventory</i>	supply or stock ( <i>n.</i> )	93.44%	... we prepare material on demand forecasting and <i>inventory</i> management and, (BL180302)
		in a fixed collocation “holding inventory” ( <i>tm.</i> )	6.56%	... thereby reducing the overhead <i>holding inventory</i> cost (OL200404)

### 5.3.3 LoSAWEL polysemous words, part of speech and register-specific collocations

#### *LoSAWEL polysemous word senses and part of speech categories*

As mentioned above, in the LoSAWEL there are 105 polysemous words, with 266 senses identified in the LAWEC. These senses were then classified by part of speech and register-specific collocation categories. Of these, 212 were classified by part of speech (see Table 5.59), and 54 were expressed through register-specific collocations which will be discussed later. The percentage of polysemous word senses in each part of speech category is also presented in Table 5.42, to explore the relationship between polysemous words and part of speech categories.

Table 5.59 Polysemous word senses and part of speech categories

Part of Speech Category	Number of Senses	Number of Senses Related to Polysemous Words	Percentage of Polysemous Word Senses in this Part of Speech Category
Noun	198	121	61.11%
Verb	114	49	42.98%
Adjective	82	38	46.34%
Adverb	15	2	13.33%
Preposition	4	2	50.00%
Total	413	212	51.33%

Table 5.59 shows that 121 of 198 noun senses (61.11%) were related to polysemous words. This further suggests that, in the Logistics academic written context, polysemous words are most likely to be related to nouns. In contrast, polysemy is much less frequently found among adverbs, since only two senses related to polysemous words were observed. These two senses actually involved only one item *further* (see Appendix D). One adverb sense of *further* is “to introduce related topics” (6.75% in citations), and the other is “greater extent or degree” (6.75% in citations). See extracts below:

(1) *further* “greater extent or degree”

..., and also at least mentioned social considerations, which could be captured by evaluating the improvement of standards of living and quality of life in the communities around the supply chain, and Chaabane et al. (2011, 2012), who *further* developed this research. (BL180102)

(2) *further* “to introduce related topics”

*Further*, this research work considers country-based cases, that is India. However, the critical SCRs and the strategies to (PDLM220201)

Regarding verb and adjective senses, in both of categories, more than 40% of senses were related to polysemous words. However, regarding prepositions, although 50% of preposition senses ( $n = 2$ ) were relevant to polysemous words, involving two LoSAWEL items, the overall number is quite small and did not play a significant role in the whole LoSAWEL list. Therefore, in the LAWEC, polysemous words are more likely to be nouns, verbs and adjectives, but infrequently found in adverbs and prepositions.

### *LoSAWEL polysemous word senses and register-specific collocations*

In section 5.2.1, it was found that 59 of 472 senses were expressed through register-specific collocations in the LAWEC, via either general list collocations (e.g., “in terms of”) or disciplinary collocations (e.g., “lead time(s)”). Of these register-specific collocations, 54 of 59 were related to polysemous words, and they involved 33 LoSAWEL items. This means that, of all 311 LoSAWEL items (monosemous and polysemous words), 10.61% were polysemous words having senses expressed through register-specific collocations. Among 105 LoSAWEL polysemous words, they accounted for 31.43%. It suggests that register-specific collocations can be an indication of polysemy. This is to be expected since a register-specific collocation is recognized as a whole multi-word unit for meaning expression, whose sense is different from that of its components.

Among 54 polysemous word senses occurring in register-specific collocations, 17 were general list collocations and 37 were disciplinary collocations. All disciplinary collocations identified in the LAWEC were involved in. This is also to be expected, since disciplinary collocations convey disciplinary concepts whose meanings are quite different from individual word meaning. I believe that these concepts of meanings will be difficult for learners or novices of Logistics field, and should be distinguished. This can be shown in the example of *lead* mentioned in previous sections, which has a sense expressed through a disciplinary collocation “lead time(s)” and the other sense of “cause something to happen” (e.g., “lead to”). Other examples like *case* (“case pack”), *means* (“means objectives”), *size* (“lot size”), *chain* (“cold chain”) were discussed in section 5.2. I shall use two other examples (*set* and *state*) to show the close relations between register-specific collocations and polysemous word senses.

*set* was identified as a polysemous word having four senses in the LAWEC (see Table 5.60). Three collocates met collocation analysis threshold ( $\text{LogDice} \geq 9$ ): “data” ( $\text{LogDice} = 9.47$ ;  $\text{FREQ} = 282$ ), “up” ( $\text{LogDice} = 9.38$ ;  $\text{FREQ} = 135$ ), and “a” ( $\text{LogDice} = 9.09$ ;  $\text{FREQ} = 1543$ ), whose concordance

lines were checked. Among the four senses, one occurred in a general list collocation “set up”, accounting for 22.62% of concordance lines.

Table 5.60 Senses of *set* identified in the LAWEC

Item	Definition	Percentage in Citations	Raw Frequency	Concordance Line Extracted	Number of Senses
<i>set</i>	in a fixed collocation “data set” ( <i>n.</i> )	42.55%	602	..., and data for direct distances was collected from the CEPII gravity data <i>set</i> . (MEL200306)	4
	refer to a group of people/things ( <i>n.</i> )	30.03%	426	Current SCs fail to engage a wider <i>set</i> of actors and institutions. (LM220301)	
	to describe “to create something”; in a fixed collocation “set up” ( <i>ph.</i> )	22.62%	320	To explore SCI's ability to differentiate between the two categories of firms (sustainable, or not), we <i>set up</i> an econometric analysis that would allow us to regress the SCI dimensions against the firm type. (BL180104)	
	to decide price/date/goal or else ( <i>v.</i> )	4.80%	69	The customer demand was measured by the average number of packages delivered per node (denoted <i>d</i> , which is <i>set</i> at 2 for a high demand level and 1.5 for a low demand level). (BL200302)	

*state*, was identified as a polysemous word having two senses in the LAWEC. One is “to describe the condition of something/someone (*n.*)”. This sense was observed in 75.48% of concordance lines ( $n = 385$ ), and in collocations like “the current state of”, “the state of system”, “transition ... from state ... to ... state ...”, “(in) steady state”, and “an/the out-of-control state” (Table 5.61).

..., instead of a diffusion limit, to approximate continuous time Markov chains in steady *state*. (NRL180102)

Section 4 develops a model of production system stop producing in the out-of-control *state* and its solution. (OL200403)

Figure 2 summarizes the current *state* of outsourcing in humanitarian logistics in a holistic framework. 4.1 (PDLM200402)

Table 5.61 Collocates of *state* in the LAWEC

Collocate	LogDice	Frequency
space	10.33	126
steady	10.05	76
out-of-control	9.56	52
transition	9.21	47
current	9.20	176
system	9.08	134

The other sense of *state* occurs in a disciplinary collocation “state space” which is a mathematical model of control engineering. It accounted for 24.52% of concordance lines, and had no relations with senses of “state” or “space”; therefore, it should be distinguished. Examples are as follows:

A second advantage of this approach, is that the part of the *state space* where the optimal policy is known, increases. (NRL180804)

In this case, the *state space* of the system becomes: Using standard results of Markov decision theory (see Ch. 6 in Ross, 1992) ... (OL210207)

This sense is not found in the BAWE or the BNC1994 NAW.

### 5.3.4 Polysemy categories in the LAWEC

As mentioned in section 4.2 on methodology of word sense analysis, collocates and concordances are essential to distinguish senses. By adopting the classification frameworks of Deignan et al (2023) and L’Homme (2021), I classified 105 LoSAWEL polysemous items in the LAWEC into three categories, based on the ways of word senses being different from each other. They are (1) nuances of senses, (2) meaning differences, and (3) lexical-grammatical use. The first category is about LAWEC polysemous words whose senses have slight difference, and the second contains polysemous words whose senses have big difference in meaning. The third category is also expected, since associations between meaning, form and lexical-grammatical use have been observed in corpus linguistics (e.g., Conrad, 2022; Greaves and Warren, 2022; Oakey, 2022). Examples of each category were given below.

### *Nuances of senses*

This category includes polysemous words whose senses may be distinguished by “Splitters” splitting senses finely (e.g., Gries, 2019; Skoufaki and Petrić, 2021), but not by “Lumpers” (Gries, 2019). In this category, senses of a polysemous word refer to the same core reference, but they may cause difficulty for learners or novices who are unfamiliar with specific contexts. This category is the largest one among three. Of 105 LoSAWEL polysemous words in the LAWEC, 63 belong to it, accounting for 60% of all. Examples selected are *ability*, *high*, *step*, *quality*, and *transportation*, presented in Table 5.62-66 below. For example, both senses of *step* have a core reference of a series of something. They had differences in collocates. The sense of “a series of actions” can be distinguished through collocates like numbers. Specifically, it was often in phrases like “Step 2” or “Step 7”, showing a step-by-step guidance in analysis or research. The other sense “stages of development” was often observed in structure like “the next step is to...”, referring to further stages of research development.

Table 5.62 Word senses of *ability* in the LAWEC

Item	Sense Definition	Percentage in Citations	Raw Frequency	Extracted Concordance	Number of Senses
<i>ability</i>	the fact that you can do something ( <i>n.</i> )	92.04%	108	..., and generally refers to a port's <b>ability</b> to optimise the allocation of various resources, industrial progress, the meeting of social needs, and other factors ... (LR190504)	2
	the quality or skills make something possible ( <i>n.</i> )	7.96%	10	Knowledge management capability KMC refers to the dynamic <b>ability</b> of the process of knowledge transferring, knowledge sharing, knowledge acquisition, knowledge application and (ML220401)	

Table 5.63 Word senses of *high* in the LAWEC

Item	Sense Definition	Percentage in Citations	Raw Frequency	Extracted Concordance	Number of Senses
<i>high</i>	greater amount and intensity ( <i>adj.</i> )	47.15%	857	... only in a limited number of cells where the network density or the demand level is <b>high</b> ). (BL200302)	3
	good in quality ( <i>adj.</i> )	28.39%	516	Real-time demand and supply expectations need to be covered with efficient processes at a <b>high</b> level of quality. (BL210108)	
	in priority ( <i>adj.</i> )	24.45%	444	..., it is to meet sustainable outcomes, meaning that the support chains have to take a <b>high</b> priority in the scheme of the business in order to deliver the appropriate value to the appropriate place. (BL180102)	

Table 5.64 Word senses of *step* in the LAWEC

Item	Sense Definition	Percentage in Citations	Raw Frequency	Extracted Concordance	Number of Senses
<i>step</i>	a series of actions ( <i>n.</i> )	80.38%	624	Step 2 If all containers are loaded into the ship, go to <b>Step</b> 7; otherwise, go to Step 3. (NRL190404)	2
	stages of development ( <i>n.</i> )	19.62%	152	these data and transferring all relevant traceability information with the pro- duct to the next stage or processing <b>step</b> . (LR221101)	

Table 5.65 Word senses of *quality* in the LAWEC

Item	Sense Definition	Percentage in Citations	Raw Frequency	Extracted Concordance	Number of Senses
<i>quality</i>	to describe how good or bad something is (n.)	60.84%	2132	In China, researchers also found that selling food products of poor <b>quality</b> could tarnish the reputation of manufacturers and reduce business sales. (LM200106)	2
	to refer to “standards” (n.)	39.16%	1373	Real-time demand and supply expectations need to be covered with efficient processes at a high level of <b>quality</b> . (BL210108)	

Table 5.66 Word senses of *transportation* in the LAWEC

Item	Sense Definition	Percentage in Citations	Raw Frequency	Extracted Concordance	Number of Senses
<i>transportation</i>	a system of transportation (n.)	84.88%	1158	..., for example, order retrieval and warehouse <b>transportation</b> systems. (BL190404)	2
	vehicles (n.)	15.12%	106	..., decreasing fill-rates and an increase of road freight <b>transportation</b> in general. (LM200206)	

### Meaning differences

The second category of polysemy contains items whose senses are quite distinguishable even by “Lumpers” (e.g., Gries, 2019). 30 of 105 LoSAWEL polysemous items belong to this, accounting for 28.57% of all polysemous items. Examples of polysemy belonging to such category are like *area*, *case*, *Figure*, *means*, *paper*, *order*, *rate*, *state*, *terms*, *lead*, and *table* which all have been discussed in previous sections. Polysemous items having senses expressed through disciplinary collocations also belong to this category. This is because disciplinary collocations express disciplinary concepts with which learners or novices may not be familiar. Here below I chose three other examples: *shows*, *value*, *function*, presented in Table 5.67 – 5.69. For example, the first three senses of *value* show differences of meaning. This item is most likely to have the sense

of “how much money it is worth”, referring to the prices or money in business trade. It may also occur with the sense of “a particular number”, which can be observed through “0.4” in the extract given in Table 5.68. It may also refer to “something is of helpfulness or importance”, which is related to the originality of research or significance of study in the given extract context.

Table 5.67 Word senses of *shows* in the LAWEC

Item	Sense Definition	Percentage in Citations	Raw Frequency	Extracted Concordance	Number of Senses
<i>shows</i>	to present something (v.)	99.97%	2084	Table 6 <i>shows</i> that if the transported volume is analyzed separately for each selected transport mode, ... (TR220716)	2
	a performance (n.)	0.03%	6	That is, increased publicity created by traditional media will expose viewers to the product (in this case, the TV <i>shows</i> ) and create curiosity toward it, which may eventually lead them to watch it. (ML220107)	

Table 5.68 Word senses of *value* in the LAWEC

Item	Sense Definition	Percentage in Citations	Raw Frequency	Extracted Concordance	Number of Senses
<i>value</i>	how much money it is worth ( <i>n.</i> )	48.69%	1425	..., and brand value is the total <b>value</b> of a brand as a distinct asset on a balance sheet (Wood et al., 2000). (ML190501)	4
	a particular number ( <i>adj.</i> )	30.23%	885	..., does predict as accurately as the optimal version when the cut-off <b>value</b> is tuned down to 0.4 (point E). (MEL190307)	
	something is of usefulness of importance ( <i>adj.</i> )	12.83%	375	Originality/ <b>value</b> – Although dynamic capabilities have been studied intensively in the last two decades, ... (LM180406)	
	in a fixed collocation “value added” ( <i>tm.</i> )	8.24%	242	In terms of <b>value</b> added services, Shi et al. (2016) study third-party purchase (3PP) as a value-added service and ... (LR200402)	

Table 5.69 Word senses of *function* in the LAWEC

Item	Sense Definition	Percentage in Citations	Raw Frequency	Extracted Concordance	Number of Senses
<i>function</i>	formula ( <i>n.</i> )	99.32%	2392	In the numerical study, we use the cost <b>function</b> as ..., which is commonly used in the literature (Baiman et al. (2000) ... (NRL220104)	2
	a particular use of something ( <i>n.</i> )	0.68%	16	Thanks to rapid economic globalization and the development of modern logistics, the <b>function</b> of ports has evolved from a traditional waterway transit node to a complex logistics center, ... (MEL200405)	

*Lexical-grammatical use*

The third category of polysemy is the smallest one among three, with 12 of 105 polysemous items classified (11.43%). Polysemous items in this category usually (but not always) have senses expressed through general list collocations. When expressing meaning, items in general list collocations show different lexical-grammatical use, which helps in distinguishing senses. Some items have been discussed in previous sections, for example:

- (1) *further* (used at the beginning of a clause with comma to introduce related topics)
- (2) *set* (“set up”)
- (3) *account* (“take ... account into”)
- (4) *knowledge* (“to the best of our/my knowledge”)
- (5) *per* (“per capita”).

I intentionally chose another example *based*, presented in Table 5.70 below. *based* was observed most frequently in “(be) based on” (82.91% in citations). It may also be combined with other nouns, in the form of “-based” (14.74%), or be a verb (past tense) (2.35%).

Table 5.70 Word senses of *based* in the LAWEC

Item	Sense Definition	Percentage in Citations	Raw Frequency	Extracted Concordance	Number of Senses
<i>based</i>	in a fixed collocation “(be) based on” ( <i>adj.</i> )	82.91%	1698	Nevertheless, <b><i>based</i></b> on the context studied here, the cases served to illustrate how the distinction of ... (BL180102)	3
	combined with nouns, in the form of “-based” ( <i>adj.</i> )	14.74%	302	Sodhi (2015) proposes the SRBV framework that extends the resource- <b><i>based</i></b> view (RBV) (Wernerfelt 1984; Barney 2001) framework to explain ... (BL190203)	
	the first thing develops from the second ( <i>v.</i> )	2.35%	49	... such as the researches of Ntayi et al. (2013) on Sub-Saharan African countries who <b><i>based</i></b> their approach on transformational leadership theory and the research of Morris et al. (2006) on South African ... (LM190101)	

### 5.3.5 Summary of polysemy in the LAWEC

Section 5.3 aims to answer RQ3 *What is the extent and nature of polysemy in the words identified?*. It focuses on LoSAWEL polysemous items and their senses identified in the disciplinary-specific context – the LAWEC. More than a third of LoSAWEL items were identified as polysemous words (35.76%;  $n = 105$ ), and more than half of senses identified in the LAWEC were related to these polysemous items (56.36%;  $n = 266$ ). Regarding the nature of polysemy in the words identified, five points were observed.

First, the likelihood of identifying polysemous items goes lower, when the general use word frequency of items goes lower. In other words, it is less likely for an item to be a polysemous word, if it is at lower general use word frequency bands, and vice versa. Second, most polysemous items were identified at LoSAWEL Level 1-3, and more than 80% of polysemous items were at Level 1-2 ( $n = 87$ ). This suggests that general use high frequency words, especially the most frequent 2,000 words were the main source of LoSAWEL polysemous items. Third, senses of polysemous items were mostly classified as nouns, verbs and adjectives, few related to adverbs and prepositions, in which only three items and four senses were involved. Fourth, register-specific collocations can be an indication of polysemy, more than a third of LoSAWEL polysemous items having senses expressed through register-specific

collocations. Fifth, polysemous items in the LAWEC showed preferences to specific senses whose percentage in citations can be more than 70%, but other senses identified in the same item played a less significant role in occurrences. 68 of 105 LoSAWEL items (64.76%) showed such preference. Moreover, three categories of polysemous words were classified: nuances of senses, meaning differences, and lexical-grammatical use, of which nuances of senses were the most frequent in the LAWEC. This suggests that polysemy in Logistics academic written context may cause difficulty for learners or novices who are unfamiliar with disciplinary context, because these senses need to be split through “Splitters” (e.g., Greis, 2019).

## 5.4 Summary of Results

This chapter presents findings of three research questions in three sections. The LoSAWEL contains 311 individual Logistics specialised word types, selected through keyness, range, dispersion and lexical word criteria. This specialised vocabulary list was divided into four levels based on general use word frequency levels, where Level 1-3 are for the general high frequency words, and Level 4 is for the rest. Most of word types (97.75%; n = 304) are from general high frequency words.

Word sense analysis of 311 word types was conducted in three corpora respectively: the LAWEC, the BAWE representing general academic context, and the BNC1994 NAW representing general use context. Word sense comparison between the LAWEC and the BAWE showed that more than half of senses identified in the LAWEC (52.97%; n = 250) were also identified in the BAWE, with similar percentage of occurrence in citations (LAWEC-BAWE shared senses (S)). 107 senses were only found in the LAWEC, and 115 were identified in both corpora but with different percentage of occurrence in citations (LAWEC-BAWE shared senses (D)). Word sense comparison between the LAWEC and the BNC1994 NAW showed that less than half of senses identified in the LAWEC (49.78%; n = 235) were also identified in the BNC1994 NAW, with similar percentage of occurrence in citations (LAWEC-BNC1994 NAW shared senses (S)). 123 senses were only found in the LAWEC but not meeting selection criteria in two other corpora, and 114 were identified in both corpora but with different percentage of occurrence in citations (LAWEC-BNC 1994 NAW shared senses (D)). Generally, the LAWEC and the BAWE had more shared senses with similar percentage of occurrence in citations. Adverb and preposition senses were generally used consistently across three corpora, but nouns, verbs and adjectives had different senses. Most different senses were observed at Level 1-3.

This research also finds that more than a third of LoSAWEL items (n = 105) were polysemous words in the LAWEC, and more than half of senses in the LAWEC were involved in (56.36%; n = 266). Polysemy in the LAWEC is closely related to general high frequency words (Level 1-3). Senses of nouns, verbs and adjectives are more likely to be related to polysemous words.

## Chapter 6 Discussion

### 6.0 Introduction

This chapter discusses each research question against the broader background of research:

Section 6.1 – RQ1: *What are the most frequently-used and evenly-distributed specialised Logistics words?*

Section 6.2 – RQ2: *To what extent are the words identified used in ways specific to Logistics?*

(a) *is their use different from the general academic context?*

(b) *is their use different from general English?*

Section 6.3 – RQ3: *What is the extent and nature of polysemy in the words identified?*

### 6.1 Specialised vocabulary in the academic written Logistics English context

#### 6.1.1 Specialised vocabulary across vocabulary categories and general use frequency levels

The specialised vocabulary list LoSAWEL was developed based on a series of quantitative word selection criteria including keyword analysis, dispersion, and range, where 311 word types were selected (see Appendix B for the full list). Two lexical features, as described in section 5.2, were observed from the analysis on these word types: one is some of words conveying both academic and discipline-related meanings, and the other is word distribution across general use frequency levels.

My findings of the feature of words conveying both academic and discipline-specific meanings, is related to the register and text-type of the specialised Logistics corpus (LAWEC) in analysis. That words in use vary in accordance with the register where they occur has been agreed in studies from both vocabulary research (e.g., Coxhead, 2020; Nation, 2022; Martinez, 2022) and corpus linguistics (e.g., Hunston, 2022; Sinclair, 1991; Stubbs, 2001; Szudarski, 2017). Therefore, it is reasonable to see that the word types in the LoSAWEL include both academic vocabulary and technical words.

Some of words seem to be easily grouped as either academic vocabulary or technical vocabulary, just based on word meaning at the first glance. For example, the academic vocabulary includes types like *study*, *studies*, *based*, *case* and *keywords*. Words such as *business*, *company*, *companies*, *distribution* and *logistics* are likely to be technical in the Logistics field, although they have differences in terms of relatedness to this field. There are also some word types that are difficult to classify, without consideration of their meaning in specific contexts. The word type *review* in the LoSAWEL is an

example. This word occurs with higher frequency in academic contexts (233.08 per million in the BAWE) than in general English (72.97 per million in the BNC1994), since the academic texts, especially for academic articles, tend to have “literature review” as an independent section. “literature review” is also found as a collocation in the LAWEC, where “literature” is the top collocates of *review*, with 54.91% of occurrence in all concordance lines. *review* also collocates with “valence”, forming the collocation “review valence” with technical meaning of “describing tones of online comments on products”. Although it occurs relatively rarely (3.95%), this technical collocation related to disciplinary knowledge is well beyond the statistical cut-off in collocational analysis set for this thesis ( $\text{LogDice} \geq 9$ ). This collocation was not found in either the BAWE or the whole BNC1994. This suggests that the occurrence of this collocation is higher than chance in the LAWEC. Thus, it is difficult, in this context, to clearly classify words like *review* as an academic word or a word belonging to technical vocabulary. The analysis of word types in the LoSAWEL list supports Nation’s (2013, 2022) classification where both academic vocabulary and technical vocabulary are subsumed under the umbrella of specialised vocabulary. It also brings corpus-derived evidence to support the existence of a blurred boundary between academic and technical vocabulary, as claimed by Nation (2019). Moreover, word types like *review* having both academic and discipline-related senses suggest general academic vocabulary can have disciplinary meanings in a specific context, which is in accordance with Hyland and Tse’s (2007) perspective and is consistent with the need of specialised vocabulary list development. The word type analysis also shows that word form in isolation can cause ambiguity of meaning, but can be disambiguated through collocation and concordance reading, as suggested by Sinclair (1991), Stubbs (2001), and Hunston (2022). The help of collocation in meaning disambiguation will be discussed further in section 6.3 on polysemy.

The other lexical feature is the tendency for specialised vocabulary to spread across general use frequency bands. It is claimed that specialised vocabulary may not consist of lower frequency words only (Nation, 2013, 2022), but rather, they can come from any level of high-, mid-, or low-frequency words (*ibid*). The analysis of the LoSAWEL is consistent with this claim through referring to general use word frequency levels, Nation’s (2012) BNC/COCA 25 based word lists. This research follows Dang, Coxhead and Webb’s (2017) method, and classified the LoSAWEL into four levels, where Levels 1-3 contain word types appearing in the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> BNC/COCA 25 base word lists respectively, and the Level 4 is for the rest. Table 6.1 below presents some examples of word types at each level (see the full list in the Appendix B).

Table 6.1 Examples of specialised Logistics vocabulary at four levels

LoSAWEL Levels	Example Items	Number of Items	Reference to BNC/COCA 25 Bands	Classification in General Use
Level 1	study, studies	106	1 <sup>st</sup> 1,000 list	High-frequency vocabulary
Level 2	knowledge, environment	127	2 <sup>nd</sup> 1,000 list	
Level 3	abstract, respectively	71	3 <sup>rd</sup> 1,000 list	
Level 4	empirical, logistics	7	4 <sup>th</sup> – 25 <sup>th</sup> 1,000 lists	Mid-/low-frequency vocabulary

It is seen that a very large amount of word types is in Levels 1-3, general high frequency words. There also exists a relatively small number of words (2.25%;  $n=7$ ) from mid-/low-frequency levels. This shows that specialised vocabulary also spreads across general use frequency levels. The issue of everyday words having specialised senses in specific contexts will be discussed later.

By referring to BNC/COCA 25 base word lists, Level 4 word types of my vocabulary list are at 4<sup>th</sup>, 5<sup>th</sup>, 6<sup>th</sup> 1,000 bands of BNC/COCA lists. Vocabulary at these frequency bands is usually considered to be mid-frequency (Schmitt and Schmitt, 2020; Vilkaite-Lozdiene and Schmitt, 2020). This suggests that, in this research, no general low-frequency words (10<sup>th</sup> – 25<sup>th</sup> 1,000 bands) become specialised Logistics vocabulary. This finding is different from previous research on other disciplines or discourses, for example, Wang, Liang, and Ge (2008) on medical academic word list, Lu (2018) on Traditional Chinese Medicine and Benson and Coxhead (2022) on rugby spoken discourse. This was not expected, since, in their studies, general low-frequency vocabulary plays a role; however, there is none in my list. There are two possible reasons for not including low-frequency general use words in the LoSAWEL: word selection criteria in word list development, and disciplinary lexical variation. Regarding word selection criteria, Nation (2016) suggests that different selection criteria can lead to different vocabulary items included. The corpus size and decision of reference corpus may also affect specific cut-off points selection criteria. Studies have shown that the choice of reference corpus (Brezina, 2018) and the decision of specific cut-off points (Brezina, 2018; Coxhead, 2000; Nation, 2016; Dang et al., 2017) lead to different items selected. General low-frequency words may also be with low range in Logistics contexts or not evenly distributed across texts. Therefore, they are not included in my list. For example, a general low frequency word “impute” (10<sup>th</sup> 1,000 band in BNC/COCA 25 lists) in the AVL was not included in the LoSAWEL, because of its low frequency and rang (Freq = 5, range = 4), and only occurring in two journals. The second reason can be lexical variation across disciplines. It is widely accepted in linguistics that vocabulary use is different across disciplines, which is a part of disciplinary convention in discourse community (Coxhead et al., 2020; Dang and Webb, 2014; Nguy and Ha, 2022). Logistics may rely on less general low frequency vocabulary to convey disciplinary concepts, when

compared to other disciplinary fields, such as Medicine, Engineering or Chemistry. To the best of my knowledge, there is no existing research on lexical profile on Logistics. However, based on lexical profile on the LoSAWEL and other disciplinary academic vocabulary lists, it is likely that this discipline has less demand on using general use lower frequency words as specialised vocabulary to convey disciplinary knowledge. Instead, this discipline may rely heavily on general use high frequency words in context. This will be discussed in the next subsection.

### 6.1.2 Specialised vocabulary focussing on the most frequent 2000 general use vocabulary

The analysis of the LoSAWEL in section 5.1 (also see Table 6.1 above) shows that much specialised Logistics vocabulary (with the exception of seven word types) is from general use high frequency words (Level 1-3). This suggests the significance of high frequency vocabulary as a foundation in vocabulary (Nation, 2013; Schmitt and Schmitt, 2020; Vilkaite-Lozdiene and Schmitt, 2020), even in specific contexts. Interestingly, my results also show that the most frequent 2,000 general use vocabulary items (Level 1-2) play an especially significant role in the LoSAWEL. The significance of these items is shown on three aspects: large proportion in the LoSAWEL, contribution to academic or disciplinary key concepts, and polysemy. This subsection will discuss the first two aspects, with the third discussed in section 6.3.

As shown in section 5.1 and Table 6.1, 233 of 311 word types (75%+) are from Level 1 and 2. Because of being selected through keywords, dispersion and range, these word types indicate much of aboutness of the academic written Logistics context, and their high proportion in the LoSAWEL suggests the preference of using high-frequency words in the Logistics context. This also supports the perspective that the most frequent 2,000 general use vocabulary can be the basis of vocabulary acquisition, both in everyday language and specialised language (Nation, 2006, 2013).

These words also help to express to academic conventions and Logistics disciplinary key concepts. They often occur as a component of longer units and collocate with at least one of other word types included in the LoSAWEL. Regarding academic convention expressions, for example, *studies*, at Level 1, collocates with *existing*, a word type at Level 2. They together form semi-fixed units like “(in/of) (the) existing studies”, with a slightly different variations on adjectives modifying *existing*. I did not classify units like “(in/of) the existing studies” as general list collocations with (*ph.*) markers, because they are not listed as an entry in the COBUILD dictionary. Therefore, they are not listed in Appendix B. However, it is a phenomenon that LoSAWEL items to form longer units with academic convention functions. Here below are more examples (emphasis added):

Future research can further explore elements potentially improving TBL sustainability to fill gaps in *existing studies*, ... (LM180410)

The factors considered in our paper are not identical to those in the *existing studies*, so here too, our results may be slightly different. (MEL210305)

This co-selection of word types in the LoSAWEL also applies to conveying disciplinary key concepts. For example, word types like *cost* (Level 1), *per* (Level 1) and *unit* (Level 2) co-select with each other and form a disciplinary technical term “cost of per unit moved”, meaning “a calculation on money/time/energy spent on moving a unit of products”. It is also a semi-fixed disciplinary collocation, since some slightly different variations were observed, including “unit cost”, “cost of unit”, and “cost per unit”. However, these combinations of words were not marked as disciplinary collocations (*tm.*) in my research, since the IWAL discipline-specific dictionary and glossary does not provide entries for them.

Indeed, these units or combinations of LoSAWEL items may provide useful information in knowing disciplinary language or for discipline-specific language learning in academic contexts. However, they are beyond of my research scope, since my research mainly focuses on specialised individual word list development and polysemy analysis.

There are three possible reasons for the significance of the most frequent 2,000 vocabulary in the LoSAWEL. The first reason is the nature of these high frequency words. Studies (e.g., Gui, 2013; Dang and Webb, 2016; Nation, 2006; Vilkaite-Lozdiene and Schmitt, 2020; Schmitt and Schmitt, 2020) have shown that general high frequency words covers 70-90% tokens in all registers. Therefore, it is likely that some of these words meet all selection criteria and are also frequently-used and evenly-distributed specialised Logistics vocabulary. The second reason can be the method used in specialised vocabulary list development. The LoSAWEL was developed from scratch, based on no other ready-made word lists (e.g., GSL, NGSL or AWL), and made no assumption about its target users' general use vocabulary knowledge. When compared to other specialised vocabulary list research (e.g., Coxhead, 2000; Coxhead and Hirsh, 2007; Liu and Han, 2015; Wang et al., 2008; Yang, 2015), the LoSAWEL does not exclude the most general high frequency words as long as the word types meet selection criteria. This decision brings items belonging to the most frequent 2,000 words to the list. This is in accordance with the method used in studies where some items of the most frequent 2,000 words are recognized as specialised vocabulary (e.g., Coxhead and Demecheleer, 2018; Coxhead et al., 2016; Dang, 2018a, 2018b; Lei and Liu, 2016; Lu, 2018; Valipouri and Nassaji, 2013). The third reason is the flexible collocability of the most frequent 2,000 words. General high frequency words as components in academic convention presentation has been recognized in a variety of corpus-based studies (Ackerman and Chen, 2013; Simpson-Valch and Ellis, 2010). The collocational analysis of the word types in the LoSAWEL is consistent with previous research, showing the co-selection of these words in forming

semi-fixed academic units (also shown in the example *studies* above). Apart from the role of the most frequent 2,000 words in academic conventions, their strong collocability in conveying disciplinary knowledge has also been noticed in specific contexts. For example, Hu et al. (2013) observe that a large proportion of business keywords are actually general high frequency words (although no specific percentage or number of words provided), and they show a tendency of co-selection for conveying business key concepts. Ward (2007) finds that the general high frequency word “time” as a component collocates with other words for engineering technical multi-word units. The analysis on the LoSAWEL also shows this co-selection (e.g., the example of *cost*, *unit* and *per* above). All these suggest that the general high frequency words, especially for the most frequent 2,000 words are important components for academic or disciplinary conventions.

### 6.1.3 Interdisciplinary feature of Logistics specialised vocabulary

The comparison between two academic corpora also shows interdisciplinary feature of Logistics specialised vocabulary. I shall discuss it from two points: individual words and disciplinary collocations. Regarding individual words, some LoSAWEL items are more related to Business and Management, and some are related to Mathematics. For example, word senses of items like *business*, *company*, *companies*, *cost*, *price*, *service*, *market*, *management*, *risk* are related to Business and Management. Senses of items like *number*, *rate*, *function* (meaning “formula”), *model*, *system(s)*, *scale* are related to Mathematics or Engineering. These suggest that Logistics has knowledge shared with these disciplines, to a certain extent. Regarding disciplinary collocations, in the LAWEC, they should be listed in the discipline-specific dictionary and glossary IWLA (2018). In the BAWE, they should only occur in one source discipline texts in the BAWE rather than in other texts. The comparison shows that “human resources” in Management, “market share” in Business, “holding inventory” in Engineering identified in the BAWE also occurred in the LAWEC. This is in accordance with the disciplinary definition on Logistics that it is an interdiscipline closely related to disciplines like Business, Economics, Engineering, Management, Mathematics and Sociology (Swanson, Goel, Francisco and Stock, 2017). It is also in line with the scope of journals collected in the LAWEC. Some topics related to these disciplines were listed in suggested research areas on journal websites (e.g., *International Journal of Logistics Research and Applications*). Different from those scholars and professionals who define Logistics as an interdiscipline based on their professional knowledge and experience, word sense comparison in this research provides evidence to support this from linguistic perspective. In addition, Hyland and Tse (2007) claimed that vocabulary in specific discipline may reveal disciplinary epistemology. Individual words and disciplinary collocations mentioned above support their idea. Specifically, business- and management-related words indicate Logistics focuses on business activities; mathematics- and engineering-related words suggest quantitative and empirical research methods may be preferred in this field, which aims

to solve real-life problems through combined methods. These features are all mentioned in Delfman et al. (2010, p.5) that “the specific insight-based interest of logistics focuses on overcoming the boundaries of established application-oriented scientific disciplines [...] through the synergistic combination of the knowledge bases of these disciplines [...] The direct and sequence of this is that [...] business management logistics, for example, or engineering logistics.

## 6.2 Word sense analysis and comparison across corpus

### 6.2.1 Word form as a more suitable counting of unit for word sense analysis in specialised context

Unit of counting of words is one of the criteria for word list development (Dang, 2020; Nation, 2016) (also see details in Chapter 2, section 2.2). Word type was chosen for developing the LoSAWEL in this research. The pilot study of this thesis (see details in section 4.2) shows that not every member of a word family or lemma has exactly the same (or disciplinary) sense in use. The analysis of word sense for items in the LoSAEEL through collocation and concordance also support this. For example, both *case* and *cases* are included in the LoSAWEL as separate items, and they are members of the same lemma CASE. However, they behave differently in collocations. *case* conveys disciplinary knowledge in a disciplinary collocation “case pack” which is not observed in the use of *cases*. For word type *case*, it has more senses and use identified in the LAWEC (see details in Appendix B on word sense):

- (1) *case* as an example or situation (e.g., the case of...)
- (2) *case* as an academic research method (e.g., case study/studies)
- (3) *case* in general list collocations (e.g., if this/it is the case, in case, in case of, a case in point)
- (4) *case* as a component of disciplinary collocations (e.g., case pack).

In contrast, *cases* is a monosemous word in the LoSAWEL. Only the sense (1) was observed for *cases*. This word type is not used with sense (2) or (3) in forming academic-related units. It is neither used to convey disciplinary knowledge in sense (4). Further checking was also made to compare senses of LoSAWEL and non-LoSAWEL items which belong to the same lemma. For example, word type *objective* of the lemma OBJECTIVE is included in the LoSAWEL (at Level 3) but “objectives” is not in the list. “objectives” was used as “aims” in this context, which had no difference between that in the BAWE and the BNC1994NAW, when the same analysis criteria as those in the LAWEC applied. Only the disciplinary collocation *objective value* was found to have disciplinary-related concept (*value* is also

a type in the LoSAWEL). Another example is word type *lead* of lemma LEAD, where “lead time” is a disciplinary collocation with high recurrence and is a strong collocation in the Logistic field, but not other members like “lead” or “led”.

The above points indicate two issues. First, a specific word type may have a specialised sense in a specific context, which does not apply to all members in the same lemma or word family. For studying word sense in a specific context, word type can be a good starting point of searching or observation. This method of word counting has been widely applied to studies on specialised vocabulary list development or lexical analysis in technical discourses (Benson and Coxhead, 2022; Chung and Nation, 2004; Coxhead and Demecheleer, 2018; Coxhead et al., 2016, 2020; Drayton and Coxhead, 2023; Ha, 2015; Lu, 2018; Quero, 2015; Ward, 2009; Waston-Todd, 2017). It supports the claim that the word sense changes when the word form is changed (Hunston, 2022; Sinclair, 1991). Each word type has its own typical sense or use which is distinguished from other types, and such differences of sense can be specific in context, deserving checking through collocation and concordance. Moreover, focusing on specific word types rather than all members in the same lemma and word family can be more helpful in pedagogy. Not checking other word members from the same lemma and word family may lead to missing senses to some extent. However, the LoSAWEL word types were those frequently and evenly distributed across the whole LAWEC corpus. They were selected based on keyness, range and dispersion. This suggests that these types were more likely to be encountered frequently and evenly by the potential word list users, when compared to those other members in the same word family or lemma. Furthermore, not all members were used frequently and evenly, and convey disciplinary specific senses (see the examples of *case*, *objective*, and *lead* above). Knowing these selected word types and their senses can generate more benefits, set more specific goals in learning, but less learning burden for potential users, when compared to knowing all word members from the same lemma and word family. Second, it is shown that the words do not occur in isolation, but in combination. These word types worked as components of general list collocations and disciplinary collocations. Especially for disciplinary collocations, collocational relationship between components is exclusive and fixed. For example, *lead* only collocates with “time” or “times” to form the disciplinary collocation “lead time(s)”. No other inflections or derivations of *lead* was observed in this disciplinary collocation. This supports the idiom principle of Sinclair (1991) that the routine of word usage is more pervasive. It also echoes with the fixedness of technical multi-word unit studies on other specialised contexts, either in spoken discourses (e.g., Benson and Coxhead, 2023; Drayton and Coxhead, 2022), or written language (Lu, 2018), or in both modes (Coxhead et al., 2020).

### 6.2.2 Core academic vocabulary, senses and register-specific collocations

With the same procedures of word sense analysis, senses of 311 LoSAWEL items were checked in the LAWEC and the BAWE for word sense comparison. The results shows that 472 and 503 senses were identified in two corpora respectively. Of these, 52.97% (n = 250) were classified as LAWEC-BAWE shared senses (S). These are word senses occurring in both two corpora and having similar percentage of occurrence in citations. They suggest the potential existence of core academic vocabulary, senses and register-specific collocations. Word sense comparison between two academic corpora, and a double-check with the academic written subcorpus of BNC1994 show much evidence on this. For example, the sense “chart (n.)” of *table*, and the sense of “difficult (adj.)” of *complex* (see details in section 5.2), were identified in all three different academic written corpora (e.g., *table* as “chart” has probability of occurrence of 99.99% in the LAWEC, 91.47% in the BAWE and 100% in the academic written subcorpus of the BNC1994 respectively). Other examples are *method*, *models*, *findings*, *purpose*, *research*, *types*, whose senses are consistent in two academic corpora, with similar percentage of occurrence. Evidence can also be observed through some general list collocations, such as *example* (“for example”), *instance* (“for instance”), *generally* (“generally speaking”), *compared* (“when compared to/with”), and *order* (“in order to”). These show that there are some words and collocations shared across disciplines. This supports the aims of corpus-derived general academic individual vocabulary and multi-word unit list development to select a group of lexical items that can benefit students from different discipline backgrounds in academic context. This is in line with research basis of studies like Ackerman and Chen (2013), Coxhead (2000), Gardner and Davies (2014), Lei and Liu (2018). Compared to their studies focusing on word form, this research provides evidence through word sense comparisons to answer the question raised by Hyland and Tse (2007) on core academic vocabulary. Indeed, there potentially exists a group of core academic vocabulary, both individual words and register-specific collocations, whose form and senses share across different disciplines.

### 6.2.3 Word senses varying in different registers

The LAWEC-BAWE comparison shows 22.67% (n = 107) senses were only found in the LAWEC but not meeting selection criteria in the BAWE. In the LAWEC-BNC comparison, 26.06% (n = 123) senses were only found in the LAWEC rather than the BNC with current selection criteria. The difference shows word senses varying in different registers, supporting Hyland and Tse’s (2007) claims. In my research, one of the examples is *lead*, which has 11 senses identified in three corpora: two in the LAWEC, four in the BAWE, and five in the BNC. However, only one sense (“to cause something to happen”) occurred in all three corpora, and this sense had a big difference in percentage of occurrence in the LAWEC and the BAWE. Table 6.2 presents senses of *lead* identified in three corpora.

Table 6.2 Word sense of *lead* in three corpora

LAWEC		BAWE		BNC1994	
Sense	Occurrence in Citations	Sense	Occurrence in Citations	Sense	Occurrence in Citations
in a fixed collocation “lead time(s)” ( <i>tm.</i> )	72.55%				
to cause something to happen ( <i>v.</i> ); in a fixed collocation “lead to”	27.45%	to cause something to happen ( <i>v.</i> ); in a fixed collocation “lead to”	95.44%	to cause something to happen ( <i>v.</i> ); in a fixed collocation “lead to”	36.83%
		to introduce to a topic ( <i>v.</i> )	3.13%		
		to take something in charge ( <i>v.</i> )	1.00%		
		a chemical component ( <i>n.</i> )	0.43%	a chemical component ( <i>n.</i> )	12.65%
				to guide someone ( <i>v.</i> )	2.15%
				as the key/first one ( <i>n.</i> )	25.48%
				in a fixed collocation “take the lead” ( <i>ph.</i> )	22.89%

Other examples include state (disciplinary collocation “state space” in the LAWEC; “country (*n.*)” and “to write in a formal way (*v.*)” in the BAWE and the BNC) and nature (“feature (*n.*)” in the LAWEC, “environment (*n.*)” in the BAWE and the BNC). These show two points. First, it shows disciplinary difference of word senses in specific contexts. This is in line with the research basis of corpus-derived discipline-specific vocabulary list development – vocabulary varies in disciplines (e.g., Liu and Han, 2015; Wang, Liang, Ge, 2008; Yang, 2015). The disciplinary examples also support Hyland and Tse’s (2007) claim that teachers need to guide students in acquisition of different vocabulary use in discipline-specific contexts.

Second, register may also be a factor affecting word senses and word use (Nation, 2022). For example, “to lead to a topic” was only identified in the BAWE, but not in the LAWEC or the BNC. This is mainly because of the difference of register in three corpora. The BAWE consists of proficient students’ writings (e.g., essays, case reports) (Heuboeck et al., 2010), in which students may use this word sense to introduce a new topic. The LAWEC collects peer-reviewed published research journal articles from experts, which may have fewer specific requirements on structures but on research design and information conveying (Hyland, 2021). Another sense of *lead* as “a chemical component” was identified in the BAWE and the BNC but not in the LAWEC. This is because *lead* as a chemical component is used in daily life, for example, lead-free petrol, and Chemistry-related disciplines are covered in the BAWE. The word senses reflect word use in contexts. The difference of word senses shown by influence of register support Nation’s (2013, 2022) framework of vocabulary knowledge, where register of vocabulary is subsumed into the part of vocabulary use knowledge.

## 6.3 Polysemy in Academic Written Logistics English Context

### 6.3.1 Polysemy and specialised Logistics vocabulary

In Chapter 3 I reviewed corpus studies on polysemy, and claimed that there is a relatively limited focus on polysemy in an academic or discipline-specific English context, with a few exceptions (e.g., Geng and Liang, 2025; Le and Miller, 2023; Skoufaki and Petrić, 2021; Wang and Nation, 2004).

In section 5.3, word sense analysis shows that polysemy is also pervasive in academic written Logistics contexts. 105 LoSAWEL word types were identified as polysemous words (35.76% in percentage), and they involved 266 senses in the LAWEC (56.36%). Such proportion of polysemous words identified among the word types in the LoSAWEL suggests that polysemous words are of high occurrence in specialised Logistics vocabulary, rather than an occasional phenomenon. This is in accordance with previous studies. In general academic contexts, Skoufaki and Petrić (2021) identified 34.48% AVL lemmas as polysemous lemmas through two general use dictionary resources. In Medical English, with the help of general use dictionary checking and native-speakers’ intuition in evaluation, more than 60% of word family headwords in the Medical Word List developed by Hsu (2013) were identified as polysemous words. Although these studies focus on different contexts with different results generated, the pervasiveness of polysemous words in academic or discipline-specific contexts are recognized.

However, the proportion of polysemous words in each context is different. There are three factors that may lead to differences in studies. The first is the word counting unit in word selection for word sense analysis. Different word counting units contain different numbers of word forms, with word type the smallest unit and the word family the biggest (Nation, 2013). Word type, the smallest one, was

selected as the unit of counting of my research in developing the LoSAWEL. The results show that this unit of counting is appropriate for my research, since it captures that discipline-specific senses have a strong tendency to be linked to a specific word type. This means that word type is more sensitive to discipline-specific sense identification in a specialised context (e.g., see examples of *case* and *lead* in section 6.2). In other words, using a larger word counting unit may lead to differences in sense identification and polysemy analysis.

The second factor is the resources used for word sense checking. Popular general use dictionaries can be good reference resources to check word senses (Kilgariff, 1992), for example, the Oxford English Dictionary (OED) and COBUILD dictionary used in those discipline-specific studies on polysemy. However, general use dictionaries have relatively limited resources for discipline-specific senses (Pantel and Lin, 2002). Disciplinary collocations of Logistics (e.g., *case pack*) are not listed as entries of general use dictionary. Instead, they are listed as entries in the Logistics specialised dictionary and glossary, IWAL (2018). This indicates that, without referencing to discipline-specific dictionaries, the number of polysemous words and word senses in a specific context may be underestimated.

The third factor is the methods of word sense analysis used in this research for identifying polysemous words. Dictionary checking is often used to distinguish senses (Kilgarriff, 1992, 2006; Skoufaki and Petrić, 2021), and users' intuition is needed to have decisions on which sense in entries fits the context best when ambiguity occurs (Dorst and Reijnierse, 2015). Corpus data and dictionary can be supplementary to each other, but the former is not widely applied to discipline-specific studies on polysemy (also see Chapter 3). Moreover, how to split senses also make differences. In this research, I followed Deignan and Love (2021) and Deignan et al. (2023) who adopted splitting narrowly, which is different from that of computational algorithms (e.g., Geng and Liang, 2024) or sense relatedness scale for homonyms (e.g., Le and Miller, 2023; Wang and Nation, 2004). Hence, different ways of splitting senses may also influence the number of word senses identified and that of polysemous words in a specific context.

### 6.3.2 Polysemy and general use high frequency words in discipline-specific contexts

In section 5.1, the analysis on LoSAWEL word types shows that a very large part of specialised Logistics vocabulary (97.75%) is general use high frequency words classified at Level 1-3, and around 75% belong to the most frequent 2,000 words. Apart from the high coverage of the general use high frequency words in the LoSAWEL, polysemy is another aspect of significance of these words. Of 105 polysemous word types identified in the LAWEC, 104 belong to general use high frequency words, and more than 80% of polysemous word types ( $n = 87$ ) are from the most frequent 2,000 general use words

(see section 5.3, Table 5.40 and Figure 5.2). Moreover, it is observed that there is a downward trend of likeness of polysemous words across general use word frequency levels. These findings suggest two points. First, general use high frequency words, especially the most frequent 2,000 words, are likely to be polysemous words, even in specialised contexts. Second, the higher general use word frequency a specific word type has, the more likely it is to be a polysemous word (Sinclair, 1991). Similar findings were also obtained in word sense analysis in general academic and everyday language (see section 5.2). A similar decreasing trend of likeliness of polysemous word existence across frequency levels was observed in these two contexts.

The findings is in accordance with previous research in everyday language (Kuiper et al., 2018; Sinclair, 1991). In general academic context, Skoufaki and Petrić (2021) found that more than 2/3 polysemous words are from the most frequent 1,000 AVL lemmas. The AVL was developed without ready-made word lists and therefore it also contains general use high frequency words which frequently recur in general academic contexts. Although there is no major study on lexical profile of items in the AVL and their general use word frequency, it is estimated that at least 70% of the most frequent 1,000 AVL lemmas are from general high frequency words, and 50% are from the most frequent 2,000 general use words. This manifests a strong relation between polysemy and general use high frequency words, regardless of being in everyday language or general academic contexts. This relation has also been noticed in other research studies in discipline-specific contexts, though fewer evidence was provided. For example, Hu et al. (2019) noticed that there exists a relation among polysemy, general use word frequency and Business English key words, but the study did not cover what extent the Business English key words are polysemous words. In medical field, Le and Miller (2023) reported that 72% polysemous words come from BNC 4<sup>th</sup> – 9<sup>th</sup> bands, classified as mid-frequency words. However, their research was based on Hsu's (2013) Medical Word List which excludes the most frequent 3,000 word families in BNC (general high frequency words). Therefore, it is not clear whether and to what extent polysemy has a certain relation with general use high frequency words in Medical English. This calls for more discipline-specific research on polysemy and frequency-based corpus-derived word lists.

Two reasons may explain why general use high frequency words tend to be polysemous words, even they are in specialised contexts. First, vocabulary development of language plays a role. It is claimed that people use familiar words for different uses in expressing meaning, rather than coining new words every time, due to the economy of expression (Falkum, 2015) and the law of least effort (Deutschman, 2021). Second and more importantly, it is the development of new disciplines and knowledge. The communication needs for new disciplines and knowledge boosts the change of vocabulary and word senses, where new senses are added to existing words or new words are coined (Gui, 2013). However, working out how general use high frequency words become polysemous words in Logistics contexts is beyond the scope of this research, but the analysis and discussion suggest that general use high frequency words in specialised context have more complicated and varied word senses in use when

compared to words at mid-, or low-frequency levels. This also suggests that, in specialised contexts, general high frequency words are of importance in discipline-specific vocabulary teaching and learning, especially for their polysemous word feature, which will be discussed further in implication section later.

### 6.3.3 Polysemous word senses, part of speech and register-specific collocations

Word sense analysis in section 5.3 shows polysemy is more frequently found in nouns, verbs, and adjectives, but infrequently in adverbs and prepositions. Of 105 polysemous LoSAWEL items, only four types were related to adverbs and prepositions. This finding is in accordance with Skoufaki and Petrić (2021) on polysemous words of the AVL in the general academic context. However, no other discipline-specific study discusses the relation between polysemous words and part of speech, which calls for more research for comparison.

Of 472 senses identified in the LAWEC, 59 occurred in register-specific collocations (12.50%): 22 general list collocations and 37 disciplinary collocations. 54 of 59 register-specific collocations were related to polysemous words, and 33 types were involved in. These 33 word types accounted for 10.61% of the whole 311-word LoSAWEL, and 31.43% of all 105 polysemous words. This suggests that register-specific collocations and polysemy are also in high occurrence in the LAWEC. Also, in cross-corpus comparison, most register-specific collocations in the LAWEC were not identified in the BAWE or the BNC, especially for disciplinary collocations. Specifically, 32 of 37 disciplinary collocations in the LAWEC were not found in the BAWE, and none of them were found in the BNC. It supports Dorset and Reijnierse (2015) that general use dictionaries may not particularly list technical senses of an item.

In general academic or discipline-specific contexts, the senses of register-specific collocations are mostly discussed in studies on multi-word units, for example, collocations (Coxhead, 2019), technical multi-word units (Benson and Coxhead, 2022; Drayton and Coxhead, 2023; Lu, 2018). However, no major research discusses polysemous word senses and register-specific collocations in detail in general academic or discipline-specific contexts. One of the reasons is the different ways of splitting senses according to differences of research purposes, as discussed in section 6.2.1. For example, Skoufaki and Petrić (2021) and Le and Miller (2023) focused on individual specialised words and their senses. For specialised words working as components of register-specific collocations and expressing meaning together with other components, they did not take it into consideration. Another possible reason is not referencing to discipline-specific dictionaries as a key resource for word sense analysis in previous research. Results in section 5.3 and discussion in this chapter have shown that about 12.5% senses identified in the LAWEC corpus were listed in the Logistics discipline-specific dictionary and glossary

IAWL (2018), rather than the COBUILD dictionary. This suggests that discipline-specific dictionaries can be a powerful supplementary to general use dictionaries when in analysis of word senses in a specialised context. Without referencing to discipline-specific dictionaries, senses, especially occurring in register-specific collocations, related to disciplinary conventions or key concepts that are different from general academia or everyday languages could be missed.

Moreover, senses occurring in register-specific collocations show that these senses are not linked to individual words, but to adjacent collocations within a span. This can be observed through collocational analysis and concordance line reading, instead of relying on dictionaries or researchers' intuition only. For example, *means* collocated with "objectives" and "objectives" was one of the top collocates of *means*. Concordance line checking further ensures that the collocation "means objectives" is beyond chance in the LAWEC corpus and should be recognized as a whole chunk because of the disciplinary concept it conveys. Through these, collocational analysis helps with extraction of frequently recurred register-specific collocations within a span, and concordances provide for further contextual information on senses. Therefore, the word senses are analysed within the contexts rather than in decontextualization when referencing to dictionaries where insufficient example sentences are provided.

Collocational analysis and concordance checking for word sense analysis suggests two points. First, the sense of a word is influenced by senses of its surrounding words, that is, its collocates. For example, the sense of *terms*, a specialised Logistics word, is rather ambiguous when it occurs in decontextualization, with probabilities of several senses. It can refer to technical words for specific concepts or terminology (e.g., a medical term), or law-related regulations and agreements (e.g., terms and conditions), or a period of schooltime (e.g., the spring term). However, disambiguity can be resolved and the sense is rather clear in collocational analysis. In the LAWEC corpus, *terms* is a polysemous word with two senses. When *terms* collocates with "in" or "of", it forms "in terms of", a general list collocation referring to "being related to". When *terms* collocates with "conditions", it refers to "commercial agreements to customers with legal effect". Second, concordance checking provides further information of word senses and collocations within context. Although the span applied to this research is +/-5 for collocational analysis, sometimes the sense of words is still ambiguous in collocations, if the concordance lines are not checked further. *suggest* is a typical example in the LAWEC. *suggest* is a polysemous word type having two senses in the LAWEC, one for "putting forward an idea for others to think about" (0.56%) and the other for "implying something" (99.44%). Both verb senses can be observed from *suggest* and one of its top collocates "these" (LogDice = 9.13). However, through concordance checking, a wide context is provided in concordance lines, as the extracted examples show below (emphasis added):

These researchers **suggest** that dominance analysis and relative weight analysis should be implemented to overcome problems of multicollinearity in future research. (BL190101)

These Figures **suggest** that there might be different types of firms exhibiting different production functions depending on their size. (LR200403)

In the first concordance, “researchers” had a specific advice in the that-clause for future studies, fitting the sense of “putting forward an idea for others to think about”. In the second concordance, “Figures” cannot have suggestions like researchers. Instead, it implies information for readers which is interpreted by writers in the that-clause. Both senses of suggest are related to its use in context. This manifests the perspective that meaning is use (Sinclair, 2004), and that meaning should be observed or inferred through multiple concordances (Sinclair, 2003). These suggest that more information can help with sense analysis in a wider context and beyond the span, showing the importance of concordance checking in word sense analysis. These extracts also show that collocations and concordances are powerful tools to identify senses. This reflects Moon’s (2008) claim that disambiguity of meaning can be resolved through collocation and concordances.

## Chapter 7 Conclusion

### 7.1 Summary of the research

This exploration of polysemy in Logistics academic written contexts and comparison with general academic and general use contexts is motivated by the fact that, first, most of existing corpus-derived specialised vocabulary lists focus on word forms, without less guidance of what senses a certain item may have in a specific context, and, second, polysemy is relatively less explored through corpus evidence. Polysemy, a pervasive linguistic phenomenon, may pose challenges to language learners (Deignan et al., 2023), especially when they are encountering with unfamiliar, specialised, technical contexts. This research, therefore, aims to answer three questions.

My first research question focuses on: RQ1: *What are the most frequently-used and evenly-distributed specialised Logistics words, and what are their formal features?*

To answer this question properly, a specialised vocabulary list on Logistics academic written English was planned for development. A specialised corpus on Logistics academic written English was compiled (the LAWEC), based on 720 texts collected from nine top journals in the field of Logistics, since there is no available corpus on Logistics. This corpus contains around 5.35 million tokens. Word type was selected as unit of counting for vocabulary list development based on previous studies (e.g., Chung and Nation, 2003; Benson and Coxhead, 2022; Lu, 2018). Keyness, range, dispersion and selection of lexical words were set as criteria. In total, 311 items were selected for the specialised vocabulary list – the LoSAWEL. The whole list provided around 16% lexical coverage of the specialised corpus. The LoSAWEL was divided into four levels, where the items at the first three, representing general high frequency vocabulary, accounted for 97.75% of all items in the list.

The second question is: RQ2: *To what extent are the words identified used in ways specific to Logistics?*

- a) *is their use different from the general academic context?*
- b) *is their use different from general English?*

Collocation analysis, concordance line reading, general use and discipline-specific dictionary checking were techniques and tools for word sense identification, following previous studies on word meaning or polysemy research, for example, Deigan et al. (2023), Le and Miller (2023), Moon (1987), Skoufaki and Petrić (2021). Pilot studies were conducted on the decision of collocation analysis statistical threshold, word sense splitting, and calculation of probability of occurrence of word senses. In total, 472 senses were identified in the LAWEC; 503 in the BAWE; 562 in the BNC1994 NAW. Each sense was provided with a mark on part of speech and a brief definition. Word senses comparison was conducted between the LAWEC and the BAWE, and the LAWEC and the BNC1994 NAW. In each

round of comparison, three kinds of word senses were noticed, based on 1.5-time word sense difference in percentage of occurrence in citations: LAWEC unique senses, shared senses (S) and shared senses (D). The LAWEC-BAWE comparison and that of LAWEC-BNC1994 NAW showed that around half of senses identified in the LAWEC were also identified in the counterpart corpus in comparison, as shared senses (D). Senses in the LAWEC and the BAWE had more similarities, compared to the BNC1994 NAW. Adverb and preposition senses were generally used consistently across three corpora. Noun, verb and adjective senses tend to be different among corpora. Moreover, unique senses and shared senses (D) were mostly identified in general use high frequency words in the LoSAWEL, showing that general high frequency words may have different senses used in different context.

The third research question focuses on: RQ3: *What is the extent and nature of polysemy in the words identified?* The results showed that, first, more than 1/3 of LoSAWEL items are polysemous words. Second, these polysemous words have relations with general use word frequency and part of speech. Specifically, items classified in levels with higher general use word frequency are more likely to be polysemous words, and vice versa. Few items have a large amount of senses in the LAWEC, and most of polysemous words have two senses. Nouns, verbs, adjectives are more likely to be polysemous words, but adverbs and prepositions are not. Third, through word sense identification, it was observed that some senses are closely associated with register-specific collocations (general list collocations and disciplinary collocations). This supports that specialised vocabulary helps with academic convention presentation and disciplinary concept expression. (Coxhead et al., 2020).

## 7.2 Contributions of the research

The first contribution of this research is its exploration of polysemy in a discipline-specific context and word sense comparison across contexts through corpus linguistics. Through word sense identification, a close association with word form and word sense is observed. It further supports that form and meaning are inseparable (Hunston, 2022). The word sense comparison shows that word sense varies according to disciplines and registers. This then provides evidence for research needs of corpus-derived specialised individual vocabulary and collocations. It is also one of the comprehensive studies to explore polysemy in Logistics English, a topic currently with relative few studies. Little research has adopted corpus analysis in exploring polysemy and conducting word sense comparison, especially in a discipline-specific context. My research fill these gaps, to some degree.

The second contribution is to combine discipline-specific dictionaries with corpus methods (e.g., collocation and concordance line analysis) for word sense identification. Previous studies (e.g., Moon, 1987, 2008; Sinclair, 2004) have shown that word sense disambiguation can be realized through collocation and concordance lines. In the technical or discipline-specific context, these tools, although

they are useful, may not be as efficiency as they are in general use contexts, due to language researchers' unfamiliarity with specialised context and vocabulary use. General use dictionaries may neither be as useful as they are in expectation, due to the scop of targeting different potential users. Discipline-specific dictionaries, with their special focus on a particular context, are a useful tool to help disambiguating discipline-specific senses that linguistic researchers may not be familiar with.

### 7.3 Implications of the research

This research has two pedagogical implications. First, it develops the first corpus-derived Logistics specialised vocabulary list. This list helps setting goals for those who have learning needs of or are interested in Logistics English. Now I can confidently answer the first question raise by my student couple of years ago: Yes, now we have a discipline-specific vocabulary list on Logistics. This list contains frequently-used and evenly-distributed vocabulary in Logistics, which can be a short-cut for vocabulary acquisition in this field. This vocabulary list also has reference to general use word frequency and is divided into four levels. Teachers and learners can allocate their attention to a specific level based on users' language proficiency level (Dang, 2017). Specifically, if a test shows that students have a vocabulary size of the most frequent 1,000 words, the teacher and the students can quickly go through items at Level 1 of the LoSAWEL, focusing on some items, if any, that are unknown to students. They can then allocate more attention to those at Level 2, 3, and 4 for instructions or deliberate learning. With percentage of word sense occurrence provided in the list, the teachers can explicitly check whether students know, or they believe that they know, the senses frequently occur in Logistics written academic context. They can check word senses through the sequence of general use word frequency levels (Level 1-4), or they can specifically focus on word senses conveying disciplinary concepts. They can set different priorities in teaching and learning, based on their aims. For example:

- (1) If they aim to test learners' vocabulary size and knowledge before ESP sessions, they can focus on word senses identified with (highest) percentage of occurrences in general use corpus (BNC1994NAW). It depends on their aims of testing only one frequent word sense or knowledge of polysemy.
- (2) If teachers aim to examine whether learners know what senses are frequently used in Logistics academic written context, they can set priorities on word senses identified with highest percentage of occurrences in the specialised corpus (the LAWEC).
- (3) If teachers aim to have investigation on whether learners know discipline-specific meanings, they can examine learners' knowledge of disciplinary collocations, marked as (*tm.*) in the LAWEC. Or they can use these as focus in teaching and learning.

- (4) If teachers aim to expand learners' vocabulary knowledge of word senses in a specific context, they can provide instructions on word senses with less percentage of occurrences in corpora, or have comparison on word senses and use across corpora.

Second, the brief definition provided and percentage of occurrence in citations calculated on each word sense provide an example of how to present word senses together with word forms. With these marks and calculations, I now can answer the second question raised by my student: Yes, now with this list, you will probably know how many senses a word has in this context, and which sense deserves your special attention. Similarities and differences of word sense used in different contexts (the BAWE and the BNC1994 NAW), and different percentage of occurrences in citations shown in the list provide both teachers and students with guidance on which senses need particular focus when they are in different contexts (i.e., general academic or general use context).

## 7.4 Limitations of the research

This research has several limitations. First, the specialised corpus – the LAWEC – only contains an academic register and focuses on written language only. This is not representative enough for students who are with academic or professional language learning needs in this field. But I believe such limitation is less important when compared to the benefits brought by the specialised vocabulary list.

Second, the Logistics specialised vocabulary list – the LoSAWEL – was not validated, either in another independent corpus with similar size and the same sample collection methods or in corpus of other registers (e.g., in a general academic corpus or a fiction corpus). Dang and Webb (2025) pointed out that vocabulary list validation is an important procedure of vocabulary list development, and should be involved in relevant studies.

Third, the collocation analysis only focuses on those  $\text{LogDice} \geq 9$ , or the top seven collocations. This is sufficient to identify some key, frequently-occurring senses, but may not sufficient to capture all senses in a specific context. There may exist the missed senses which may be useful and particular to Logistics students. In addition, although concordance lines were sampled and checked, far more amount of concordance lines were not read in this research. Moreover, the word sense identification results deserve peer validation. Indeed, splitting word sense is a subjective process, which should depend on research aims (Moon, 1987). It can be less subjective if it had experienced cross-checking.

## 7.5 Directions for future research

Polysemy is a pervasive and interesting linguistic phenomenon; however, more corpus studies are needed to explore it in specific contexts. Here are some recommendations for future research.

First, in terms of specialised corpora, larger and more representative specialised corpora can be developed, covering a wide range of registers of a particular context, both in written and spoken languages. Taking Logistics as an example, texts of different registers in this field can be collected, such as books, written report, official documents, professional lectures, business conference minutes, warehouse inventory or check list. In this way, the results of corpus-derived specialised vocabulary list and word sense identification would be more representative to the students who aim to have professional development in this field. It may also bring benefits to other stakeholders.

Second, more collocations and concordance lines can be checked in further analysis. Specifically, a lower LogDice threshold can be adopted, for example,  $\text{LogDice} \geq 4/5$ , which, in my pilot studies, generated hundreds of collocations. If time permits, checking more collocations and concordance lines can bring more insights to word senses analysis in specialised context and polysemy research. Or we can follow Stubbs (2001) to check all collocations and concordance lines in a relatively small but representative corpus.

Third, in word sense identification, peer validation can be a method to be employed, by following Le and Miller (2023) with a slight adjustment. Specifically, two more peers who have background of Logistics professional working or research experience can be invited to form a panel of three. The panel reaches a consensus on how to split senses and co-work on a group of randomly selected items to see how the procedure works. They then can work independently, and then exchange their results with each other. If disagreeing appears, the panel could discuss collaboratively for disambiguation tasks. If not solved, they can consult with another experienced researcher or expert in Logistics field.

Fourth, LLMs have potentials in disambiguate word senses in specific contexts. As shown in 3.3, studies (e.g., Geng and Liang, 2024; Gong et al., 2025) examined word senses in general academic context through LLMs, which could save a lot of labour and time in projects. Although my research finished in 2025, it started from early 2022, and at that time studies on LLMs were at their beginning stage. With time passing, LLMs demonstrate their ability in helping word sense disambiguation. I believe that using LLMs to investigate word senses in discipline-specific contexts can be a further direction of relevant studies.

## References

- Aarts, B., Chalker, S. and Weiner, E. 2015. *Oxford Dictionary of English Grammar*. Oxford: Oxford University Press.
- Ackermann, K. and Chen, Y. 2013. Developing the academic collocation list (ACL) – a corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*. **12**(4), pp.235–247.
- Aitchison, J. 2012. *Words in the Mind*. 4<sup>th</sup> edition. Oxford: Wiley-Blackwell.
- Al-Masrai, A. and Milton, J. 2012. The vocabulary knowledge of university students in Saudi Arabia. *Perspectives (TESOL Arabia)*. **19**(3), pp.13-19.
- Almeman, F. and Schockaert, S. 2024. WordNet under scrutiny: Dictionary examples in the era of large language models. In: Calzolari, N. ed. *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, 20-25<sup>th</sup> May 2024, Turin Italy*. ELRA and ICCL, pp.17683-17695.
- 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), pp.112-121.
- Andreoli-Versbach, P. and Gans, J. 2024. Interplay between Amazon store and logistics. *European Competition Journal*. **20**(2), pp.460-497.
- Anthony, L. 2022. What can corpus software do?. In: O’Keeffe, A. and McCarthy, M. ed. *The Routledge handbook of corpus linguistics*. 2<sup>nd</sup> ed. London: Routledge, pp. 103-125.
- Anthony, L. 2024. AntConc (4.3.1). [Software]. [Accessed 24<sup>th</sup> August 2025].
- Archer, D. 2009. *What's in a word-list?: investigating word frequency and keyword extraction*. Ashgate Publishing, Ltd.
- Ashrafzadeh, A. 2021. DEVELOPING A CORPUS-BASED ACADEMIC WORD LIST AND COLLOCATION LIST IN MEDICINE. *Journal of Language and Communication*. **8**(2), pp.239-300.
- Baker, M. 1988. Sub-technical vocabulary and the ESP teacher: An analysis of some rhetorical terms in medical journal articles. *Reading in a Foreign Language*. **4**, pp.47-64.
- Baker, P. 2006. *Using corpora in discourse analysis*. 1<sup>st</sup> ed. London: Routledge.
- Baker, P. 2023. *Using corpora in discourse analysis*. 2<sup>nd</sup> ed. London: Routledge.
- Barnbrook, G., Mason, O. and Krishnamurthy, R. 2013. *Collocation: Applications and Implications*. London: Palgrave Macmillan.

- Barraza, A., Eckstein, G., Lambert, Z., Rogers, C. and Vincent, J. 2024. The Applied Linguistics Academic Word List (ALAWL): The creation of a domain-specific word list for applied linguistics. *Journal of Academic Language and Learning*. **18**(1), pp.21-48.
- Bauer, L. and Nation, P. 1993. Word families. *International journal of Lexicography*. **6**(4), pp.253-279.
- Bauer, L. and Nation, I. S. P. 2020. *English morphology for the language teaching profession*. London: Routledge.
- Benson, S. and Coxhead, A., 2022. Technical single and multiword unit vocabulary in spoken rugby discourse. *English for Specific Purposes*. **66**, pp.111-130.
- Berdicevskis, A. 2020. Older English Words are More Polysemous. In: Ravnani, A. et al. ed. *The Evolution of Language. Proceedings of the 13<sup>th</sup> International Conference on the Evolution of Language, 14-17<sup>th</sup> April 2020, Vrije Universiteit Brussel*. Nijmegen: The Evolution of Language Conference, pp.14-21.
- Biber, D. 1993. Representativeness in corpus design. *Literary and linguistic computing*. **8**(4), pp.243-257.
- BNC Consortium. *The British National Corpus (BNC1994)*. <https://www.sketchengine.eu>
- Brezina, V. 2018. *Statistics in corpus linguistics: A practical guide*. Cambridge: Cambridge University Press.
- Brezina, V. and Fox, L. 2021. Adjective + noun collocations in L2 and L1 speech: Evidence from the Trinity Lancaster Corpus and the spoken BNC2014. In: S. Granger. ed. *Perspectives on the L2 phrasicon: The view from learner corpora*. Bristol: Multilingual Matters, pp. 152–177.
- Brezina, V. and Gablasova, D., 2015. Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics*. **36**(1), pp.1-22.
- Brezina, V., Platt, W. and Tangham, E. 2021. #LancsBox 6.0. [Software]. [Accessed 23<sup>rd</sup> April 2023].
- Brown, C., Culligan, B., & Phillips, J. 2013a. *New Academic Word List*. Retrieved from <http://www.newgeneralservicelist.org/nawl-new-academic-word-list/>.
- Browne, C., Culligan, B. and Phillips, J. 2013b. *The new general service list*. Retrieved from: [New General Service List ProjectNew General Service List Project](#).
- Brown, D., Stoeckel, T., Mclean, S. and Stewart, J. 2022. The most appropriate lexical unit for L2 vocabulary research and pedagogy: A brief review of the evidence. *Applied Linguistics*. **43**(3), pp.596-602.

- Burnard, L. 2007. Une introduction au British National Corpus dans son édition XML. *Texte et corpus*. **3**, pp.17-34.
- Cao, D. and Deignan, A. 2019. Using an online collocation dictionary to support learners' L2 writing. *Voices and practices in applied linguistics: Diversifying a discipline*, pp.233-249.
- Cao, D. T. P., Pho, P. D. and Chi, D. N. A. 2022. The treatment of collocations in English textbooks for Vietnamese students. *TEFLIN Journal*. **33**(2), pp.220-239.
- Clarivate. 2022. *Journal Citation Report*. Available from: [Clarivate Names World's Leading Journals with the 2022 Journal Citation Reports | Clarivate](#).
- Chanasattru, S. and Tangkiengsirisin, S. 2016. Developing of a high frequency word list in social sciences. *Journal of Studies in the English Language*. **11**, pp.41-87.
- Chen, L. 陈玲. 2019. Qianxi wuliu yingyu zhong yiciduoyi xianxiang ji fanyi. 浅析物流英语中一词多义现象及翻译. *Yingyu guangchang*. 英语广场. **108**, pp.44-45.
- Chen, H. and Lei, G. 2019. Developing a Technical Words List for Research Articles in Computer Science Discipline. *English Language Teaching*. **12**(10), pp.131-141.
- Chung, T. and Nation, I. S. P. 2003. Technical vocabulary in specialised texts. *Reading in a Foreign Language*. **15**(2), pp.103-116.
- Chung, T. M. and Nation, I. S. P. 2004. Identifying technical vocabulary. *System*. **32**(2), pp.251-263.
- CLPP. 2019. *Jiaoyubu gaodengxuexiao wuliuguanli yu gongchenglei zhuan ye jiaoxue zhidao weiyuanhui*. 教育部高等学校物流管理与工程类专业教学指导委员会. [Online]. [Accessed 26<sup>th</sup> July 2022]. Available from: <http://www.clpp.org.cn>
- Collins COBUILD Dictionary*. 10<sup>th</sup> ed. 2023.
- Conrad, S. 2001. Variation among discipline texts: A comparison of textbooks and journal articles in biology and history. In: S. Conrad. and D. Biber. ed. *Variation in English: Multidimensional studies*. Harlow: Longman, pp.94-107.
- Contract Logistics Observatory, 2023. Logistics Global HR Trends. [Online]. [Accessed 7<sup>th</sup> August 2025]. Available from: <https://www.gigroupholding.com>.
- Coxhead, A. 1998. *An academic word list*. School of Linguistics and Applied Language Studies, Victoria University of Wellington.
- Coxhead, A. 2000. A new academic word list. *TESOL Quarterly*. **34**(2), pp.213-238.

- Coxhead, A. 2011. The academic word list 10 years on: Research and teaching implications. *Tesol Quarterly*. **45**(2), pp.355-362.
- Coxhead, A. 2012. Vocabulary and ESP. In: Paltridge, B. and Starfield, S. ed. *The handbook of English for specific purposes*. Hoboken: Wiley-Blackwell.
- Coxhead, A., 2016. Reflecting on Coxhead (2000)," A New Academic Word List". *Tesol Quarterly*. **50**(1), pp.181-185.
- Coxhead, A. 2018. *Vocabulary and English for Specific Purposes Research*. New York: Routledge.
- Coxhead, A. 2021. Vocabulary in English in tertiary contexts: Connecting research and learning. *LEARN Journal: Language Education and Acquisition Research Network*. **14**(1), pp.1-14.
- Coxhead, A. 2024. Understanding teachers' and learners' perceptions of incidental vocabulary learning: Another piece of the puzzle. In Teng, M. F. and Reynolds, B. ed. *Researching incidental vocabulary learning in a second language*. London: Routledge, pp. 119-130.
- Coxhead, A. and Demecheleer, M. 2018. Investigating the technical vocabulary of plumbing. *English for Specific Purposes*. **51**, pp.84-97.
- Coxhead, A., & Hirsh, D. 2007. A pilot science word list for EAP. *Revue Française de linguistique appliquée*. **XII** (2) 65–78.
- Coxhead, A., Parkinson, J., Mackay, J. and McLaughlin, E. 2020. *English for vocational purposes: Language use in trades education*. New York: Routledge.
- Crossley, S., Salsbury, T. and McNamara, D. 2010. The development of polysemy and frequency use in English second language speakers. *Language Learning*. **60**(3), pp.573-605.
- Cruse, D. A. 2002. *Lexical Semantics*. Cambridge: Cambridge University Press.
- Csábi, S. 2004. A cognitive linguistic view of polysemy in English and its implications for teaching. In: Achard, M. and Niemeier, S. ed. *Cognitive Linguistics, Second Language Acquisition, and Foreign Language Teaching*. Berlin: De Gruyter, pp.233-256.
- Dai, L. 戴丽娟. 2020. Yidaiyilu changyi xia gaozhixuanxiao wuliuyingyu jiaoxue gaige tanxi. “一带一路”倡议下高职院校物流英语教学改革探析. *Yingyu guangchang. 英语广场*. **124**, pp.94-97.
- Dang, T. N. Y. 2018a. A Hard Science Spoken Word List. *ITL - International Journal of Applied Linguistics*. **169**(1), pp. 44-71.

- Dang, T. N. Y. 2018b. The nature of vocabulary in academic speech of hard and soft-sciences. *English for Specific Purposes*. **51**, pp.69-83.
- Dang, T. N. Y. 2020. Corpus-based Word Lists in Second Language Vocabulary Research, Learning and Teaching. In: Webb, S. ed. *The Routledge Handbook of Vocabulary Studies*. London: Routledge, pp.288-303.
- Dang, T. N. Y. and Webb, S. 2025. Applications of word lists in second language learning and teaching. *Language Teaching*. **58**, pp.291-311.
- Dang, T. N. Y., Coxhead, A. and Webb, S. 2017. The academic spoken word list. *Language Learning*. **67**(4), pp.959-997.
- Dang T. N. Y., Webb, S., Coxhead, A. 2022. Evaluating lists of high-frequency words: Teachers' and learners' perspectives. *Language Teaching Research*. **26**(4), pp. 617-641.
- Davis, M. 2020. *The Concise Dictionary of Crime and Justice*. London: SAGE Publishing.
- Davies, M. 2008. The Corpus of Contemporary American English (COCA). [www.english-corpora.org/coca/](http://www.english-corpora.org/coca/).
- Daxueshengbibeiwang. 2025. Which university has Logistics as a major? [Online]. [Accessed 27<sup>th</sup> August 2025]. Available from: <https://www.dxsbb.com/11463.html>.
- de Chazal., E. 2014. *English for Academic Purposes*. Oxford: Oxford University Press.
- Deignan, A., Candarli, D. and Oxley, F. 2023. *The Linguistic challenge of the transition to secondary school*. London: Routledge.
- Deignan, A. and Love, R. 2021. Using corpus methods to identify subject specific uses of polysemous words in English secondary school science materials. *Corpora*. **16**(2), pp.165-189.
- Deignan, A. and Jone, M. 2025. *New Words, New Meanings*. London: Routledge.
- Deignan, A., Semino, E. and Paul, S. A. 2019. Metaphors of climate science in three genres: Research articles, educational texts, and secondary school student talk. *Applied Linguistics*. **40**(2), pp.379-403.
- Delfmann, W., Dangelmaier, W., Gunthner, W., Klaus, P., Overmeyer, L., Rothengatter, W. 2010. Towards a science of logistics: cornerstones of a framework of understanding of logistics as an academic discipline. *Logistics Research*. **2**(2), pp.57-63.
- Deutchman, S. M. 2022. Polysemy and homonymy should be considered when teaching vocabulary. Website: <http://pansig.org>/Facebook: <https://www.facebook.com/JALTpansig>, p.311-317.

- Dorst, A. and Reijnierse, W. 2015. A dictionary gives definitions, not decision: On using a dictionary to identify the basic senses of words. *Metaphor and the Social World*. **5**(1), pp.137-144.
- Drayton, J. and Coxhead, A. 2023. The development, evaluation and application of an aviation radiotelephony specialised technical vocabulary list. *English for Specific Purposes*. **69**, pp.51-66.
- Durrant, P., Siyanova-Chanturia, A., Kremmel, B. and Sonbul, S. 2022. *Research Methods in Vocabulary Studies*. Amsterdam: John Benjamins.
- Earl, R. and Nicholson, J. 2021. *The Concise Oxford Dictionary of Mathematics*. 6<sup>th</sup> ed. Oxford: Oxford University Press.
- Egbert, J., Larsson, T. and Biber, D. 2020. *Doing linguistics with a corpus: Methodological considerations for the everyday user*. Cambridge: Cambridge University Press.
- Egbert, J., Biber, D. and Gray, B. 2022. *Designing and evaluating language corpora: A practical framework for corpus representativeness*. Cambridge University Press.
- Ellis, N. 2002. Frequency effects in language processing. *Studies in Second Language Acquisition*. **24**, pp.143-188.
- Enesi, M., Vrapı, F. and Trifoni, A., 2021. Challenges of teaching and learning English language for ESP courses. *Journal of Educational and Social Research*. 11(4), pp.213-226.
- Falkum, I. L. 2015. The how and why of polysemy: A pragmatic account. *Lingua*. **157**, pp.83-99.
- Firth, J. 1957. Modes of meaning. In: *Papers in Linguistics, 1934-1951*. Oxford: Oxford University Press.
- Francis, W. N. E. L. S. O. N., and Kučera, H. 1982. *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.
- Fraser, S. 2012. Factors affecting the learnability of technical vocabulary: findings from a specialized corpus. *広島外国語教育研究*. **15**, pp.123-142.
- Fraser, S., Higa, M. K. and Davies, W., 2025. Delivering an ESP Pedagogic Word List: Integrating Corpus Analysis, Materials Design, and Software Development. *Languages*. **10**(3), p.46-58.
- Freund, Z. K. 2014. A corpus analysis of grant guidelines: The education and training programme word list (ETPWL). *Journal of Teaching English for Specific and Academic Purposes*. **2**(3), pp.501-514.
- Friginal, E. 2018. *Corpus Linguistics for English Teachers*. London: Routledge.

- Gablasova, D. and Brezina, V. 2021. Words that matter in L2 research and pedagogy: A corpus-linguistics perspective. *Studies in Second Language Acquisition*. **43**(5), pp.958-961.
- Gablasova, D., Brezina, V. and McEnery, T. 2017. Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language learning*. **67**(1), pp.155-179.
- Gardner, D. 2007. Validating the Construct of Word in Applied Corpus-based Vocabulary Research: A Critical Survey. *Applied Linguistics*. **28**(2), pp.241-265.
- Gardner, D. and Davies, M. 2014. A new academic vocabulary list. *Applied linguistics*. **35**(3), pp.305-327.
- Geng, W. and Liang, M. 2024. From words to senses: A sense-based approach to quantitative polysemy detection across disciplines. *Journal of English for Academic Purposes*. **72**, 101449.
- Ghadessy, P., 1979. Frequency counts, word lists, and materials preparation: A new approach. *English Teaching Forum*. **17**(1), pp.24-27.
- Gong, T., Liu, L., Shi, J. and Guo, Y. 2025. A meaning-based academic vocabulary list. *Journal of English for Academic Purposes*. **77**, p.101557.
- Gong, Y. and Song, Y. 龚艳茹、宋玉佳. 2022. Jiyu hulianwang+ shiyu de wuliuyingyu jiaoxuetixi jianshe. The Construction of Logistics English Teaching System Based on the “Internet+” Perspective – Take the Course of “Logistics Professional English” as An Example. *Wuliu keji. Logistics Sci-Tech*. **2022(June)**, pp.172-173, 181.
- Green, C. and Lambert, J. 2018. Advancing disciplinary literacy through English for academic purposes: Discipline-specific wordlists, collocations and word families for eight secondary subjects. *Journal of English for Academic Purposes*. **35**, pp.105-115.
- Greene, J.W. and Coxhead, A. 2015. *Academic vocabulary for middle school students. Research-Based Lists and Strategies for Key Content Areas*. Baltimore: Brookes Publishing Co.
- Gries, S.T., 2008. Dispersions and adjusted frequencies in corpora. *International journal of corpus linguistics*. **13**(4), pp.403-437.
- Gries, S., 2021. Analyzing dispersion. In Paquot, A. Gries, S. ed: *A practical handbook of corpus linguistics*. London: Springer International Publishing, pp. 99-118.
- Gulhayo, P. 2024. WordNet – A lexical database for linguistic ontologies. *Current Research Journal of Philological Sciences*. **5**(12), pp.27-32.
- Györi, G. 2002. Semantic change and cognition. *Cognitive linguistics*. **13**(2), pp.123-167.

- Haahr, M. 1998. *Random selection website*. Available from: [RANDOM.ORG - True Random Number Service](http://RANDOM.ORG - True Random Number Service).
- Ha, Y. H. 2015. *Technical vocabulary in finance: a corpus-based study of annual reports and earnings calls*. HKU Theses Online (HKUTO).
- Hanks, P. 2013. *Lexical analysis: Norms and exploitations*. Cambridge, USA: MIT Press.
- Hassel, A. and Sieker, F. 2022. The platform effect: How Amazon changed work in logistics in Germany, the United States and the United Kingdom. *European Journal of Industrial Relations*. **28**(3), pp.363-382.
- Havwini, T., Ratnasari, D., Fadilah, N., Feby, F. and Firdaus, F. 2024. English vocabulary mastery in vocational higher education: An analysis of challenges and feasible strategies. *Journal on Education*. 6(4), pp.19110-19118.
- He, X. and Godfroid, A. 2019. Choosing words to teach: A novel method for vocabulary selection and its practical application. *Tesol Quarterly*. **53**(2), pp.348-371.
- Heuboeck, A., Holmes, J. and Nesi, H. 2010. *The BAWE corpus manual*. Available from: <http://www.reading.ac.uk/internal/appling/bawe/BAWE>.
- Hoey, M. 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- Hou, H. I. 2014. Teaching Specialized Vocabulary by Integrating a Corpus-Based Approach: Implications for ESP Course Design at the University Level. *English Language Teaching*. **7**(5), pp.26-37.
- Hsu, W. 2011. The vocabulary thresholds of business textbooks and business research articles for EFL learners. *English for Specific Purposes*. **30**(4), pp.247-257.
- Hsu, W. 2013. Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *English for Specific Purposes*. **33**, pp.54-65.
- Hsu, W. 2013. Bridging the vocabulary gap for EFL medical undergraduates: The establishment of a medical word list. *Language Teaching Research*. **17**(4), pp.454-484.
- Hu, C., Chen, L. He, A. 胡春雨、陈丽丹、何安平. 2019. *Yuliaoku fuzhu de shangwu yingyu duanyu jiaoxue yanjiu. 语料库辅助的商务英语短语教学研究*. Beijing: FLTRP.
- Hulstijn, J.H. and Laufer, B. 2001. Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language learning*. **51**(3), pp.539-558.
- Hunston, S. 2002. *Corpora in Applied Linguistics*. 1<sup>st</sup> ed. Cambridge: Cambridge University Press.

- Hunston, S. 2022. *Corpora in Applied Linguistics*. 2<sup>nd</sup> ed. Cambridge: Cambridge University Press.
- Hyland, K. 2022. *Teaching and Researching Writing*. 4<sup>th</sup> ed. London: Routledge.
- Hyland, K. and Tse, P. 2007. Is there an “academic vocabulary”? *TESOL Quarterly*. **41**(2), pp.235-253.
- It-ngam, T. and Phoocharoensil, S. 2015. The development of science academic word list. *Indonesian Journal of Applied Linguistics*. **8**(3), pp.657-667.
- IWLA. 2018. *Warehouse Supply Chain & Logistics Terms and Glossary*. Available from: [IWLA-Glossary-and-Terminology.pdf](#).
- Jacobs, G. 1989. Word usage misconceptions among first-year university physics students. *International Journal of Science Education*. **11**(4), pp.395-399.
- Jin, Q. 2021. Supporting indigenous students in science and STEM education: A systematic review. *Education Sciences*. **11**(9), p.555-570.
- Jurafsky, D. and Martin, J. 2025. *Speech and Language Processing*. Stanford University. Available from: [Speech and Language Processing](#)
- Khani, R. and Tazik, K. 2013. Towards the development of an academic word list for applied linguistics research articles. *RELC journal*. **44**(2), pp.209-232.
- Kilgarriff, A. 1992. Dictionary word sense distinctions: An enquiry into their nature. *Computers and the Humanities*. **26**(5), pp.365-387.
- Kilgarriff, A. 2006. Word senses. *Word Sense Disambiguation*. London: Springer, pp.29-46.
- Kilgarriff, A., Baisa, V., Busta, J., Jakubicek, M., Kovar, V., Michelfeit, J., Rychly, P. and Suchomel, V. 2014. The Sketch Engine: ten years on. *Lexicography*. **1**, pp.7-36.
- Kitano, M.L. 2018. The High Frequency Word Gap: What Words Don't Students Know?. *Concern*. **5**, pp.19-24.
- Knight, D. and Adolphs, S. 2022. Building a spoken corpus: what are the basics?. In: O’Keffee, A. and McCarthy, M. ed. *The Routledge handbook of corpus linguistics*. 2<sup>nd</sup> ed. London: Routledge, pp.21-34.
- Koester, A. 2022. Building Small Specialised Corpora. In: O’Keffee, A. and McCarthy, M. ed. *The Routledge Handbook of Corpus Linguistics*. 2<sup>nd</sup> ed. London: Routledge, pp.48-61.
- Konstantakis, N., 2007. Creating a business word list for teaching business English. *Elia*. **7**(1), pp.79-102.

- Kremmel, B. 2021. Selling the (word) family silver?: A response to Webb's "lemma dilemma". *Studies in Second Language Acquisition*. **43**(5), pp.962-964.
- Kruawong, T. and Phoocharoensil, S. 2020. Developing an English zoology academic word list: A corpus-based study. *Thoughts*. **2**, pp.63-78.
- Kuiper, K., Fromont, R. and Gerhard, D. 2018. Polysemy and word frequency. *Journal of Research Design and Statistics in Linguistics and Communication Science*. **4**(2), pp.144-155.
- Kwary, D. A. and Artha, A. F. 2017. The academic article word list for social sciences. *Mextesol Journal*. **41**(4), pp.1-11.
- Laosrirattanachai, P. and Ruangjaroon, S. 2021. Corpus-Based Creation of Tourism, Hotel, and Airline Business Word Lists. *LEARN Journal: Language Education and Acquisition Research Network*. **14**(1), pp.50-86.
- Laufer, B. 1997. "The lexical plight in second language reading: words you don't know, words you think you know and words you can't guess". In: Coady, J. and Huckin, T. ed. *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*. Cambridge: Cambridge University Press. pp. 20-34.
- Laufer, B. 2021. Lemmas, flemmas, word families, and common sense. *Studies in Second Language Acquisition*. **43**(5), pp.965-968.
- Law, J. 2022. *Oxford Dictionary of Law. 10<sup>th</sup> edition*. Oxford: Oxford University Press.
- Law, J. and Martin, E. 2020. *Concise Medical Dictionary. 10<sup>th</sup> edition*. Oxford: Oxford University Press.
- Le, C. N. N. and Miller, J. 2023. A core meaning-based analysis of English semi-technical vocabulary in the medical field. *English for Specific Purposes*. **70**, pp.252-266.
- Lee, Y. J. 2021. Polysemous words in English movies, learning obstacles or gifted talent. *Journal of English Teaching through Movies and Media*. **22**(4), pp.14-26.
- Lee, H. K. and Kim, H. O. 2020. Constructing a vocabulary list of English textbooks in semantics and pragmatics. *영어학*. **20**, pp.384-410.
- Lei, L. and Liu, D., 2016. A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for academic purposes*. **22**, pp.42-53.
- L'Homme, M. C., 2021. Revisiting polysemy in terminology. *EURALEX XIX*, pp.415-427.
- Li, X. 李秀敏. 2018. Wuliuzhuanye yingyu jiaoxue gaige shizheng yanjiu. 物流专业英语教学改革实证研究. *Journal of Liaoning Teachers College*. *辽宁师专学报*. **117**, pp.55-57.

- Lijffijt, J., & Gries, S. T. 2012. Correction to “Dispersions and adjusted frequencies in corpora”. *International Journal of Corpus Linguistics*. **17**(1), 147–149.
- Liu, D. and Lei, L. 2020. Technical Vocabulary. In: Webb, S. ed. *The Routledge Handbook of Vocabulary Studies*. London: Routledge, pp.111-124.
- Liu, J. and Han, L. 2015. A corpus-based environmental academic word list building and its validity test. *English for Specific Purposes*. **39**, pp.1-11.
- Liu, L., Gong, T., Shi, J, and Guo, Y. 2025. Calculating Semantic Frequency of GSL words Using a BERT model in large corpora. *Sage*, **15**(2), available from: <https://doi.org/10.1177/21582440251333182>.
- Long, X. 2023. *WeChat personal communication*. 23<sup>rd</sup> August, 2023.
- Lu, C. 2018. *Investigating knowledge and use of technical vocabulary in Traditional Chinese Medicine* (Doctoral dissertation, Open Access Te Herenga Waka-Victoria University of Wellington).
- Lu, C. and Dang, T. N. Y. 2022. Vocabulary in EAP learning materials: What can we learn from teachers, learners, and corpora?. *System*. **106**, 102791.
- Luo, J., Rong, Y. and Zheng, H. 2020. Impacts of Logistics information on sales: Evidence from Alibaba. *Naval Research Logistics*. **67**(8), pp.646-669.
- Lynn, R. W. 1973. Preparing word-lists: a suggested method. *RELC journal*. **4**(1), pp.25-28.
- MacArthur, F. 2015. On using a dictionary to identify the basic senses of words. *Metaphor and the Social World*. **5**(1), pp.124-136.
- Martínez, I. A., Beck, S. C. and Panza, C. B. 2009. Academic vocabulary in agriculture research articles: A corpus-based study. *English for specific purposes*. **28**(3), pp.183-198.
- McCarthy, M. 2023. *English Vocabulary: the basics*. London: Routledge.
- McEnery, T. and Brookes, G., 2022. Building a written corpus: what are the basics?. In: O’Keffee, A. and McCarthy, M. ed. *The Routledge handbook of corpus linguistics*. 2<sup>nd</sup>. ed. London: Routledge, pp.35-47.
- McLean, S. and Stoeckel, T., 2021. Lexical mastery thresholds and lexical units: A reply to Laufer. *Reading in a Foreign Language*. **33**(2), pp.247-259.
- Mićović, D. N. and Beko, L. V. 2022. Polysemy-Related Problems in ESP Students—A Case Study. *Зборник радова Филозофског факултета у Приштини*. **52**(3), pp.123-144.

- Milenkov, M. A., Sokolović, V. S., Milovanović, V. R. and Milić, M. D. 2020. Logistics – its role, significance and approaches. *Vojnotehnicki glasniki/Military Technical Courier*. **68**(1), pp.79-106.
- Ministry of Education. 中华人民共和国教育部. 2022. *Zhongwai hezuo banxue. 中外合作办学*. [Online]. [Accessed 17<sup>th</sup> April 2022]. Available from: <https://www.crs.jsj.edu.cn/approval/orglists/1>
- Mojela, V. M. 2007. Polysemy and homonymy: Challenges relating to lexical entries in the Sesotho sa Leboa–English Bilingual Dictionary. *Lexikos*. **17**. pp.433-439.
- Moon, R., 1987. The analysis of meaning. In: Sinclair, J. ed. *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. pp.86-103.
- Moon, R. 2008. Conventionalized as-similes in English: A problem case. *International Journal of Corpus Linguistics*. **13**(1), pp.3-37.
- Mozaffari, A. and Moini, R. 2014. Academic words in education research articles: A corpus study. *Procedia-Social and Behavioral Sciences*. **98**. pp.1290-1296.
- Nagy, W. and Townsend, D. 2012. Words as tools: Learning academic vocabulary as language acquisition. *Reading research quarterly*. **47**(1), pp.91-108.
- Nation, I. S. P. 2001. *Learning Vocabulary in Another Language*. 1<sup>st</sup> edition. Cambridge: Cambridge University Press.
- Nation, I. S. P. 2012. The BNC/COCA word family lists. Retrieved from: [www.victoria.ac.nz/lals/about/staff/pau;-nation](http://www.victoria.ac.nz/lals/about/staff/pau;-nation).
- Nation, I. S. P. 2013. *Learning Vocabulary in Another Language*. 2<sup>nd</sup> edition. Cambridge: Cambridge University Press.
- Nation, I. S. P. 2016. *Making and Using Word Lists for Language Learning and Testing*. Amsterdam: John Benjamins.
- Nation, I. S. P. 2021. Thoughts on word families. *Studies in Second Language Acquisition*. **43**(5), pp.969-972.
- Nation, I. S. P. 2022. *Learning Vocabulary in Another Language*. 3<sup>rd</sup> edition. Cambridge: Cambridge University Press.
- Nation, I. S. P. and Coxhead, A. 2021. *Measuring native-speaker vocabulary size*. New York: John Benjamins.

- Nation, I. S. P. and Kobleeva, P., 2016. Proper names. In: Nation, I. S. P. ed. *Making and using word lists for language learning and testing*. Amsterdam: John Benjamins Publishing Company, pp. 55-64.
- Nation, I. S. P., Coxhead, A., Wang, T. and Quero, B. 2016. Specialized word lists. In: Nation, I. S. P. ed. *Making and Using Word Lists in Language Learning and Teaching*. Amsterdam: John Benjamins, pp.145-151.
- Nation, I. S. P. and Gu, P. Y. 2019. *Teaching and learning vocabulary in EFL*. Beijing: Foreign Language Teaching and Research Press.
- Nation, I. S. P. and Webb, S. A. 2011. *Researching and analyzing vocabulary*. Boston, MA: Heinle, Cengage Learning.
- Nesi, H. 2002, August. An English spoken academic word list. In: *Proceedings of the tenth EURALEX International Congress*. Vol. 1. Copenhagen, Denmark: EURALEX, pp. 351-358.
- Nordin, N.R., Stapa, S.H. and Darus, S. 2013. Developing a specialized vocabulary word list in a composition culinary course through lecture notes. *Advances in Language and Literary Studies*. 4(1), pp.78-88.
- Oakey, D. 2009. Fixed collocational patterns in isolexical and isotextual versions of a corpus. In: P. Baker. ed. *Contemporary Corpus Linguistics*. London: Continuum.
- Oakey, D. 2022. What can a corpus tell us about lexis?. In O’Keffee, A. and McCarthy. M. ed. *The Routledge handbook of corpus linguistics*. 2<sup>nd</sup> ed. London: Routledge, pp. 185-203.
- O’Flynn, J. A. 2019. An Economics Academic Word List (EAWL): Using online resources to develop a subject-specific word list and associated teaching-learning materials. *Journal of academic language and learning*. 13(1), pp.A28-A87.
- O’Keeffe, A., McCarthy, M. and Carter, R. 2007. *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- O’Neil, D. 2018. English as the lingua franca of international publishing. *World Englishes*. 37(2), pp.146-165.
- Oxford English Dictionary*. [Online]. Available from: [https://www.oed.com/dictionary/bank\\_n1?tab=etymology](https://www.oed.com/dictionary/bank_n1?tab=etymology)
- Pan, G. 潘桂娟. 2008. Wuliuzhuanye yingyu jiaoxue wenti yu duice. Problems and Countermeasures on the Teaching of Logistics ESP. *Zhongguoshichang*. 中国市场. 8, pp.140-141.

- Pantel, P. and Lin, D. 2002. July. Discovering word senses from text. In: *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 613-619.
- Paquot, M. 2007. Towards a productively-oriented academic word list. In: Walinski, J., Kredens, K., Gozdz-Roszkowski, S. ed. *Corpora and ICT in language studies*, pp.127-140.
- Parent, K. 2009. *Polysemy: A second language pedagogical concern* (Doctoral dissertation, Open Access Te Herenga Waka-Victoria University of Wellington).
- Partington, A. 1998. *Patterns and meanings*. Amsterdam: John Benjamins.
- Pinto, P. T., Crosthwaite, P., de Carvalho, C. T., Spinelli, F., Serpa, T., Garcia, W., & Ottaiano, A. O. 2023. *Using language data to learn about language: A teachers' guide to classroom corpus use*. University of Queensland. Available from: <https://doi.org/10.14264/3bbe92d>.
- Piribabadi, A. and Rahmany, R. 2014. The effect of the keyword method and word-list method instruction on ESP vocabulary learning. *Journal of Language Teaching and Research*. **5**(5), pp.1110-1115.
- Pojanapunya, P. and Watson Todd, R. 2021. The influence of the benchmark corpus on keyword analysis. *Register Studies*. **3**(1), pp.88-114.
- Powers, D.M. 1998. Applications and explanations of Zipf's law. In: *New methods in language processing and computational natural language learning*. pp.151-160.
- Praninskas, J. 1972. *American university word list*. New York: Longman.
- Precedence Research. 2025. The Logistics Market Share, Size and Trends 2025 to 2034. [Online]. [Accessed 7<sup>th</sup> August 2025]. Available from: <https://www.precedenceresearch.com/logistics-market>
- Quero, B. 2015. *Estimating the vocabulary size of L1 Spanish ESP learners and the vocabulary load of medical textbooks* (Doctoral dissertation, Open Access Te Herenga Waka-Victoria University of Wellington).
- Quigley, A., 2018. *Closing the vocabulary gap*. London: Routledge.
- Quyet, N. X. and Phung, T. K. 2023. Comparative analysis of information security policies at Big 4 Vietnamese logistics companies. *International Journal of Multidisciplinary Research and Growth Evaluation*. **4**(6), pp.683-690.
- Rayson, P. 2019. Corpus analysis of key words. In: Chappelle, A. ed: *The concise encyclopedia of applied linguistics*. London: Wiley. pp.320-326.

- Reimann, S. and Scheffler, T. 2025. Using Large Language Models to Perform MIPVU-Inspired Automatic Metaphor Detection. *The 2nd Workshop on Analogical Abstraction in Cognition, Perception, and Language (Analogy-Angle II)*. Kerrville: Association for Computer Linguistics, pp.10-21.
- Rennie, R. and Law, J. 2019. *Oxford Dictionary of Physics*. Oxford: Oxford University Press.
- Reppen, R. and Simpson-Valch, R. 2020. Corpus Linguistics. In: Schmitt, N. and Rogers, M. ed. *An Introduction to Applied Linguistics*. London: Taylor & Francis, pp.91-108.
- Riccobono, P. S. 2019. Towards English for Summer Olympics Purposes: A Corpus-Based Vocabulary Analysis. *KOTESOL Proceedings 2019*. pp.123-136.
- Riccobono, P. S. 2020. Triangulating diamond talk: Identifying technical spoken vocabulary in English for baseball purposes. *ESP Today*. **8**(2), pp.114-140.
- Risberg, A. 2023. A systematic literature review on e-commerce logistics: towards an e-commerce and omni-channel decision framework. *The International Review of Retail, Distribution, and Consumer Research*. **33**(1), pp.67-91.
- Robinson, P. 1980. *ESP (English for Specific Purposes)*. Oxford: Pergamon Press.
- Roesler, D., 2020. *A computer science academic vocabulary list* (Master's thesis, Portland State University).
- Ruhl, C. 1989. *On monosemy: A study in linguistic semantics*. SUNY Press.
- Rychlý, P., 2008. *A Lexicographer-Friendly Association Score*. RASLAN, pp.6-9.
- Saeedi, M., Khany, R. and Tazik, K. 2023. Towards the Development of a Word List for English Newspapers. *REFlections*. **30**(2), pp.506-525.
- Safari, M. 2019. English vocabulary for equine veterans: How different from GSL and AWL words. *Iranian Journal of English for Academic Purposes*. **8**(2), pp.51-65.
- Sardarabady, N. J. and Durst, S. 2024. A systematic literature review on the economic impact of digitalization technologies in transport logistics. *Transport Economics and Management*. **2**, pp.76-89.
- Schmitt, N. 2010. *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.
- Schmitt, N. and Schmitt, D. 2020. *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Scott, M. 1997. PC analysis of key words—and key key words. *System*. **25**(2), pp.233-245.

- Simpson-Vlach, R. and Ellis, N.C. 2010. An academic formulas list: New methods in phraseology research. *Applied linguistics*. **31**(4), pp.487-512.
- Sinclair, J. M. 1966. 'Beginning the study of lexis'. In: C. Bazell et al. ed. *In Memory of J.R. Firth*. London: Longman, 410–430.
- Sinclair, J. 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing*. London: Collins ELT.
- Sinclair, J. M. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. M., 2003. *Reading Concordances: An Introduction*. London: Longman.
- Sinclair, J. M. 2004. *Trust the Text: Language, Corpus and Discourse*. London: Routledge.
- Sinclair, J., Jones, S. and Daley, R. 1970. *English collocation studies: the OSTI report*. London: Continuum.
- Skoufaki, S. and Petrić, B. 2021. Exploring polysemy in the Academic Vocabulary List: a lexicographic study. *Journal of English for Academic Purposes*. **54**, 10138.
- Smith, S. 2021. EAP Foundation. [Online]. [Accessed 17<sup>th</sup> April 2021]. Available from: <https://www.eapfoundation.com>
- Snoder, P. and Laufer, B. 2021. *EFL learners' receptive knowledge of derived words: The case of Swedish adolescents* [Manuscript submitted for publication]. Department of Language Education, Stockholm University and Department of English Language and Literature, University of Haifa.
- Stefanowitsch, A. 2020. *Corpus linguistics: A guide to the methodology*. Berlin: Language Science Press.
- Stubbs, M. 2001. *Words and phrases: Corpus studies of lexical semantics*. New York: John Wiley & Sons.
- Sultanova, U. 2025. Semantic Analysis of English Polysemous Words. *Pubmedia Jurnal Pendidikan Bahasa Inggris*. **2**(3), pp.1-7.
- Swanson, D., Goel, L., Francisco, K. and Stock, J. 2017. Applying theories from other disciplines to logistics and supply chain management: a systematic literature review. *Transportation Journal*. **56**(3), pp.299-356.
- Szudarski, P. 2022. *Corpus Linguistics for Vocabulary*. London: Routledge.

- Tangpijaikul, M. 2014. Preparing business vocabulary for the ESP classroom. *RELC Journal*. **45**(1), pp.51-65.
- Tabrizi, H. and Jamalzadeh, M. 2020. Academic vocabulary in tourism research articles: A corpus-based study. *Journal of Language and Discourse Practice*. **1**(2), pp.25-44.
- Thompson, P. and Hunston, S. 2019. *Interdisciplinary research discourse: Corpus investigations into environment journals*. London: Routledge.
- Thornbury, S. 2002. *How to teach vocabulary*. New York: Pearson Longman.
- Tongpoon-Patanasorn, A. 2018. Developing a frequent technical words list for finance: A hybrid approach. *English for Specific Purposes*. **51**, pp.45-54.
- University of Warwick. 2017. *Technical vocabulary*. [Online]. [Accessed 27<sup>th</sup> August 2025]. Available from:  
<https://www2.warwick.ac.uk/fac/soc/al/globalpad/openhouse/academicenglishskills/vocabulary/tech/>
- Valipouri, L. and Nassaji, H. 2013. A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes*. **12**(4), pp.248-263.
- Veenstra, J. and Sato, Y. 2018. Creating an institution-specific science and engineering academic word list for university students. *Journal of Asia TEFL*. **15**(1), p.148-166.
- Viana, V. 2022. *Teaching English with corpora: A resource book*. London: Taylor & Francis.
- Vilkaitė-Lozdienė, L. and Schmitt, N., 2020. Frequency as a guide for vocabulary usefulness. In: Webb. S. ed. *Routledge Handbook of Vocabulary Studies*. London: Routledge, pp.81-96.
- Vongpumivitch, V., Huang, J.Y. and Chang, Y.C. 2009. Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes*. **28**(1), pp.33-41.
- Wang, J., Liang, S.L. and Ge, G.C. 2008. Establishment of a medical academic word list. *English for Specific Purposes*. **27**(4), pp.442-458.
- Wang, K. 王科贝. 2022. Renwu qudong jiaoxuefa zai wuliuyingyu zhuanke ketangjiaoxue zhong de yingyong. 任务驱动教学法在物流英语专业课堂教学法中的应用. *Xiandai yingyu. 现代英语*. **15**, pp.41-44.
- Wang M, Nation P. 2004. Word meaning in academic English: homography in the academic word list. *Applied Linguistics*. **25**(3): 291-314.

- Wang, Y. and Zhao, J. 2022. Research on Jing Dong's Self-built Logistics Based on Technology Acceptance Model. *BCP Business & Management*. **38**, pp.614-628.
- Ward, J. 2009. A basic engineering English word list for less proficient foundation engineering undergraduates. *English for specific purposes*. **28**(3), pp.170-182.
- Watson-Todd, R. 2017. An opaque engineering word list: Which words should a teacher focus on?. *English for Specific Purposes*. **45**, pp.31-39.
- Webb, S. 2021. The lemma dilemma: How should words be operationalized in research and pedagogy?. *Studies in Second Language Acquisition*. **43**(5), pp.941-949.
- Webb, S. and Nation, I. S. P. 2017. *How vocabulary is learned*. Oxford: Oxford University Press.
- Wei, N. 卫乃兴. 2001. *Ciyu dapei de jieding yu yanjiu tixi. 词语搭配的界定与研究体系*. Shanghai: Shanghai Jiaoda Press.
- Wei, X. and Lou, Y. 2015. A study of word meaning acquisition of polysemous words for English majors. *Creative Education*. **6**(18), pp.1993-2001.
- Weisser, M. 2015. *Practical corpus linguistics: An introduction to corpus-based language analysis*. John Wiley & Sons.
- West, M. 1953. *A general service list of English words, with semantic frequencies and a supplementary word-list for the writing of popular science and technology*. London: Longman.
- Wilkins, D. A. 1972. *Linguistics in Language Teaching*. London: Hodder Arnold.
- Williams, J. N. 1992. Processing polysemous words in context: Evidence for interrelated meanings. *Journal of Psycholinguistic Research*. **21**(3), pp.193-218.
- Woodward-Kron, R. W. 2008. More than just jargon: The nature and role of specialist language in learning disciplinary knowledge. *Journal of English for Academic Purposes*. **7**, pp.234-249.
- Wu, N. 吴妮. 2016. Wuliu yingyu rencai zhiyengli xuqiu diaocha yanjiu. Survey on Professional Competence Demand of Logistics English Talents. *Journal of Guangxi Normal University for Nationalities. 广西民族师范学院学报*. **33**(5), pp.131-133.
- Xhaferi, B. 2009. Teaching and learning ESP vocabulary. *LFE: Revista de lenguas para fines especificos*. **15**, pp.229-255.
- Xie, Y. 谢燕. 2022. Guojiwuliu beijing xia wuliuzhichangyingyunengli xuqiu yanjiu. 国际物流背景下物流职场英语能力需求研究. *Wuliupingtai. 物流平台*. **967**, pp.46-48.

- Xie, Z. and Xiong, L. 谢志辉、熊力游. 2012. Wuliuyingyu de cihui tezheng jiqi fanyi. 物流英语的词汇特征及其翻译. *Chinese Science and Technology Translation Journal*. 中国科技翻译. **25**(1), pp.30-32, 46.
- Xodabande, I. and Xodabande, N. 2020. Academic Vocabulary in Psychology Research Articles: A Corpus-Based Study. *MEXTESOL Journal*. **44**(3), pp.1-20.
- Xue, G. and Nation, I. S. P. 1984. A University Word List. *Language Learning and Communication*. **3**(2), pp.215-229.
- Yang, H. 1986. A New Technique for Identifying Scientific/Technical Terms and Describing Science Texts (An Interim Report). *Literary and Linguistic Computing*. **1**(2), pp.93-103.
- Yang, M. N. 2015. A nursing academic word list. *English for specific purposes*. **37**, pp.27-38.
- Zamfir, T. 2022. On Vocabulary Learning Strategies in ESP: A Students' Perspective. *Bulletin of the Transilvania University of Brasov*. Series V: Economic Sciences. pp.97-104.
- Zhang, L. and Wang, Y. H. 张凌凌、王玉华. 2019. Qianxi gaozhigaozhuan wuliuyingyu kecheng gaige. 浅析高中高专“物流英语”课程改革. *Study of Science and Engineering at RTVU*. 电大理工. **280**, pp.53-55.
- Zhao, B. 赵贝. 2018. Gaoxiao zhuanmenyongtuyingyu jiaocai pingjia – yi wuliuyingyu weili. 高校专门用途英语教材评价 – 以《物流英语》为例. *Yingyu guangchang*. 英语广场. **86**, pp.68-70.
- Zhou, Z. 2019. *A Study of English Logistics Textbook from the Perspective of Phraseology*. Master Dissertation, Guangdong University of Foreign Studies.
- Zipf, G. 1949. *Human behaviour and principle of last effort*. Cambridge. MA: Addison-Wesley.



## Appendix A: Function Word List

This word list is used to exclude function words in the corpus. The list (306 items in total) is developed based on West (1953) *The General Service List*, Biber, Johansson, Leech and Conrad (1999) *Longman Grammar of Spoken and Written English*, Brezina and Gablasova (2015) *The New General Service List*, Nation (2022) *Learning Vocabulary in Another Language 3<sup>rd</sup> edition* (Function Word List).

Category	Words				
Others	there	not	so		
Auxiliary verbs (including contracted forms)	am	are	aren't	be	been
	being	can	can't	could	couldn't
	did	didn't	do	does	doesn't
	doing	done	don't	get	gets
	getting	got	had	hadn't	has
	hasn't	have	haven't	having	he'd
	he'll	he's	I'd	I'll	I'm
	is	I've	isn't	it's	may
	might	mist	mustn't	ought	oughtn't
	shall	shan't	she'd	she'll	she's
	should	shouldn't	that's	they'd	they'll
	they're	was	wasn't	we'd	we'll
	were	we're	weren't	we've	will
	won't	would	wouldn't	you'd	you'll
you're	you've				

Conjunctions	after	albeit	although	and	as
	because	before	but	for	how
	however	if	neither	nor	or
	since	so	that	though	till
	unless	until	what	whatever	when
	whenever	where	whereas	wherever	while
	whether	which	who	whoever	whom
	whose	why	yet		
Prepositions	about	above	across	against	along
	among	amongst	around	at	behind
	below	beneath	beside	besides	between
	beyond	by	despite	down	during
	except	from	in	into	like
	minus	near	notwithstanding	of	off
	on	onto	over	per	plus
	round	through	throughout	towards	under
	underneath	unlike	up	upon	via
	with	within	without		
Adverb particles	aside	away	back	forth	home
	out	still	whence		
Determiners	a	all	an	another	any
	both	certain	each	either	enough
	every	few	fewer	half	less

	many	more	most	much	no
	other	others	several	some	such
	the	these	this	those	
Pronouns	he	her	hers	herself	him
	himself	his	I	it	its
	me	mine	my	myself	our
	ours	ourselves	she	their	theirs
	them	themselves	they	us	we
	you	your	yourself	yourselves	anybody
	anyone	anything	everybody	everyone	everything
	nobody	none	nothing	somebody	someone
	something				
Numerals	billion	billionth	eight	eighteen	eighteenth
	eighth	eightieth	eighty	eleven	eleventh
	fifteen	fifteenth	fifth	fiftieth	fifty
	first	five	fortieth	forty	four
	fourteenth	fourth	hundred	hundredth	last
	million	millionth	next	nine	nineteen
	nineteenth	ninetieth	ninety	ninth	once
	one	second	seven	seventeen	seventeenth
	seventh	seventieth	seventy	six	sixteen
	sixteenth	sixth	sixtieth	sixty	ten
	tenth	third	thirteen	thirteenth	thirtieth

	thirty	thousand	thousandth	three	thrice
	twelfth	twelve	twentieth	twenty	twice
	two	zero			

## Appendix B LoSAWEL and Word Sense Identification in the LAWEC, the BAWE and the BNC1994 NAW

Level 1	LAWEC		BAWE		BNC	
Item	Sense/Use	Occurrence (Percentage)	Sense/Use	Occurrence (Percentage)	Sense/Use	Occurrence (Percentage)
ability	the fact that you can do it ( <i>n.</i> )	92.04%	the fact that you can do it ( <i>n.</i> )	100%	the fact that you can do it ( <i>n.</i> )	31.53%
	quality or skills ( <i>n.</i> )	7.96%			quality or skills ( <i>n.</i> )	68.47%
addition	in addition (to) ( <i>ph.</i> )	100%	in addition (to) ( <i>ph.</i> )	100%	in addition (to) ( <i>ph.</i> )	100%
additional	extra things apart from the one presented ( <i>adj.</i> )	100%			extra things apart from the one presented ( <i>adj.</i> )	100%
			in a fixed collocation "additional protocol" ( <i>tm.</i> ) [Law]	100%		
address	try to understand a problem or deal with it ( <i>v.</i> )	100%				
			in a fixed collocation "email address" ( <i>ph.</i> )	100%		
					referring to places ( <i>n.</i> )	100%
allows	create opportunities for something to happen ( <i>v.</i> )	100%	create opportunities for something to happen ( <i>v.</i> )	98.72%		
			to have permission to do something ( <i>v.</i> )	1.82%	to have permission to do something ( <i>v.</i> )	100%
area	a piece of land ( <i>n.</i> )	97.14%	a piece of land ( <i>n.</i> )	100%		
	a particular topic or subject ( <i>n.</i> )	2.86%			a particular topic or subject ( <i>n.</i> )	100%
areas	a piece of land ( <i>n.</i> )	100%	a particular topic or subject ( <i>n.</i> )	100%	a particular topic or subject ( <i>n.</i> )	100%
based	in a fixed collocation "based on" ( <i>adj.</i> )	82.91%	in a fixed collocation "based on/upon" ( <i>adj.</i> )	85.16%	in a fixed collocation "based on/upon" ( <i>adj.</i> )	99.87%

	combined with nouns, in the form of “-based” ( <i>adj.</i> )	14.74%	combined with nouns, in the form of “-based” ( <i>adj.</i> )	14.74%	combined with nouns, in the form of “-based” ( <i>adj.</i> )	0.13%
	the first thing develops from the second ( <i>v.</i> )	2.35%	the first thing develops from the second ( <i>v.</i> )	0.10%		
becomes	something changes gradually ( <i>v.</i> )	100%	something changes gradually ( <i>v.</i> )	100%	something changes gradually ( <i>v.</i> )	100%
business	work relating to the production, buying and selling goods or services ( <i>n.</i> )	100%	work relating to the production, buying and selling goods or services ( <i>n.</i> )	100%	work relating to the production, buying and selling goods or services ( <i>n.</i> )	100%
case	a particular situation used as an example ( <i>n.</i> )	53.48%	a particular situation used as an example ( <i>n.</i> )	71.03%	a particular situation used as an example ( <i>n.</i> )	37.23%
	in a fixed collocation “case study” ( <i>n.</i> )	42.95%	in a fixed collocation “case study” ( <i>n.</i> )	19.23%		
	in a fixed collocation “if this is the case” ( <i>ph.</i> )	2.41%	in a fixed collocation “(if this) is the case” ( <i>ph.</i> )	4.64%		
	in a fixed collocation “in case of” ( <i>ph.</i> )	0.78%	in a fixed collocation “in case of” ( <i>ph.</i> )	2.18%	in a fixed collocation “in case of” ( <i>ph.</i> )	0.79%
	in a fixed collocation “in case” ( <i>ph.</i> )	0.27%	in a fixed collocation “in case” ( <i>ph.</i> )	1.14%	in a fixed collocation “in case” ( <i>ph.</i> )	1.78%
	in a fixed collocation “case pack” ( <i>tm.</i> )	0.08%				
	in a fixed collocation “a case in point” ( <i>ph.</i> )	0.03%				
			in a fixed collocation “in any case” ( <i>ph.</i> )	1.61%		
			law-related issue ( <i>n.</i> )	0.17%	law-related issue ( <i>n.</i> )	0.99%
					in a fixed collocation “in any case” ( <i>ph.</i> )	49.14%
					in a fixed collocation “in the case of” ( <i>ph.</i> )	10.12%
cases	a particular situation used as an example ( <i>n.</i> )	100%	a particular situation used as an example ( <i>n.</i> )	63.51%	a particular situation used as an example ( <i>n.</i> )	94.27%

			a professional situation to deal with ( <i>tm.</i> ) [Med]	36.49%	a professional situation to deal with ( <i>n.</i> )	2.58%
					law-related issue ( <i>n.</i> )	1.68%
certain	referring to a particular thing or people ( <i>adj.</i> )	100%	referring to a particular thing or people ( <i>adj.</i> )	100%	referring to a particular thing or people ( <i>adj.</i> )	99.71%
					to show that something is true ( <i>adj.</i> )	0.29%
changes	make something different ( <i>v.</i> )	100%	make something different ( <i>v.</i> )	91.53%	make something different ( <i>v.</i> )	100%
			climate change ( <i>n.</i> )	8.47%		
companies	business organizations ( <i>n.</i> )	100%	business organizations ( <i>n.</i> )	100%	business organizations ( <i>n.</i> )	100%
company	a business organization ( <i>n.</i> )	100%	a business organization ( <i>n.</i> )	100%	a business organization ( <i>n.</i> )	100%
consider	to think about something carefully ( <i>v.</i> )	100%	to think about something carefully ( <i>v.</i> )	100%	to think about something carefully ( <i>v.</i> )	100%
considered	to think about something carefully ( <i>v.</i> )	100%	to think about something carefully ( <i>v.</i> )	47.44%	to think about something carefully ( <i>v.</i> )	82.14%
			to have an opinion on something ( <i>v.</i> )	52.56%	to have an opinion on something ( <i>v.</i> )	17.86%
considering	to think about something carefully ( <i>v.</i> )	100%	to think about something carefully ( <i>v.</i> )	100%	to think about something carefully ( <i>v.</i> )	100%
control	in a fixed collocation “control variable” ( <i>n.</i> )	50.74%				
	in a fixed collocation “inventory control” ( <i>tm.</i> )	30.54%				
	the power to make all important decisions; in a collocation “control ... over” ( <i>n.</i> )	18.72%			the power to make all important decisions; in a collocation “control ... over” ( <i>n.</i> )	99.70%

			want somebody to do (v.)	28.40%	want somebody to do (v.)	0.30%
			in a fixed collocation “biological control” (tm.) (Biology)	36.85%		
			to have important decisions (v.)	34.75%		
cost	the amount of money to buy or do something (n.)	100%	the amount of money to buy or do something (n.)	98.43%	the amount of money to buy or do something (n.)	100%
			to “spend” money (n.)	1.57%		
degree	the extent of something (n.)	83.60%	the extent of something (n.)	100%		
	In a fixed collocation “degree centrality” (tm.)	16.40%				
					diploma (n.)	100%
difference	the way in which two things are unlike each other (n.)	100%	the way in which two things are unlike each other (n.)	76.55%	the way in which two things are unlike each other (n.)	40.57%
			the quantity of something that is less than others (n.)	23.45%		
					the amount is less (n.)	1.17%
					disagree (n.)	0.38%
					in a fixed collocation “make a difference” (ph.)	57.88%
different	two things that are not like each other in one or more ways (adj.)	100%	two things that are not like each other in one or more ways (adj.)	93.73%	two things that are not like each other in one or more ways (adj.)	100%
			to describe something that is unusual (adj.)	6.27%		
expected	believe something will happen (v.)	100%	believe something will happen (v.)	100%	believe something will happen (v.)	100%
figure	diagrams to show information (n.)	99.99%	diagrams to show information (n.)	99.93%	diagrams to show information (n.)	98.00%

	a particular number or statistics ( <i>n.</i> )	0.01%	a particular number or statistics ( <i>n.</i> )	0.07%	a particular number or statistics ( <i>n.</i> )	1.10%
					referring to “people” ( <i>n.</i> )	0.90%
findings	research results ( <i>n.</i> )	100%	research results ( <i>n.</i> )	100%	research results ( <i>n.</i> )	99.60%
					the court’s decision after investigation ( <i>n.</i> )	0.40%
finally	in speech or writing to introduce the last point ( <i>adv.</i> )	100%	in speech or writing to introduce the last point ( <i>adv.</i> )	100%	in speech or writing to introduce the last point ( <i>adv.</i> )	0.64%
					later than in expectation ( <i>adv.</i> )	99.36%
following	the thing about to mention ( <i>adj.</i> )	97.34%	the thing about to mention ( <i>adj.</i> )	100%	the thing about to mention ( <i>adj.</i> )	9.39%
	after that event ( <i>prep.</i> )	2.27%			after that event ( <i>prep.</i> )	19.23%
	act the way after someone’s advice ( <i>v.</i> )	0.39%			act the way after someone’s advice ( <i>v.</i> )	0.52%
					the next day/year ( <i>adj.</i> )	70.48%
					one after another ( <i>v.</i> )	0.24%
					go after with someone ( <i>v.</i> )	0.10%
					follow the path ( <i>v.</i> )	0.01%
					take the interest ( <i>v.</i> )	0.06%
follows	one after another ( <i>v.</i> )	100%	one after another ( <i>v.</i> )	8.49%	one after another ( <i>v.</i> )	100%
			in a fixed collocation “as follows” ( <i>ph.</i> )	91.51%		
found	see or learn something ( <i>v.</i> )	98.96%	see or learn something ( <i>v.</i> )	59.27%	see or learn something ( <i>v.</i> )	1.17%
	usually in passive voice, to show something exists ( <i>v.</i> )	1.04%	usually in passive voice, to show something exists ( <i>v.</i> )	40.73%		
					notice someone in a particular situation ( <i>v.</i> )	96.91%

					to realise something (v.)	1.81%
					successfully to have time (v.)	0.11%
further	something additional (adj.)	86.50%	something additional (adj.)	90.68%	something additional (adj.)	99.65%
	to introduce some related topics (adv.)	6.75%	to introduce some related topics (adv.)	8.00%	to introduce some related topics (adv.)	0.17%
	greater extent or degree (adv.)	6.75%	greater extent or degree (adv.)	1.32%	greater extent or degree (adv.)	0.18%
general	in a fixed collocation "in general" (ph.)	69.61%				
	to talk about something without much detail (adj.)	30.13%				
	to describe something is quite common (adj.)	0.26%				
			to involve a lot of people (adj.)	100%	to involve a lot of people (adj.)	61.93%
					general secretary (n.)	37.23%
					an officer in the army (n.)	0.84%
generally	to give a summary of a situation (adv.)	100%	to give a summary of a situation (adv.)	100%	to give a summary of a situation (adv.)	100%
greater	larger in number (adj.)	100%	larger in number (adj.)	80.93%	larger in number (adj.)	96.89%
			to describe something that is more important (adj.)	19.07%	to describe something that is more important (adj.)	1.00%
					to describe some place around a city (adj.)	1.17%
					successful (adj.)	0.30%
					larger in size (adj.)	0.40%
high	greater amount and intensity (adj.)	47.15%	greater amount and intensity (adj.)	99.82%	greater amount and intensity (adj.)	44.66%
	good in quality (adj.)	28.39%			good in quality (adj.)	25.88%

	in priority ( <i>adj.</i> )	24.45%			in priority ( <i>adj.</i> )	0.06%
			surface of water ( <i>adj.</i> )	0.18%		
					good reputation ( <i>adj.</i> )	0.03%
					to modify sea ground level ( <i>adj.</i> )	0.30%
					high street ( <i>n.</i> )	21.67%
					have great influence on something ( <i>adj.</i> )	7.40%
higher	greater in number and intensity ( <i>adj.</i> )	84.91%	greater in number and intensity ( <i>adj.</i> )	95.51%	greater in number and intensity ( <i>adj.</i> )	64.13%
	good in quality ( <i>adj.</i> )	15.09%			good in quality ( <i>adj.</i> )	0.31%
			to describe something that is more important ( <i>adj.</i> )	2.25%	to describe something that is more important ( <i>adj.</i> )	0.68%
			to describe something that has more ingredient ( <i>adj.</i> )	1.22%		
			surface of water ( <i>adj.</i> )	1.02%		
					to describe something that is advanced or complex ( <i>adj.</i> )	34.73%
					to describe sound volume ( <i>adj.</i> )	0.02%
					to describe aims ( <i>adj.</i> )	0.05%
					to describe emotions ( <i>adj.</i> )	0.05%
					to describe social class ( <i>adj.</i> )	0.05%
highly	very ( <i>adv.</i> )	100%	very ( <i>adv.</i> )	100%	very ( <i>adv.</i> )	100%
important	vital ( <i>adj.</i> )	100%	vital ( <i>adj.</i> )	100%	vital ( <i>adj.</i> )	100%
information	something consisting of facts ( <i>n.</i> )	86.57%	something consisting of facts ( <i>n.</i> )	99.67%	something consisting of facts ( <i>n.</i> )	79.94%
	in a fixed collocation “information technology” ( <i>n.</i> )	5.67%			in a fixed collocation “information technology” ( <i>n.</i> )	20.06%

	facts and pictures stored in computers ( <i>n.</i> ); in a fixed collocation “information flow”	4.26%	facts and pictures stored in computers ( <i>n.</i> )	0.33%		
	in a fixed collocation “asymmetric information” ( <i>tm.</i> )	3.50%				
issues	something that is worth discussing ( <i>n.</i> )	100%	something that is worth discussing ( <i>n.</i> )	100%	something that is worth discussing ( <i>n.</i> )	100%
key	the most important one ( <i>adj.</i> )	100%	the most important one ( <i>adj.</i> )	100%		
					special metal for locks ( <i>n.</i> )	100%
keywords	important words in journal article abstracts ( <i>n.</i> )	100%	important words in journal article abstracts ( <i>n.</i> )	100%	important words in journal article abstracts ( <i>n.</i> )	100%
large	to describe number or size ( <i>adj.</i> )	67.62%	to describe number or size ( <i>adj.</i> )	100%	to describe number or size ( <i>adj.</i> )	99.31%
	to describe severity of problem ( <i>adj.</i> )	32.38%			to describe severity of problem ( <i>adj.</i> )	0.09%
					to modify business ( <i>adj.</i> )	0.59%
larger	to describe number or size ( <i>adj.</i> )	99.29%	to describe number or size ( <i>adj.</i> )	100%	to describe number or size ( <i>adj.</i> )	92.56%
	to describe importance of something ( <i>adj.</i> )	0.71%			to describe importance of something ( <i>adj.</i> )	0.32%
					to describe business ( <i>adj.</i> )	7.12%
lead	in a fixed collocation “lead time(s)” ( <i>tm.</i> )	72.55%				
	to cause something to happen ( <i>v.</i> ); in a fixed collocation “lead to”	27.45%	to cause something to happen ( <i>v.</i> ); in a fixed collocation “lead to”	95.44%	to cause something to happen ( <i>v.</i> ); in a fixed collocation “lead to”	36.83%
			to introduce to a topic ( <i>v.</i> )	3.13%		
			to take something in charge ( <i>v.</i> )	1.00%		

			a chemical component ( <i>n.</i> )	0.43%	a chemical component ( <i>n.</i> )	12.65%
					to guide someone ( <i>v.</i> )	2.15%
					as the key/first one ( <i>n.</i> )	25.48%
					in a fixed collocation “take the lead” ( <i>ph.</i> )	22.89%
leads	to cause something to happen ( <i>v.</i> ); in a fixed collocation “lead to”	100%	to cause something to happen ( <i>v.</i> ); in a fixed collocation “lead to”	85.12%	to cause something to happen ( <i>v.</i> ); in a fixed collocation “lead to”	29.45%
			to introduce a topic ( <i>v.</i> )	14.88%	in a fixed collocation “leads and lags” ( <i>n.</i> )	6.16%
					to guide someone or something ( <i>v.</i> )	64.38%
level	referring to “scale” to measure something ( <i>n.</i> )	100%	referring to “scale” to measure something ( <i>n.</i> )	99.46%	referring to “scale” to measure something ( <i>n.</i> )	99.04%
			surface of water ( <i>n.</i> )	0.54%	surface of water ( <i>n.</i> )	0.96%
levels	referring to “scale” to measure something ( <i>n.</i> )	100%	referring to “scale” to measure something ( <i>n.</i> )	99.49%	referring to “scale” to measure something ( <i>n.</i> )	97.45%
			surface of water ( <i>n.</i> )	0.51%	surface of water ( <i>n.</i> )	2.55%
low	to describe number or size ( <i>adj.</i> )	83.72%	to describe number or size ( <i>adj.</i> )	89.49%	to describe number or size ( <i>adj.</i> )	68.93%
	to describe poor quality ( <i>adj.</i> )	16.28%			to describe poor quality ( <i>adj.</i> )	1.26%
			surface of water ( <i>adj.</i> )	0.34%	surface of water ( <i>adj.</i> )	1.46%
			to describe something that is less important ( <i>adj.</i> )	0.17%	to describe something that is less important ( <i>adj.</i> )	0.20%
					to modify sound volume ( <i>adj.</i> )	21.11%
					to describe social class ( <i>adj.</i> )	0.25%
					to describe “aims” ( <i>adj.</i> )	0.20%

					to describe height ( <i>adj.</i> )	3.97%
					in a fixed collocation “high and low” ( <i>ph.</i> )	0.15%
					to describe emotions ( <i>adj.</i> )	0.05%
					to show something has less ingredients ( <i>adj.</i> )	2.26%
main	to describe the most important thing ( <i>adj.</i> )	100%	to describe the most important thing ( <i>adj.</i> )	100%	to describe the most important thing ( <i>adj.</i> )	100%
mainly	to describe most of something ( <i>adv.</i> )	100%	to describe most of something ( <i>adv.</i> )	100%	to describe most of something ( <i>adv.</i> )	45.55%
					to describe the main reason ( <i>adv.</i> )	54.45%
major	to describe something that is more important ( <i>adj.</i> )	100%	to describe something that is more important ( <i>adj.</i> )	100%	to describe something that is more important ( <i>adj.</i> )	1.25%
					British Army Official ( <i>n.</i> )	0.60%
					as a surname ( <i>n.</i> )	98.15%
management	to control business organization ( <i>n.</i> )	100%	to control business organization ( <i>n.</i> )	100%	to control business organization ( <i>n.</i> )	100%
managerial	related to work of a manager ( <i>adj.</i> )	100%	related to work of a manager ( <i>adj.</i> )	100%	related to work of a manager ( <i>adj.</i> )	95.16%
					in a fixed collocation “managerial grid” ( <i>ph.</i> )	4.84%
managers	people who control the business ( <i>n.</i> )	100%	people who control the business ( <i>n.</i> )	45.28%	people who control the business ( <i>n.</i> )	100%
					in a fixed collocation “expatriate manager” ( <i>tm.</i> ) (Hospitality)	54.72%
market	referring to how many people would buy the product ( <i>n.</i> )	35.45%	referring to how many people would buy the product ( <i>n.</i> )	19.16%	referring to how many people would buy the product ( <i>n.</i> )	11.55%
	in a fixed collocation “market share” ( <i>tm.</i> )	31.96%	in a fixed collocation “market share” ( <i>tm.</i> )	46.33%	in a fixed collocation “market share” ( <i>tm.</i> )	31.60%

	in a fixed collocation “spot market” ( <i>tm.</i> )	11.79%				
	goods in production ( <i>n.</i> )	11.58%	goods in production ( <i>n.</i> )	2.88%		
	in a fixed collocation “market power” ( <i>tm.</i> )	9.21%				
			in a fixed collocation “labour market” ( <i>n.</i> )	31.63%		
					referring to “people who want job” ( <i>n.</i> )	29.36%
					referring to stock market ( <i>n.</i> )	31.60%
means	to indicate something ( <i>v.</i> )	78.69%	to indicate something ( <i>v.</i> )	98.71%	to indicate something ( <i>v.</i> )	61.85%
	in a fixed collocation “means objectives” ( <i>tm.</i> )	20.62%				
	in a collocation “this (is) by no means” ( <i>ph.</i> )	0.34%	in a collocation “this (is) by no means” ( <i>ph.</i> )	0.74%	in a collocation “this (is) by no means” ( <i>ph.</i> )	9.51%
	to describe methods to do something ( <i>n.</i> )	0.34%	to describe methods to do something ( <i>n.</i> )	0.55%	to describe methods to do something ( <i>n.</i> )	11.28%
					the first thing leads to the second thing to happen ( <i>v.</i> )	13.59%
					intend to do something (mean to do something) ( <i>v.</i> )	0.10%
					in a fixed collocation “by means of” ( <i>ph.</i> )	3.67%
nature	the feature of something ( <i>n.</i> )	100%	the feature of something ( <i>n.</i> )	96.85%	the feature of something ( <i>n.</i> )	0.17%
			referring to the natural world or environment ( <i>n.</i> )	3.15%	referring to the natural world or environment ( <i>n.</i> )	73.52%
					human nature ( <i>n.</i> )	25.98%
					the nature of someone (to show	0.33%

					how they behave) ( <i>n.</i> )	
necessary	to describe something is needed ( <i>adj.</i> )	100%	to describe something is needed ( <i>adj.</i> )	89.25%	to describe something is needed ( <i>adj.</i> )	100%
			in a fixed collocation “if necessary” ( <i>ph.</i> )	10.75%		
note			to write something down ( <i>v.</i> )	0.26%	to write something down ( <i>v.</i> )	0.49%
	to have attention to something ( <i>v.</i> )	100%	to have attention to something ( <i>v.</i> )	74.67%	to have attention to something ( <i>v.</i> )	96.06%
			in a fixed collocation “additional note” ( <i>tm.</i> ) [Law]	25.07%		
					something for remind ( <i>n.</i> )	2.96%
					referring to additional information ( <i>n.</i> )	0.49%
number	amount/size ( <i>n.</i> )	89.97%	amount/size ( <i>n.</i> )	49.88%	amount/size ( <i>n.</i> )	99.39%
	in a fixed collocation “a number of” ( <i>n.</i> )	10.03%				
			in a fixed collocation “Reynolds number” ( <i>tm.</i> ) [Engineering]	30.72%		
					a specific number ( <i>n.</i> )	0.61%
			a series of number ( <i>n.</i> )	19.40%		
order	to request something ( <i>v.</i> )	49.46%	to request something ( <i>v.</i> )	0.23%		
	in a fixed collocation “in order to” ( <i>ph.</i> )	37.68%	in a fixed collocation “in order to” ( <i>ph.</i> )	96.47%		
	in a fixed collocation “order fulfilment” ( <i>tm.</i> )	6.24%				
	in a fixed collocation “order picking” ( <i>tm.</i> )	5.42%				

	in a particular sequence ( <i>n.</i> )	0.85%	in a particular sequence ( <i>n.</i> )	0.85%		
	to book something in advance ( <i>v.</i> )	0.85%	to book something in advance ( <i>v.</i> )	0.61%		
			a situation that exists when everything is in correct order ( <i>n.</i> )	1.14%		
			the legal instruction ( <i>n.</i> )	0.70%		
					used in parliament debate ( <i>n.</i> )	100%
paper	research report ( <i>n.</i> )	99.71%	research report ( <i>n.</i> )	98.27%		
	material ( <i>n.</i> )	0.29%	material ( <i>n.</i> )	1.73%	material ( <i>n.</i> )	63.39%
					official report ( <i>n.</i> )	36.31%
particular	in a fixed collocation “in particular” ( <i>ph.</i> )	99.54%	in a fixed collocation “in particular” ( <i>ph.</i> )	43.77%	in a fixed collocation “in particular” ( <i>ph.</i> )	10.59%
	to describe something that is special ( <i>adj.</i> )	0.46%	to describe something that is special ( <i>adj.</i> )	37.90%	to describe something that is special ( <i>adj.</i> )	89.41%
			to describe something is more intense than usual ( <i>adj.</i> )	18.33%		
particularly	apply to a specific situation ( <i>adv.</i> )	100%	apply to a specific situation ( <i>adv.</i> )	32.51%	apply to a specific situation ( <i>adv.</i> )	29.91%
			more than usual ( <i>adv.</i> )	67.49%	more than usual ( <i>adv.</i> )	70.09%
possible	to describe something is likely to happen ( <i>adj.</i> )	100%	to describe something is likely to happen ( <i>adj.</i> )	100%	to describe something is likely to happen ( <i>adj.</i> )	100%
price	amount of money ( <i>n.</i> )	100%	amount of money ( <i>n.</i> )	100%	amount of money ( <i>n.</i> )	100%
present	to show something ( <i>v.</i> )	54.24%	to show something ( <i>v.</i> )	1.50%		
	to describe something that is in current situation ( <i>adj.</i> )	45.76%	to describe something that is in current situation ( <i>adj.</i> )	83.23%	to describe something that is in current situation ( <i>adj.</i> )	11.07%
					in a fixed collocation “at present” ( <i>ph.</i> )	88.93%

			as a term in language ( <i>tm.</i> ) [Language]	15.27%		
presented	to show something ( <i>v.</i> )	100%	to show something ( <i>v.</i> )	100%	to show something ( <i>v.</i> )	57.36%
					to give something to someone ( <i>v.</i> )	42.64%
problem	an unsatisfied situation ( <i>n.</i> )	92.66%	an unsatisfied situation ( <i>n.</i> )	100%	an unsatisfied situation ( <i>n.</i> )	100%
	in a fixed collocation “problem instance” ( <i>tm.</i> )	4.07%				
	in a fixed collocation “routing problem(s)” ( <i>tm.</i> )	3.27%				
problems	an unsatisfied situation ( <i>n.</i> )	88.19%	an unsatisfied situation ( <i>n.</i> )	100%	an unsatisfied situation ( <i>n.</i> )	100%
	in a fixed collocation “routing problem(s)” ( <i>tm.</i> )	11.81%				
rate	to describe a number in a specific period ( <i>n.</i> )	85.22%	to describe a number in a specific period ( <i>n.</i> )	12.74%	to describe a number in a specific period ( <i>n.</i> )	32.94%
	referring to “speed” ( <i>n.</i> )	14.78%	referring to “speed” ( <i>n.</i> )	54.73%	referring to “speed” ( <i>n.</i> )	21.69%
			amount of money ( <i>n.</i> )	32.10%	amount of money ( <i>n.</i> )	45.29%
			in a fixed collocation “at any rate” ( <i>ph.</i> )	0.43%		
					to evaluate something ( <i>v.</i> )	0.07%
recent	a time near present ( <i>adj.</i> )	100%	a time near present ( <i>adj.</i> )	100%	a time near present ( <i>adj.</i> )	100%
related	a connection between two things ( <i>adj.</i> )	73.70%	a connection between two things ( <i>adj.</i> )	100%		
	two things are derived from one ( <i>adj.</i> ); in a fixed collocation “be related to”	26.30%			two things are derived from one ( <i>adj.</i> ); in a fixed collocation “be related to”	100%

service	a particular work ( <i>n.</i> )	100%	a particular work ( <i>n.</i> )	88.95%		
			part of economy ( <i>n.</i> )	7.84%		
			in a fixed collocation “National Health Service” ( <i>tm.</i> )	3.21%		
					public needs ( <i>n.</i> )	100%
services	a particular work ( <i>n.</i> )	100%	a particular work ( <i>n.</i> )	72.29%	a particular work ( <i>n.</i> )	43.52%
			public needs ( <i>n.</i> )	27.41%	public needs ( <i>n.</i> )	56.48%
			part of economy ( <i>n.</i> )	0.30%		
set	in a fixed collocation “data set” ( <i>ph.</i> )	42.55%	in a fixed collocation “data set” ( <i>ph.</i> )	19.62%		
	to refer to a group of people/things ( <i>n.</i> )	30.03%				
	in a fixed collocation “set up” to describe “create something” ( <i>ph.</i> )	22.62%	in a fixed collocation “set up” to describe “create something” ( <i>ph.</i> )	57.75%	in a fixed collocation “set up” to describe “create something” ( <i>ph.</i> )	19.70%
	to decide price/date/goal ( <i>v.</i> )	4.80%				
			to construct or build something ( <i>v.</i> )	0.13%		
			in a fixed collocation “set out”, to explain something ( <i>ph.</i> )	5.63%	in a fixed collocation “set out”, to explain something ( <i>ph.</i> )	9.64%
			in a fixed collocation “set out”, try to do ( <i>ph.</i> )	6.50%	in a fixed collocation “set out”, try to do ( <i>ph.</i> )	9.64%
			in a fixed collocation “display” ( <i>ph.</i> )	10.00%		
			in a fixed collocation “set out”, start a journey” ( <i>ph.</i> )	0.37%	in a fixed collocation “set out”, start a journey” ( <i>ph.</i> )	9.64%
					in a fixed collocation “set up” for emotion ( <i>ph.</i> )	0.33%
					in a fixed collocation “set off” to mean “happen” ( <i>ph.</i> )	5.72%

					in a fixed collocation “set off” for journey ( <i>ph.</i> )	8.75%
					in a fixed collocation “set off” for alarms ( <i>ph.</i> )	5.33%
					in a fixed collocation “set off” for contrast ( <i>ph.</i> )	3.64%
					in a fixed collocation “set off” for calming down ( <i>ph.</i> )	0.78%
					in a fixed collocation “set off” for cancelling payment ( <i>ph.</i> )	0.65%
					in a fixed collocation “set off” for turning off ( <i>ph.</i> )	0.26%
					in a fixed collocation “set off” for making a balance ( <i>ph.</i> )	2.86%
					in a fixed collocation “set aside” ( <i>ph.</i> )	23.05%
several	referring to “some” ( <i>adj.</i> )	100%	referring to “some” ( <i>adj.</i> )	100%	referring to “some” ( <i>adj.</i> )	100%
show	to present something ( <i>v.</i> )	99.96%	to present something ( <i>v.</i> )	97.76%	to present something ( <i>v.</i> )	67.37%
	as a performance ( <i>n.</i> )	0.02%			as a performance ( <i>n.</i> )	17.78%
	as a behaviour ( <i>v.</i> )	0.02%			as a behaviour ( <i>v.</i> )	3.55%
			to display ( <i>v.</i> )	2.24%	to display ( <i>v.</i> )	3.16%
					lead someone to somewhere ( <i>v.</i> )	0.43%
					do something and watch ( <i>v.</i> )	4.93%
					public exhibition ( <i>n.</i> )	27.78%
shows	to present something ( <i>v.</i> )	99.97%	to present something ( <i>v.</i> )	99.04%	to present something ( <i>v.</i> )	100%

	as a performance ( <i>n.</i> )	0.03%				
			to display ( <i>v.</i> )	0.96%		
shown	to present something ( <i>v.</i> )	100%	to present something ( <i>v.</i> )	99.94%	to present something ( <i>v.</i> )	100%
			to display ( <i>v.</i> )	0.06%		
single	referring to “one” ( <i>adj.</i> )	100%	referring to “one” ( <i>adj.</i> )	49.48%	referring to “one” ( <i>adj.</i> )	100%
			in a fixed collocation “EU Single Market” ( <i>tm.</i> ) (Business)	50.52%		
situation	referring to what is happening in a specific time ( <i>n.</i> )	100%	referring to what is happening in a specific time ( <i>n.</i> )	100%	referring to what is happening in a specific time ( <i>n.</i> )	100%
situations	referring to what is happening in a specific time ( <i>n.</i> )	100%	referring to what is happening in a specific time ( <i>n.</i> )	100%	referring to what is happening in a specific time ( <i>n.</i> )	100%
size	to describe how big or small a thing is ( <i>n.</i> )	59.76%	to describe how big or small a thing is ( <i>n.</i> )	100%	to describe how big or small a thing is ( <i>n.</i> )	100%
	as a graded measure ( <i>n.</i> )	24.48%				
	in a fixed collocation “lot size” ( <i>tm.</i> )	15.76%				
state	to describe the situation of something ( <i>n.</i> )	75.48%	to describe the situation of something ( <i>n.</i> )	27.97%		
	in a fixed collocation “state space” ( <i>tm.</i> )	24.52%				
			referring to country or a region ( <i>n.</i> )	48.25%	referring to country or a region ( <i>n.</i> )	100%
			in a fixed collocation “state intervention” ( <i>tm.</i> ) [Politics/Economics]	23.78%		
step	to refer to a series of actions ( <i>n.</i> )	80.62%	to refer to a series of actions ( <i>n.</i> )	89.53%	to refer to a series of actions ( <i>n.</i> )	17.35%
	to refers to stages of development of something ( <i>n.</i> )	19.38%			to refers to stages of development of something ( <i>n.</i> )	60.88%

			in a fixed collocation “step by step” ( <i>ph.</i> )	10.47%	in a fixed collocation “step by step” ( <i>ph.</i> )	21.77%
studied	to refer to the action of doing research ( <i>v.</i> )	100%	to refer to the action of doing research ( <i>v.</i> )	100%	to refer to the action of doing research ( <i>v.</i> )	100%
studies	to refer to a piece of research ( <i>n.</i> )	94.98%	to refer to a piece of research ( <i>n.</i> )	100%	to refer to a piece of research ( <i>n.</i> )	100%
	in a fixed collocation “case studies” ( <i>n.</i> )	5.02%				
study	to refer to a piece of research ( <i>n.</i> )	82.60%	to refer to a piece of research ( <i>n.</i> )	62.71%	to refer to a piece of research ( <i>n.</i> )	51.95%
	in a fixed collocation “case study” ( <i>n.</i> )	11.60%	in a fixed collocation “case study” ( <i>n.</i> )	37.28%		
	to refer to the action of doing research ( <i>v.</i> )	5.80%			to refer to the action of doing research ( <i>v.</i> )	8.70%
					to spend time learning something ( <i>v.</i> )	39.02%
					a room for reading, writing and studying ( <i>n.</i> )	0.34%
suggest	to imply something ( <i>v.</i> )	99.44%	to imply something ( <i>v.</i> )	100%	to imply something ( <i>v.</i> )	77.52%
	to show that you believe something is true ( <i>v.</i> )	0.56%			to show that you believe something is true ( <i>v.</i> )	1.34%
					to put forward something ( <i>v.</i> )	21.14%
suggested	to imply something ( <i>v.</i> )	100%	to imply something ( <i>v.</i> )	86.12%	to imply something ( <i>v.</i> )	46%
			to believe something that is true ( <i>v.</i> )	13.12%	to believe something that is true ( <i>v.</i> )	4.32%
			to recommend ( <i>v.</i> )	0.76%		
					to put forward something like a plan ( <i>v.</i> )	49.67%
suggests	to imply something ( <i>v.</i> )	100%	to imply something ( <i>v.</i> )	100%	to imply something ( <i>v.</i> )	99.55%

					to put forward something like a plan (v.)	0.45%
system	to refer to the framework or structure of a complex thing (n.)	100%	to refer to the framework or structure of a complex thing (n.)	100%	to refer to the framework or structure of a complex thing (n.)	0.73%
					to refer to administration (n.)	0.95%
					used in electronic device (n.)	98.32%
systems	to refer to the framework or structure of a complex thing (n.)	100%	to refer to the framework or structure of a complex thing (n.)	100%	to refer to the framework or structure of a complex thing (n.)	16.17%
					used in electronic device (n.)	56.72%
					used in cooperation name (n.)	27.11%
table	chart (n.)	99.99%	chart (n.)	91.47%	chart (n.)	47.61%
	in a fixed collocation "table wine" (n.)	0.01%				
					furniture (n.)	52.39%
			in a fixed collocation "table of content(s)" (ph.)	8.53%		
terms	in a fixed collocation "in terms of" (ph.)	97.01%	in a fixed collocation "in terms of" (ph.)	72.06%	in a fixed collocation "in terms of" (ph.)	28.26%
	to refers to technical words (n.)	2.24%	to refers to technical words (n.)	19.83%	to refers to technical words (n.)	6.81%
	in a fixed collocation "in practical/general terms" (ph.)	0.75%	in a fixed collocation "in practical/general terms" (ph.)	7.84%		
			a period of time (n.)	0.25%		
					agreement and conditions (n.)	16.81%

					in a fixed collocation “terms of reference” ( <i>n.</i> )	21.10%
					period of learning at school ( <i>n.</i> )	0.01%
					in a fixed collocation “come to terms” ( <i>ph.</i> )	27.35%
test	in order to find what condition it is ( <i>v.</i> )	56.95%	in order to find what condition it is ( <i>v.</i> )	4.32%	in order to find what condition it is ( <i>v.</i> )	0.31%
	a deliberate action or experiment to know whether it work or not ( <i>n.</i> )	43.05%	a deliberate action or experiment to know whether it work or not ( <i>n.</i> )	95.68%	a deliberate action or experiment to know whether it work or not ( <i>n.</i> )	54.76%
					sport match ( <i>n.</i> )	34.99%
					body test ( <i>n.</i> )	0.78%
					to reveal quality of something, usually in “a test of” ( <i>n.</i> )	0.16%
total	the number or cost of all parts ( <i>adj.</i> )	100%	the number or cost of all parts ( <i>adj.</i> )	100%	the number or cost of all parts ( <i>adj.</i> )	100%
type	to refer to a particular kind ( <i>n.</i> )	100%	to refer to a particular kind ( <i>n.</i> )	100%	to refer to a particular kind ( <i>n.</i> )	100%
types	to refer to a particular kind ( <i>n.</i> )	100%	to refer to a particular kind ( <i>n.</i> )	100%	to refer to a particular kind ( <i>n.</i> )	100%
understanding	to know how it works ( <i>n.</i> )	84.79%	to know how it works ( <i>n.</i> )	99.63%	to know how it works ( <i>n.</i> )	80.47%
			being kind and forgiving ( <i>n.</i> )	0.37%		
	to know why or how something happens ( <i>v.</i> )	15.21%				
					informal agreement ( <i>n.</i> )	19.34%
					referring to trust ( <i>n.</i> )	0.19%
use	to achieve a particular result ( <i>v.</i> )	84.39%	to achieve a particular result ( <i>v.</i> )	43.56%	to achieve a particular result ( <i>v.</i> )	46.20%

	something has a particular way to be used ( <i>n.</i> )	15.61%	something has a particular way to be used ( <i>n.</i> )	56.44%	something has a particular way to be used ( <i>n.</i> )	42.65%
					in a fixed collocation “make use of” ( <i>ph.</i> )	11.15%
used	to achieve a particular result ( <i>v.</i> )	99.97%	to achieve a particular result ( <i>v.</i> )	99.85%	to achieve a particular result ( <i>v.</i> )	76.87%
	in a fixed collocation “used to be” ( <i>ph.</i> )	0.03%	in a fixed collocation “used to be” ( <i>ph.</i> )	0.15%	in a fixed collocation “used to be” ( <i>ph.</i> )	23.13%
using	to achieve a particular result ( <i>v.</i> )	100%	to achieve a particular result ( <i>v.</i> )	100%	to achieve a particular result ( <i>v.</i> )	78.48%
					something has a particular way to be used ( <i>n.</i> )	21.52%

Level 2	LAWEC		BAWE		BNC1994	
Item	Sense/Use	Occurrence (Percentage)	Sense/Use	Occurrence (Percentage)	Sense/Use	Occurrence (Percentage)
according	in a fixed collocation “according to” ( <i>ph.</i> )	100%	in a fixed collocation “according to” ( <i>ph.</i> )	100%	in a fixed collocation “according to” ( <i>ph.</i> )	100%
account	in a fixed collocation “take ... into account” ( <i>ph.</i> )	99.91%	in a fixed collocation “take ... into account” ( <i>ph.</i> )	99.51%	in a fixed collocation “take ... into account” ( <i>ph.</i> )	99.99%
	in a fixed collocation “interest bearing account” ( <i>n.</i> ); a type of banking account	0.09%	a type of banking account ( <i>n.</i> )	0.28%		
			a spoken or written report ( <i>n.</i> )	0.14%		
			other accounts in business ( <i>n.</i> )	0.07%	other accounts in business ( <i>n.</i> )	0.01%
activities	events ( <i>n.</i> )	100%	events ( <i>n.</i> )	100%	events ( <i>n.</i> )	100%
affect	to lead to changes of something ( <i>v.</i> )	100%	to lead to changes of something ( <i>v.</i> )	100%	to lead to changes of something ( <i>v.</i> )	100%
applied	to use rules/skills in an activity ( <i>v.</i> )	100%	to use rules/skills in an activity ( <i>v.</i> )	83.60%	to use rules/skills in an activity ( <i>v.</i> )	34.30%
			the subject or study has a practical use ( <i>adj.</i> )	16.40%	the subject or study has a practical use ( <i>adj.</i> )	26.09%
					ask formally ( <i>v.</i> )	21.26%
					be relevant to someone or something ( <i>v.</i> ) (apply to)	18.36%
application	referring to “software” ( <i>n.</i> )	63.04%	referring to “software” ( <i>n.</i> )	58.92%	referring to “software” ( <i>n.</i> )	99.67%
	referring to “use” ( <i>n.</i> )	26.96%	referring to “use” ( <i>n.</i> )	41.07%	referring to “use” ( <i>n.</i> )	0.33%
approach	method ( <i>n.</i> )	100%	method ( <i>n.</i> )	100%	method ( <i>n.</i> )	99.85%
					get close to somewhere ( <i>v.</i> )	0.15%
approaches	method ( <i>n.</i> )	100%	method ( <i>n.</i> )	100%	method ( <i>n.</i> )	86.47%
					get close to somewhere ( <i>v.</i> )	13.53%

associated	in a fixed collocation “be associated with” ( <i>ph.</i> )	87.73%	in a fixed collocation “be associated with” ( <i>ph.</i> )	100%	in a fixed collocation “be associated with” ( <i>ph.</i> )	100%
	referring to “jointly” ( <i>v.</i> )	12.27%				
assume	to think something as a true thing ( <i>v.</i> )	100%	to think something as a true thing ( <i>v.</i> )	100%	to think something as a true thing ( <i>v.</i> )	54.44%
					to take responsibility ( <i>v.</i> )	25.81%
					in a fixed collocation “let us assume” ( <i>ph.</i> )	19.675%
assumed	to think something as a true thing ( <i>v.</i> )	99.28%	to think something as a true thing ( <i>v.</i> )	100%	to think something as a true thing ( <i>v.</i> )	43.27%
	to begin to behave like some way ( <i>v.</i> )	0.72%			to begin to behave like some way ( <i>v.</i> )	13.35%
					to take responsibility ( <i>v.</i> )	43.27%
average	referring to “mean” number ( <i>adj.</i> )	100%	referring to “mean” number ( <i>adj.</i> )	100%	referring to “mean” number ( <i>adj.</i> )	100%
benefits	the help you get from something or the advantage results from it ( <i>n.</i> )	58.47%	the help you get from something or the advantage results from it ( <i>n.</i> )	24.24%	the help you get from something or the advantage results from it ( <i>n.</i> )	31.40%
	non-wage compensation provided by employers to their workers ( <i>n.</i> )	41.53%	non-wage compensation provided by employers to their workers ( <i>n.</i> )	49.49%	non-wage compensation provided by employers to their workers ( <i>n.</i> )	68.46%
			referring to money provided by governments ( <i>n.</i> )	26.27%		
					receiving helps ( <i>v.</i> )	0.14%
chain	in a fixed collocation “supply chain (management)” ( <i>tm.</i> )	94.56%	in a fixed collocation “supply chain (management)” ( <i>tm.</i> )	25.91%		
	to describe a series of actions/objects ( <i>n.</i> )	4.88%	to describe a series of actions/objects ( <i>n.</i> )	27.91%	to describe a series of actions/objects ( <i>n.</i> )	49.43%

	in a fixed collocation “cold chain logistics” ( <i>tm.</i> )	0.56%				
			in a fixed collocation “polymerase chain reaction” ( <i>tm.</i> ) [Biology]	33.22%		
			in a fixed collocation “side chain” ( <i>tm.</i> ) [Biology]	8.63%		
			in a fixed collocation “the chain of causation” ( <i>tm.</i> ) [Law]	4.32%		
					chain store ( <i>n.</i> )	27.29%
					metal ( <i>n.</i> )	23.28%
common	to describe something which can be found in large numbers ( <i>adj.</i> )	59.31%	to describe something which can be found in large numbers ( <i>adj.</i> )	0.52%	to describe something which can be found in large numbers ( <i>adj.</i> )	0.58%
	in a fixed collocation “common method bias” ( <i>tm.</i> )	23.53%				
	in a fixed collocation “common agency” ( <i>tm.</i> )	9.64%				
	in a fixed collocation “common method variance” ( <i>tm.</i> )	7.52%				
			in a fixed collocation “common law” ( <i>n.</i> )	34.79%		
			something shared in general ( <i>adj.</i> )	43.01%	something shared in general ( <i>adj.</i> )	4.32%
			in a fixed collocation “common market” ( <i>n.</i> )	10.31%		
			in a fixed collocation “common sense” ( <i>n.</i> )	11.37%	in a fixed collocation “common sense” ( <i>n.</i> )	95.10%
compared	in a fixed collocation “(when) compared with/to” ( <i>ph.</i> )	100%	in a fixed collocation “(when) compared with/to” ( <i>ph.</i> )	100%	in a fixed collocation “(when) compared with/to” ( <i>ph.</i> )	100%

comparison	to consider two things to find their differences ( <i>n.</i> )	100%	to consider two things to find their differences ( <i>n.</i> )	100%	to consider two things to find their differences ( <i>n.</i> )	100%
create	to cause something to happen ( <i>v.</i> )	100%	to cause something to happen ( <i>v.</i> )	100%	to cause something to happen ( <i>v.</i> )	100%
current	referring to “present” or “now” ( <i>adj.</i> )	100%	referring to “present” or “now” ( <i>adj.</i> )	75.51%	referring to “present” or “now” ( <i>adj.</i> )	100%
			electricity ( <i>tm.</i> ) [Engineering]	24.49%		
customer	people who would like to buy products ( <i>n.</i> )	100%	people who would like to buy products ( <i>n.</i> )	100%	people who would like to buy products ( <i>n.</i> )	100%
customers	people who would like to buy products ( <i>n.</i> )	100%	people who would like to buy products ( <i>n.</i> )	100%	people who would like to buy products ( <i>n.</i> )	100%
decision	the act of deciding something ( <i>n.</i> )	38.53%	the act of deciding something ( <i>n.</i> )	94.98%	the act of deciding something ( <i>n.</i> )	27.32%
	In a fixed collocation “decision variable” ( <i>tm.</i> )	30.21%				
	process of reaching decisions ( <i>n.</i> ); in “decision making”	21.33%	process of reaching decisions ( <i>n.</i> ); in “decision making”	5.02%	process of reaching decisions ( <i>n.</i> ); in “decision making”	72.68%
	in a fixed collocation “Markov decision chain/process” ( <i>tm.</i> )	5.30%				
	in a fixed collocation “decision tree” ( <i>tm.</i> )	4.62%				
demand	referring to how many people would buy products in the market ( <i>n.</i> )	63.53%			referring to how many people would buy products in the market ( <i>n.</i> )	100%
	referring to a firm request ( <i>v.</i> )	36.47%	referring to a firm request ( <i>v.</i> )	100%		
described	to say what people/things are like ( <i>v.</i> )	100%	to say what people/things are like ( <i>v.</i> )	100%	to say what people/things are like ( <i>v.</i> )	100%
design	to plan the way of doing something ( <i>n.</i> )	58.51%	to plan the way of doing something ( <i>n.</i> )	100%	to plan the way of doing something ( <i>n.</i> )	63.78%

	to plan a survey or research (v.)	40.38%				
	to make a detailed drawing from which things can be made (v.)	1.11%			to make a detailed drawing from which things can be made (v.)	16.40%
					the process of art design and drawing (n.)	19.82%
determine	to discover a fact as a result of investigation (v.)	93.89%	to discover a fact as a result of investigation (v.)	90.68%		
	to decide or settle something (v.)	6.11%	to decide or settle something (v.)	9.32%	to decide or settle something (v.)	71.78%
					a particular factor causes something to happen (v.)	28.22%
determining	to decide or settle something (v.)	100%	to decide or settle something (v.)	100%	to decide or settle something (v.)	100%
develop	to provide something with more details (v.)	59.20%	to provide something with more details (v.)	26.62%		
	to make something better (v.)	40.80%	to make something better (v.)	44.81%	to make something better (v.)	100%
			to have a specific idea (v.)	20.78%		
			to make something more profitable (v.)	6.49%		
			to design a new product and produce it (v.)	1.30%		
developed	to provide something with more details (v.)	85.75%			to provide something with more details (v.)	18.42%
	to describe a country/place with wealth (adj.)	14.25%	to describe a country/place with wealth (adj.)	100%	to describe a country/place with wealth (adj.)	62.32%
					to describe something that is more advanced (adj.)	19.26%
developing	to describe a country/place is not in wealth (adj.)	100%	to describe a country/place is not in wealth (adj.)	100%	to describe a country/place is not in wealth (adj.)	100%

development	the gradual growth or formation of something ( <i>n.</i> )	56.16%				
	to make the basic design better and advanced ( <i>n.</i> )	22.56%	to make the basic design better and advanced ( <i>n.</i> )	20.82%	to make the basic design better and advanced ( <i>n.</i> )	100%
	an event or incident that happened recently and is likely to have influence on present situation ( <i>n.</i> )	21.12%	an event or incident that happened recently and is likely to have influence on present situation ( <i>n.</i> )	36.91%		
	the process to make an area more useful and profitable ( <i>n.</i> )	0.16%	the process to make an area more useful and profitable ( <i>n.</i> )	42.27%		
direct	without changing directions or stopping ( <i>adj.</i> )	85.97%	without changing directions or stopping ( <i>adj.</i> )	21.46%		
	only involve necessary people, actions or activities ( <i>adj.</i> ); in collocation like “direct selling”	14.03%	only involve necessary people, actions or activities ( <i>adj.</i> )	77.45%	only involve necessary people, actions or activities ( <i>adj.</i> )	71.19%
			straightforward ( <i>adj.</i> )	1.09%		
					to lead or guide someone to somewhere ( <i>v.</i> )	28.81%
discussion	a piece of writing or lecture to talk something in detail ( <i>n.</i> )	100%	a piece of writing or lecture to talk something in detail ( <i>n.</i> )	100%	a piece of writing or lecture to talk something in detail ( <i>n.</i> )	91.90%
					a writing document containing something to discuss ( <i>n.</i> )	9.10%
discussed	to write or talk something in detail ( <i>v.</i> )	100%	to write or talk something in detail ( <i>v.</i> )	100%	to write or talk something in detail ( <i>v.</i> )	100%
discusses	to write or talk something in detail ( <i>v.</i> )	100%	to write or talk something in detail ( <i>v.</i> )	100%	to write or talk something in detail ( <i>v.</i> )	100%
economic	concerning with the organization of money,	100%	concerning with the organization of money,	100%	concerning with the organization of money,	100%

	industry, or trade of a country/region ( <i>adj.</i> )		industry, or trade of a country/region ( <i>adj.</i> )		industry, or trade of a country/region ( <i>adj.</i> )	
effect	changes of the first thing cause the second thing to happen ( <i>n.</i> )	82.67%	changes of the first thing cause the second thing to happen ( <i>n.</i> )	100%	changes of the first thing cause the second thing to happen ( <i>n.</i> )	77.83%
	in a fixed collocation “moderating effect” ( <i>tm.</i> )	7.86%				
	in a fixed collocation “mediating effect” ( <i>tm.</i> )	5.60%				
	in a fixed collocation “interaction effect” ( <i>tm.</i> )	3.57%				
	in a fixed collocation “ripple effect” ( <i>tm.</i> )	2.73%				
					greenhouse effect ( <i>n.</i> )	22.17%
effects	changes of the first thing cause the second thing to happen ( <i>n.</i> )	92.38%	changes of the first thing cause the second thing to happen ( <i>n.</i> )	66.67%	changes of the first thing cause the second thing to happen ( <i>n.</i> )	86.22%
	in a fixed collocation “moderating effects” ( <i>tm.</i> )	3.84%				
	in a fixed collocation “interaction effects” ( <i>tm.</i> )	3.78%				
					ill effects ( <i>n.</i> )	13.78%
			in a fixed collocation “side effect” ( <i>n.</i> )	33.33%		
environment	circumstance of something ( <i>n.</i> )	83.72%	circumstance of something ( <i>n.</i> )	100%		
	human-made conditions ( <i>n.</i> ); in a collocation “built environment”	16.28%				
					nature ( <i>n.</i> )	100%
example	to show a particular situation ( <i>n.</i> )	57.20%	to show a particular situation ( <i>n.</i> )	65.07%	to show a particular situation ( <i>n.</i> )	72.32%
	in a fixed collocation “for example” ( <i>ph.</i> )	42.80%	in a fixed collocation “for example” ( <i>ph.</i> )	34.93%	in a fixed collocation “for example” ( <i>ph.</i> )	27.68%

existing	to describe something is present ( <i>adj.</i> )	100%	to describe something is present ( <i>adj.</i> )	100%	to describe something is present ( <i>adj.</i> )	100%
future	the time period after the present time ( <i>adj.</i> )	100%	the time period after the present time ( <i>adj.</i> )	85.66%	the time period after the present time ( <i>adj.</i> )	35.35%
			as a term in language ( <i>tm.</i> ) [Language]	14.34%		
					the time period after the present time ( <i>n.</i> ) (the future)	64.65%
including	containing ( <i>prep.</i> )	100%	containing ( <i>prep.</i> )	100%	containing ( <i>prep.</i> )	100%
increase	become greater in number ( <i>n.</i> )	53.76%	become greater in number ( <i>n.</i> )	95.07%	become greater in number ( <i>n.</i> )	100%
	become greater in number ( <i>v.</i> )	46.24%	become greater in number ( <i>v.</i> )	4.93%		
increases	become greater in number ( <i>v.</i> )	100%	become greater in number ( <i>v.</i> )	48.84%	become greater in number ( <i>v.</i> )	3.92%
			become greater in number ( <i>n.</i> )	51.16%	become greater in number ( <i>n.</i> )	96.08%
increasing	become greater in number ( <i>v.</i> )	100%	become greater in number ( <i>v.</i> )	100%	become greater in number ( <i>v.</i> )	100%
increasingly	to describe a situation where the quality is becoming greater in intensity ( <i>adv.</i> )	100%	to describe a situation where the quality is becoming greater in intensity ( <i>adv.</i> )	100%	to describe a situation where the quality is becoming greater in intensity ( <i>adv.</i> )	100%
increased	become greater in number ( <i>v.</i> )	100%	become greater in number ( <i>v.</i> )	100%	become greater in number ( <i>v.</i> )	100%
individual	relating to one person or thing ( <i>adj.</i> )	100%	relating to one person or thing ( <i>adj.</i> )	69.43%	relating to one person or thing ( <i>adj.</i> )	73.29%
			referring to a human ( <i>n.</i> )	28.11%	referring to a human ( <i>n.</i> )	26.68%
			to describe someone or something as unusual ones ( <i>n.</i> )	2.46%	to describe someone or something as unusual ones ( <i>n.</i> )	0.03%
identified	discover or notice the existence of something ( <i>v.</i> )	100%	discover or notice the existence of something ( <i>v.</i> )	19.95%	discover or notice the existence of something ( <i>v.</i> )	100%

			to recognize the existence of something (v.)	80.05%		
identifies	discover or notice the existence of something (v.)	100%	discover or notice the existence of something (v.)	61.54%	discover or notice the existence of something (v.)	100%
			to recognize the existence of something (v.)	38.46%		
identify	discover or notice the existence of something (v.)	100%	discover or notice the existence of something (v.)	16.87%	discover or notice the existence of something (v.)	100%
			to recognize the existence of something (v.)	83.13%		
improve	to make something better (v.)	100%	to make something better (v.)	100%	to make something better (v.)	100%
include	it has the other thing as its part (v.)	100%	it has the other thing as its part (v.)	100%	it has the other thing as its part (v.)	100%
included	it has the other thing as its part (v.)	100%	it has the other thing as its part (v.)	100%	it has the other thing as its part (v.)	100%
includes	it has the other thing as its part (v.)	100%	it has the other thing as its part (v.)	100%	it has the other thing as its part (v.)	100%
indicate	the first thing shows that the second is true or exists (v.)	100%	the first thing shows that the second is true or exists (v.)	11.94%	the first thing shows that the second is true or exists (v.)	100%
			it is a sign of another thing (v.)	86.57%		
			to mention something indirectly (v.)	1.49%		
indicates	it is a sign of another thing (v.)	100%	it is a sign of another thing (v.)	88.08%	it is a sign of another thing (v.)	100%
			the first thing shows that the second is true or exists (v.)	6.95%		
			to mention something indirectly (v.)	4.97%		
industry	a particular trade (n.)	14.55 %	a particular trade (n.)	100%	a particular trade (n.)	13.70%

	the work and process to produce products in factories ( <i>n.</i> )	85.45%			the work and process to produce products in factories ( <i>n.</i> )	86.30%
influence	to affect what people do or what situation happens ( <i>n.</i> )	83.02%	to affect what people do or what situation happens ( <i>n.</i> )	38.55%	to affect what people do or what situation happens ( <i>n.</i> )	84.59%
	have an effect on something/someone ( <i>v.</i> )	16.98%	have an effect on something/someone ( <i>v.</i> )	61.45%	have an effect on something/someone ( <i>v.</i> )	15.41%
instance	in a fixed collocation “for instance” ( <i>ph.</i> )	100%	in a fixed collocation “for instance” ( <i>ph.</i> )	100%	in a fixed collocation “for instance” ( <i>ph.</i> )	100%
introduction	a piece of writing at the beginning to tell what the rest of writing is about ( <i>n.</i> )	99.27%	a piece of writing at the beginning to tell what the rest of writing is about ( <i>n.</i> )	100%	a piece of writing at the beginning to tell what the rest of writing is about ( <i>n.</i> )	30.72%
	a new product becomes available ( <i>n.</i> )	0.73%			a new product becomes available ( <i>n.</i> )	69.18%
investigate	try to find out what happened or what is the truth ( <i>v.</i> )	100%	try to find out what happened or what is the truth ( <i>v.</i> )	100%	try to find out what happened or what is the truth ( <i>v.</i> )	100%
investigated	try to find out what happened or what is the truth ( <i>v.</i> )	100%	try to find out what happened or what is the truth ( <i>v.</i> )	100%	try to find out what happened or what is the truth ( <i>v.</i> )	100%
knowledge	in a fixed collocation “to our best knowledge” ( <i>ph.</i> )	48.77%				
	the information or understanding of a subject or a topic ( <i>n.</i> )	51.23%	the information or understanding of a subject or a topic ( <i>n.</i> )	100%	the information or understanding of a subject or a topic ( <i>n.</i> )	100%
lack	to describe something is of insufficiency ( <i>n.</i> )	100%	to describe something is of insufficiency ( <i>n.</i> )	100%	to describe something is of insufficiency ( <i>n.</i> )	91.87%
					to describe something is of insufficiency ( <i>v.</i> )	9.13%
likely	probably to do something ( <i>adj.</i> )	100%	probably to do something ( <i>adj.</i> )	100%	probably to do something ( <i>adj.</i> )	100%
limitations	only can do something, not others ( <i>n.</i> )	100%	only can do something, not others ( <i>n.</i> )	20.51%	only can do something, not others ( <i>n.</i> )	99.95%

			a fact or situation that only makes some actions impossible ( <i>n.</i> )	79.49%		
					the process or action to controlling something or reducing something else ( <i>n.</i> )	0.05%
limited	not very great in number/size/range ( <i>adj.</i> )	100%			not very great in number/size/range ( <i>adj.</i> )	75.99%
			to describe limited companies ( <i>adj.</i> )	100%	to describe limited companies ( <i>adj.</i> )	24.01%
lower	less quality or intense ( <i>adj.</i> )	56.76%			less quality or intense ( <i>adj.</i> )	100%
	to refer to bottom of something ( <i>adj.</i> )	43.24%	to refer to bottom of something ( <i>adj.</i> )	6.66%		
			smaller amount ( <i>adj.</i> )	68.49%		
			division of society ( <i>adj.</i> )	23.37%		
			to reduce ( <i>v.</i> )	1.48%		
measure	to discover or judge how great the amount is ( <i>v.</i> )	86.76%	to discover or judge how great the amount is ( <i>v.</i> )	87.77%	to discover or judge how great the amount is ( <i>v.</i> )	20.00%
	as an official/standard measurement ( <i>n.</i> )	13.26%	as an official amount ( <i>n.</i> )	11.70%	as an official amount ( <i>n.</i> )	60.98%
			a large amount of ( <i>n.</i> )	0.53%		
					tape measure ( <i>n.</i> )	19.02%
measures	as an official/standard measurement ( <i>n.</i> )	97.72%	as an official amount ( <i>n.</i> )	100%	as an official amount ( <i>n.</i> )	100%
	to discover or judge how great the amount is ( <i>v.</i> )	2.28%				
model	a system that is being used ( <i>n.</i> )	98.81%	a system that is being used ( <i>n.</i> )	39.11%	a system that is being used ( <i>n.</i> )	100%
	to make an accurate theoretical descriptions to understand the framework ( <i>v.</i> )	1.19%				

			in a fixed collocation “Bohr’s model” ( <i>tm.</i> ) [Physics]	60.89%		
models	a system that is being used ( <i>n.</i> )	100%	a system that is being used ( <i>n.</i> )	100%	a system that is being used ( <i>n.</i> )	68.28%
					role model ( <i>n.</i> )	31.72%
observed	to watch something carefully ( <i>v.</i> )	100%	to watch something carefully ( <i>v.</i> )	44.00%	to watch something carefully ( <i>v.</i> )	92.86%
			to notice something ( <i>v.</i> )	56.00%	to notice something ( <i>v.</i> )	7.14%
operations	to run a business or company ( <i>n.</i> )	100%	to run a business or company ( <i>n.</i> )	100%	to run a business or company ( <i>n.</i> )	100%
per	to express rates and ratios ( <i>prep.</i> )	79.43%	to express rates and ratios ( <i>prep.</i> )	52.74%	to express rates and ratios ( <i>prep.</i> )	83.59%
	in a fixed collocation “per cent” ( <i>n.</i> )	16.37%	in a fixed collocation “per cent” ( <i>n.</i> )	25.73%	in a fixed collocation “per cent” ( <i>n.</i> )	16.41%
	in a fixed collocation “per capita” ( <i>adj.</i> )	3.64%	in a fixed collocation “per capita” ( <i>adj.</i> )	17.42%		
			in a fixed collocation “per se” ( <i>adv.</i> )	4.11%		
period	a specific time ( <i>n.</i> )	67.19%	a specific time ( <i>n.</i> )	6.19%	a specific time ( <i>n.</i> )	100%
	referring to a time period ( <i>n.</i> )	33.81%	referring to a time period ( <i>n.</i> )	93.81%		
positive	to describe the situation is pleasant ( <i>adj.</i> )	52.39%	to describe the situation is pleasant ( <i>adj.</i> )	74.58%	to describe the situation is pleasant ( <i>adj.</i> )	29.57%
	to show something is evident ( <i>adj.</i> )	27.37%	to show something is evident ( <i>adj.</i> )	17.80%	to show something is evident ( <i>adj.</i> )	8.12%
	to indicate something/measure is helpful ( <i>adj.</i> )	18.42%			to indicate something/measure is helpful ( <i>adj.</i> )	12.75%
	to describe a number larger than zero ( <i>n.</i> )	1.83%	to describe a number larger than zero ( <i>n.</i> )	0.42%	to describe a number larger than zero ( <i>n.</i> )	10.43%
			to describe electrons ( <i>adj.</i> )	7.20%	to describe electrons ( <i>adj.</i> )	7.54%
					to be used in medical test ( <i>adj.</i> )	31.59%
practical	based on real situation ( <i>adj.</i> )	100%	based on real situation ( <i>adj.</i> )	100%	based on real situation ( <i>adj.</i> )	100%

practice	to do something regularly ( <i>n.</i> )	100%	to do something regularly ( <i>n.</i> )	100%	to do something regularly ( <i>n.</i> )	27.71%
					in a fixed collocation “in practice” ( <i>ph.</i> )	47.09%
					code of practice ( <i>n.</i> )	18.58%
					used in law-related or medical contexts ( <i>n.</i> )	6.62%
previous	referring to “past” ( <i>adj.</i> )	100%	referring to “past” ( <i>adj.</i> )	100%	referring to “past” ( <i>adj.</i> )	100%
process	a series of actions ( <i>n.</i> )	100%	a series of actions ( <i>n.</i> )	100%	a series of actions ( <i>n.</i> )	83.05%
					peace process ( <i>n.</i> )	16.95%
processes	a series of action ( <i>n.</i> )	100%	a series of actions ( <i>n.</i> )	100%	a series of actions ( <i>n.</i> )	72.18%
					about climate change ( <i>n.</i> )	27.72%
product	the goods for selling ( <i>n.</i> )	100%	the goods for selling ( <i>n.</i> )	100%		
					in a fixed collocation “gross domestic/national product” ( <i>n.</i> )	100%
products	the goods for selling ( <i>n.</i> )	100%	the goods for selling ( <i>n.</i> )	100%	the goods for selling ( <i>n.</i> )	100%
production	the action or process of making products ( <i>n.</i> )	100%	the action or process of making products ( <i>n.</i> )	100%	the action or process of making products ( <i>n.</i> )	100%
propose	suggest a plan for others to think about it ( <i>v.</i> )	81.90%	suggest a plan for others to think about it ( <i>v.</i> )	61.22%	suggest a plan for others to think about it ( <i>v.</i> )	71.43%
	state that something like a theory could be true ( <i>v.</i> )	18.10%	state that something like a theory could be true ( <i>v.</i> )	28.57%		
			intend to do something ( <i>v.</i> )	10.21%		
					ask someone to drink ( <i>v.</i> )	28.57%
proposed	state that something like a theory could be true ( <i>v.</i> )	87.90%	state that something like a theory could be true ( <i>v.</i> )	100%		
	suggest a plan for others to think about it ( <i>v.</i> )	12.10%			suggest a plan for others to think about it ( <i>v.</i> )	100%

provide	to make something available (v.)	100%	to make something available (v.)	100%	to make something available (v.)	100%
provided	to make something available (v.)	100%	to make something available (v.)	100%	to make something available (v.)	100%
provides	to make something available (v.)	100%	to make something available (v.)	100%	to make something available (v.)	100%
providing	to make something available (v.)	100%	to make something available (v.)	100%	to make something available (v.)	100%
purpose	the aim of doing something (n.)	100%	the aim of doing something (n.)	100%	the aim of doing something (n.)	100%
quality	to describe how good or bad something is (n.)	60.84%	to describe how good or bad something is (n.)	60.12%	to describe how good or bad something is (n.)	100%
	referring to a standard (n.)	39.16%	referring to a standard (n.)	39.88%		
range	a number of different things in the general kind (n.)	94.05%	a number of different things in the general kind (n.)	100%	a number of different things in the general kind (n.)	100%
	in a fixed collocation "middle range theory" (tm.)	5.95%				
reduce	to make something smaller in size or amount (v.)	100%	to make something smaller in size or amount (v.)	100%	to make something smaller in size or amount (v.)	100%
reduced	to make something smaller in size or amount (v.)	100%	to make something smaller in size or amount (v.)	100%	to make something smaller in size or amount (v.)	100%
reducing	to make something smaller in size or amount (v.)	100%	to make something smaller in size or amount (v.)	100%	to make something smaller in size or amount (v.)	100%
regarding	to indicate that the subject is being talked about (prep.)	100%	to indicate that the subject is being talked about (prep.)	100%	to indicate that the subject is being talked about (prep.)	100%
represent	to show the meaning of something like a symbol (v.)	100%	to show the meaning of something like a symbol (v.)	100%	to show the meaning of something like a symbol (v.)	16.09%
					on behalf of someone or something (v.)	83.91%

represents	to show the meaning of something like a symbol (v.)	100%	to show the meaning of something like a symbol (v.)	100%	to show the meaning of something like a symbol (v.)	100%
required	something is necessary or needed (v.)	100%	something is necessary or needed (v.)	100%	something is necessary or needed (v.)	86.55%
					used in context related to laws, firmly requiring (v.)	13.45%
requires	referring to “have to do something” (v.)	100%				
			something is necessary or needed (v.)	100%	something is necessary or needed (v.)	100%
research	a piece of scientific study (n.)	100%	a piece of scientific study (n.)	100%	a piece of scientific study (n.)	100%
researchers	people who conduct research (n.)	100%	people who conduct research (n.)	100%	people who conduct research (n.)	100%
respect	in a fixed collocation “with respect to” (ph.)	100%			in a fixed collocation “with respect to” (ph.)	31.44%
			to have a good opinion on something (n.)	70.83%	to have a good opinion on something (n.)	46.51%
			to have a good opinion on something (v.)	29.17%	to have a good opinion on something (v.)	6.11%
					to respect one’s wishes to avoid doing something that the people dislike (v.)	15.94%
result	something happens because other things happened (n.); in a collocation “as a result (of)”	55.77%	something happens because other things happened (n.); in a collocation “as a result (of)”	100%	something happens because other things happened (n.); in a collocation “as a result (of)”	100%
	as a final end (n.)	30.54%				
	cause a situation (v.); result in	12.85%				
	caused by a situation (v.); result from	0.84%				
resulting	cause a situation (v.)	100%	cause a situation (v.)	10.67%	cause a situation (v.)	58.87%

			caused by a situation (v.); result from	89.33%	caused by a situation (v.); result from	41.13%
results	something happens because other things happened (n.)	99.15%	something happens because other things happened (n.)	100%	something happens because other things happened (n.)	100%
	cause a situation (v.); result in	0.73%				
	caused by a situation (v.); result from	0.12%				
risk	in a fixed collocation “risk management” (tm.)	51.61%				
	may have bad results (n.)	27.65%	may have bad results (n.)	61.38%	may have bad results (n.)	24.45%
	in a fixed collocation “risk mitigation” (tm.)	11.51%				
	something may have bad results (n.)	9.59%	something may have bad results (n.)	38.62%	something may have bad results (n.)	75.55%
role	to have a particular position (n.)	100%	to have a particular position (n.)	99.81%	to have a particular position (n.)	98.68%
			in a fixed collocation “role play” (n.)	0.19%		
					as a character (n.)	1.32%
scale	size or extent (n.)	47.26%	size or extent (n.)	57.25%		
	a set of levels or numbers to measure (n.)	52.74%	a set of levels or numbers to measure (n.)	42.75%	a set of levels or numbers to measure (n.)	100%
section	a part of writing (n.)	100%	a part of writing (n.)	28.52%	a part of writing (n.)	100%
			a part of law (tm.) [Law]	31.64%		
			a part of something (n.)	39.84%		
selected	to choose something (v.)	100%	to choose something (v.)	100%	to choose something (v.)	91.90%
					something that has many desirable features but is only available for some people (adj.)	8.1%
similar	two things are almost the same in feature (adj.)	100%	two things are almost the same in feature (adj.)	100%	two things are almost the same in feature (adj.)	97.35%

					in a fixed collocation “in a similar vein” ( <i>ph.</i> )	2.65%
similarly	to indicate that two things are close in relation ( <i>adv.</i> )	100%	to indicate that two things are close in relation ( <i>adv.</i> )	100%	to indicate that two things are close in relation ( <i>adv.</i> )	100%
specific	referring to a particular problem or situation ( <i>adj.</i> )	100%	referring to a particular problem or situation ( <i>adj.</i> )	100%	referring to a particular problem or situation ( <i>adj.</i> )	90.37%
					to provide a precise description ( <i>adj.</i> )	9.63%
specifically	to add something more precise ( <i>adv.</i> )	100%	to add something more precise ( <i>adv.</i> )	100%	to add something more precise ( <i>adv.</i> )	100%
standard	something to judge the quality or else ( <i>n.</i> )	89.64%	something to judge the quality or else ( <i>n.</i> )	43.27%	something to judge the quality or else ( <i>n.</i> )	11.62%
	the level or quality to achieve ( <i>n.</i> )	10.36%	the level or quality to achieve ( <i>n.</i> )	23.48%	the level or quality to achieve ( <i>n.</i> )	56.59%
			in a fixed collocation “Standard English” ( <i>tm.</i> ) (Language)	33.25%	in a fixed collocation “Standard English” ( <i>adj.</i> ) (Language)	15.74%
					standard gauge ( <i>n.</i> )	6.05%
					as a newspaper Evening Standard ( <i>n.</i> )	10.00%
supply	in a fixed collocation “supply chain (management)” ( <i>tm.</i> )	85.46%	in a fixed collocation “supply chain (management)” ( <i>tm.</i> )	23.94%		
	to provide something ( <i>v.</i> )	14.55%			to provide something ( <i>v.</i> )	23.76%
			referring to “goods” ( <i>n.</i> )	29.93%		
			referring to something available ( <i>n.</i> )	39.08%	referring to something available ( <i>n.</i> )	76.24%
			in a fixed collocation “aggregate supply” ( <i>tm.</i> ) [Economics]	7.05%		
traditional	existing for a long time ( <i>adj.</i> )	100%	existing for a long time ( <i>adj.</i> )	100%	existing for a long time ( <i>adj.</i> )	100%
value	how much money it worth ( <i>n.</i> )	48.69%			how much money it worth ( <i>n.</i> )	100%
	a particular number ( <i>n.</i> )	30.23%	a particular number ( <i>n.</i> )	100%		

	something is of usefulness or importance ( <i>n.</i> )	12.83%				
	in a fixed collocation “value added” ( <i>tm.</i> )	8.24%				
values	a particular number ( <i>n.</i> )	100%	a particular number ( <i>n.</i> )	100%		
					referring to moral ( <i>n.</i> )	83.09%
					referring to importance ( <i>n.</i> )	16.91%
variable	a value used in technical contexts ( <i>n.</i> )	83.89%	a value used in technical contexts ( <i>n.</i> )	22.31%		
	factor ( <i>n.</i> )	12.68%	factor ( <i>n.</i> )	77.69%	factor ( <i>n.</i> )	100%
	changeable ( <i>adj.</i> )	2.48%				
variables	a value used in technical contexts ( <i>n.</i> )	100%	a value used in technical contexts ( <i>n.</i> )	14.84%		
			factor ( <i>n.</i> )	85.16%	factor ( <i>n.</i> )	100%
various	there are several different things of the type mentioned ( <i>adj.</i> )	100%	there are several different things of the type mentioned ( <i>adj.</i> )	100%	there are several different things of the type mentioned ( <i>adj.</i> )	100%

Level 3	LAWEC		BAWE		BNC1994	
Item	Sense/Use	Occurrence (Percentage)	Sense/Use	Occurrence (Percentage)	Sense/Use	Occurrence (Percentage)
abstract	as a section of research papers or writings ( <i>n.</i> )	100%			as a section of research papers or writings ( <i>n.</i> )	7.59%
			in a fixed collocation “abstract labour” ( <i>tm.</i> ) (sociology)	61.40%		
			something that is not practical ( <i>adj.</i> )	38.60%		
					about arts ( <i>adj.</i> )	92.41%
achieve	to success in doing something ( <i>v.</i> )	100%	to success in doing something ( <i>v.</i> )	100%	to success in doing something ( <i>v.</i> )	100%
adopted	begin to have a plan ( <i>v.</i> )	100%	begin to have a plan ( <i>v.</i> )	93.68%	begin to have a plan ( <i>v.</i> )	100%
			to get children from others ( <i>v.</i> )	6.32%		
analyse	to consider it carefully or use statistical method to fully understand it ( <i>v.</i> )	100%	to consider it carefully or use statistical method to fully understand it ( <i>v.</i> )	100%	to consider it carefully or use statistical method to fully understand it ( <i>v.</i> )	100%
analysis	a specific method of processing data ( <i>n.</i> )	65.53%	a specific method of processing data ( <i>n.</i> )	23.98%		
	processing of working with data ( <i>n.</i> )	34.47%	processing of working with data ( <i>n.</i> )	76.02%	processing of working with data ( <i>n.</i> )	100%
appropriate	being suitable or acceptable for a particular situation ( <i>adj.</i> )	100%	being suitable or acceptable for a particular situation ( <i>adj.</i> )	100%	being suitable or acceptable for a particular situation ( <i>adj.</i> )	100%
aspects	a part or nature of something ( <i>n.</i> )	100%	a part or nature of something ( <i>n.</i> )	100%	a part or nature of something ( <i>n.</i> )	100%
authors	writers ( <i>n.</i> )	100%	writers ( <i>n.</i> )	100%	writers ( <i>n.</i> )	100%
capacity	the volume the container to hold ( <i>n.</i> )	53.67%	the volume the container to hold ( <i>n.</i> )	53.62%	the volume the container to hold ( <i>n.</i> )	44.48%
	ability of factories to produce ( <i>n.</i> )	42.53%	ability of factories to produce ( <i>n.</i> )	2.90%	ability of factories to produce ( <i>n.</i> )	38.80%

	ability of doing something ( <i>n.</i> )	3.80%			ability of doing something ( <i>n.</i> )	16.72%
			in a fixed collocation “antioxidant capacity” ( <i>tm.</i> ) [Food Science]	43.48%		
characteristics	quality or features of something ( <i>n.</i> )	100%	quality or features of something ( <i>n.</i> )	100%	quality or features of something ( <i>n.</i> )	100%
competitive	firmly compete with each other ( <i>adj.</i> )	100%	firmly compete with each other ( <i>adj.</i> )	19.18%	firmly compete with each other ( <i>adj.</i> )	100%
			to describe successful person ( <i>adj.</i> )	80.82%		
complex	something that has different parts and therefore is difficult to understand ( <i>adj.</i> )	100%	something that has different parts and therefore is difficult to understand ( <i>adj.</i> )	98.44%	something that has different parts and therefore is difficult to understand ( <i>adj.</i> )	84.33%
			a group of objects or buildings ( <i>n.</i> )	1.56%	a group of objects or buildings ( <i>n.</i> )	0.44%
					leisure complex ( <i>n.</i> )	13.02%
					referring to a system ( <i>n.</i> )	2.21%
comprehensive	to include everything relevant ( <i>adj.</i> )	100%	to include everything relevant ( <i>adj.</i> )	100%	to include everything relevant ( <i>adj.</i> )	56.97%
					comprehensive school ( <i>n.</i> )	43.03%
conclusion	the ending of something ( <i>n.</i> )	100%	the ending of something ( <i>n.</i> )	100%		
					final thoughts ( <i>n.</i> )	100%
conclusions	final thoughts ( <i>n.</i> )	86.23%	final thoughts ( <i>n.</i> )	100%	final thoughts ( <i>n.</i> )	100%
	the ending of something ( <i>n.</i> )	13.77%				
conducted	to organize something and carry it out ( <i>v.</i> )	100%	to organize something and carry it out ( <i>v.</i> )	100%	to organize something and carry it out ( <i>v.</i> )	100%
consistent	things are not contradict ( <i>adj.</i> )	100%	things are not contradict ( <i>adj.</i> )	13.33%	things are not contradict ( <i>adj.</i> )	100%
			in a fixed collocation “consistent mass	86.67%		

			matrix” [Engineering] ( <i>tm.</i> )			
context	general conditions ( <i>n.</i> )	100%	general conditions ( <i>n.</i> )	100%	general conditions ( <i>n.</i> )	100%
corresponding	a close relation with something else ( <i>adj.</i> )	100%	a close relation with something else ( <i>adj.</i> )	100%	a close relation with something else ( <i>adj.</i> )	100%
data	information or facts or statistics ( <i>n.</i> )	84.91%	information or facts or statistics ( <i>n.</i> )	89.04%	information or facts or statistics ( <i>n.</i> )	100%
	information stored in computers or electronic devices ( <i>n.</i> )	15.09%	information stored in computers or electronic devices ( <i>n.</i> )	10.96%		
decreases	become smaller in amount ( <i>v.</i> )	99.53%	become smaller in amount ( <i>v.</i> )	100%	become smaller in amount ( <i>v.</i> )	64.44%
	become smaller in amount ( <i>n.</i> )	0.47%			become smaller in amount ( <i>n.</i> )	35.56%
defined	to show, describe and state clearly what it is ( <i>v.</i> ); in collocations “defined as”, “defined as follows”, “as defined”	100%	to show, describe and state clearly what it is ( <i>v.</i> )	100%		
					its outline is clear and strong ( <i>adj.</i> )	100%
distribution	giving or delivering thing to many people and things ( <i>n.</i> )	57.91%			giving or delivering thing to many people and things ( <i>n.</i> )	49.70%
	how much of something that each time, place, or people has ( <i>n.</i> )	42.09%	how much of something that each time, place, or people has ( <i>n.</i> )	100%	how much of something that each time, place, or people has ( <i>n.</i> )	50.30%
efficiency	ability to do tasks ( <i>n.</i> )	88.21%	ability to do tasks ( <i>n.</i> )	43.06%	ability to do tasks ( <i>n.</i> )	41.44%
	engine ratio ( <i>n.</i> )	11.79%	engine ratio ( <i>n.</i> )	40.28%	engine ratio ( <i>n.</i> )	58.56%
			in a fixed collocation “thermal efficiency” ( <i>tm.</i> ) [Engineering]	16.67%		
ensure	to make it happen ( <i>v.</i> )	100%	to make it happen ( <i>v.</i> )	100%	to make it happen ( <i>v.</i> )	100%
evaluate	to assess something ( <i>v.</i> )	100%	to assess something ( <i>v.</i> )	100%	to assess something ( <i>v.</i> )	100%

factor	influential aspects ( <i>n.</i> )	100%	influential aspects ( <i>n.</i> )	74.44%	influential aspects ( <i>n.</i> )	100%
			in a fixed collocation “friction factor” ( <i>tm.</i> ) [Engineering]	16.39%		
			in a fixed collocation “transcription factor” ( <i>tm.</i> ) [Biology]	9.17%		
factors	influential aspects ( <i>n.</i> )	100%	influential aspects ( <i>n.</i> )	100%	influential aspects ( <i>n.</i> )	100%
focus	to concentrate on something ( <i>v.</i> )	58.50%	to concentrate on something ( <i>v.</i> )	83.17%	to concentrate on something ( <i>v.</i> )	66.15%
	special attention to pay ( <i>n.</i> )	41.50%	special attention to pay ( <i>n.</i> )	16.83%	special attention to pay ( <i>n.</i> )	33.85%
focused	to concentrate on something ( <i>v.</i> )	100%	to concentrate on something ( <i>v.</i> )	100%	to concentrate on something ( <i>v.</i> )	100%
focuses	to concentrate on something ( <i>v.</i> )	100%	to concentrate on something ( <i>v.</i> )	100%	to concentrate on something ( <i>v.</i> )	100%
framework	a particular set of rules to deal with problems ( <i>n.</i> )	100%	a particular set of rules to deal with problems ( <i>n.</i> )	100%	a particular set of rules to deal with problems ( <i>n.</i> )	97.58%
					in a fixed collocation “framework knitters” ( <i>n.</i> ) in history	2.42%
frameworks	a particular set of rules to deal with problems ( <i>n.</i> )	100%	a particular set of rules to deal with problems ( <i>n.</i> )	100%	a particular set of rules to deal with problems ( <i>n.</i> )	100%
function	formula ( <i>n.</i> )	99.32%			formula ( <i>n.</i> )	26.64%
	as a particular use of something ( <i>n.</i> )	0.68%	as a particular use of something ( <i>n.</i> )	100%	as a particular use of something ( <i>n.</i> )	73.36%
functions	formula ( <i>n.</i> )	100%	formula ( <i>n.</i> )	17.59%		
			as a particular use of something ( <i>n.</i> )	82.41%	as a particular use of something ( <i>n.</i> )	100%
furthermore	to introduce more information to support previous mentioned one ( <i>adv.</i> )	100%	to introduce more information to support previous mentioned one ( <i>adv.</i> )	100%	to introduce more information to support previous mentioned one ( <i>adv.</i> )	100%
global	referring to “world” ( <i>adj.</i> )	100%	referring to “world” ( <i>adj.</i> )	100%	referring to “world” ( <i>adj.</i> )	100%

impact	a sudden or powerful effect ( <i>n.</i> )	99.92%	a sudden or powerful effect ( <i>n.</i> )	100%	a sudden or powerful effect ( <i>n.</i> )	100%
	to affect ( <i>v.</i> )	0.08%				
implications	what is implied is the case ( <i>n.</i> )	100%	what is implied is the case ( <i>n.</i> )	100%	what is implied is the case ( <i>n.</i> )	100%
importance	something of significance ( <i>n.</i> )	100%	something of significance ( <i>n.</i> )	100%	something of significance ( <i>n.</i> )	100%
independent	things are separated and not connected ( <i>adj.</i> )	100%	things are separated and not connected ( <i>adj.</i> )	78.79%	things are separated and not connected ( <i>adj.</i> )	100%
			an organization has its own finance ( <i>n.</i> )	21.21%		
initial	at the beginning of stages ( <i>adj.</i> )	100%	at the beginning of stages ( <i>adj.</i> )	100%	at the beginning of stages ( <i>adj.</i> )	72.64%
					in a fixed collocation "initial margin" ( <i>n.</i> )	7.21%
					in a fixed collocation "initial outlay" ( <i>n.</i> )	6.72%
					in a fixed collocation "initial public offering" ( <i>adj.</i> )	13.43%
insights	an accurate and deep understanding of something ( <i>n.</i> )	100%	an accurate and deep understanding of something ( <i>n.</i> )	100%	an accurate and deep understanding of something ( <i>n.</i> )	100%
literature	research papers or study ( <i>n.</i> )	100%	research papers or study ( <i>n.</i> )	85.51%	research papers or study ( <i>n.</i> )	0.52%
			novels, plays and poetry ( <i>n.</i> )	14.49%	novels, plays and poetry ( <i>n.</i> )	99.48%
method	a particular way of doing something ( <i>n.</i> )	100%	a particular way of doing something ( <i>n.</i> )	100%	a particular way of doing something ( <i>n.</i> )	100%
methodology	a system of methods ( <i>n.</i> )	100%	a system of methods ( <i>n.</i> )	100%	a system of methods ( <i>n.</i> )	100%
methods	a particular way of doing something ( <i>n.</i> )	100%	a particular way of doing something ( <i>n.</i> )	100%	a particular way of doing something ( <i>n.</i> )	100%
moreover	to add more information to support previous one ( <i>adv.</i> )	100%	to add more information to support previous one ( <i>adv.</i> )	100%	to add more information to support previous one ( <i>adv.</i> )	100%

negative	something unpleasant ( <i>adj.</i> )	87.49%	something unpleasant ( <i>adj.</i> )	96.70%	something unpleasant ( <i>adj.</i> )	32.60%
	to mean “no” ( <i>adj.</i> )	9.78%				
	number smaller than zero ( <i>n.</i> )	1.87%	number smaller than zero ( <i>n.</i> )	0.94%	number smaller than zero ( <i>n.</i> )	11.29%
	not evident in experiment ( <i>adj.</i> )	0.86%	not evident in experiment ( <i>adj.</i> )	0.47%	not evident in experiment ( <i>adj.</i> )	8.78%
			to describe electricity ( <i>adj.</i> )	1.89%	to describe electricity ( <i>adj.</i> )	8.15%
					used in medical contexts ( <i>adj.</i> )	0.94%
network	a large number of something having connections ( <i>n.</i> )	100%	a large number of something having connections ( <i>n.</i> )	100%	a large number of something having connections ( <i>n.</i> )	68.86%
					a large number of lines or roads crossing to meet at some points ( <i>n.</i> )	31.14%
objective	something to achieve ( <i>n.</i> )	100%			something to achieve ( <i>n.</i> )	86.26%
			based on facts and not subjective ( <i>adj.</i> )	100%	based on facts and not subjective ( <i>adj.</i> )	13.74%
obtain	to acquire something ( <i>v.</i> )	100%	to acquire something ( <i>v.</i> )	100%	to acquire something ( <i>v.</i> )	100%
obtained	to acquire something ( <i>v.</i> )	100%	to acquire something ( <i>v.</i> )	100%	to acquire something ( <i>v.</i> )	100%
overall	to talk about something in general ( <i>adj.</i> )	100%	to talk about something in general ( <i>adj.</i> )	100%	to talk about something in general ( <i>adj.</i> )	100%
perspective	a particular way of thinking ( <i>n.</i> )	100%	a particular way of thinking ( <i>n.</i> )	100%	a particular way of thinking ( <i>n.</i> )	56.06%
					In a fixed collocation “perspective cue” ( <i>n.</i> )	43.94%
potential	something or someone is capable of developing into the particular things or people ( <i>adj.</i> )	65.40%	something or someone is capable of developing into the particular things or people ( <i>adj.</i> )	100%	something or someone is capable of developing into the particular things or people ( <i>adj.</i> )	82.59%

	to have the necessary abilities ( <i>n.</i> )	34.60%				
					may be possible to do something ( <i>n.</i> ) (have potential for)	17.41%
relatively	to a certain degree ( <i>adv.</i> )	100%	to a certain degree ( <i>adv.</i> )	100%	to a certain degree ( <i>adv.</i> )	100%
relevant	something is of significance or importance in a particular situation ( <i>adj.</i> )	100%	something is of significance or importance in a particular situation ( <i>adj.</i> )	100%	something is of significance or importance in a particular situation ( <i>adj.</i> )	100%
respectively	in the same order of discussing things as mentioned ( <i>adv.</i> )	100%	in the same order of discussing things as mentioned ( <i>adv.</i> )	100%	in the same order of discussing things as mentioned ( <i>adv.</i> )	100%
resources	materials to use ( <i>n.</i> )	69.77%	materials to use ( <i>n.</i> )	70.81%	materials to use ( <i>n.</i> )	63.91%
	in a fixed collocation “human resources” ( <i>tm.</i> )	30.23%	in a fixed collocation “human resources” ( <i>tm.</i> ) [Management]	28.11%		
			referring to natural resources ( <i>n.</i> )	1.08%		
					things to increase wealth ( <i>n.</i> )	36.09%
review	formal explanation made by people in authority ( <i>n.</i> )	97.34%	formal explanation made by people in authority ( <i>n.</i> )	37.50%		
	to consider it carefully ( <i>v.</i> )	2.66%				
			in a fixed collocation “judicial review” ( <i>tm.</i> ) [Law]	62.50%	in a fixed collocation “judicial review” ( <i>n.</i> ) [Law]	60.18%
					a summary of books or articles ( <i>n.</i> )	39.82%
sample	a small number of people chosen from a large group for testing to provide	100%	a small number of people chosen from a large group for testing to provide	100%	a small number of people chosen from a large group for testing to provide	100%

	information about this group ( <i>n.</i> )		information about this group ( <i>n.</i> )		information about this group ( <i>n.</i> )	
significant	to have large effects on something ( <i>adj.</i> )	83.76%			to have large effects on something ( <i>adj.</i> )	84.88%
	being of importance ( <i>adj.</i> )	16.24%	being of importance ( <i>adj.</i> )	37.24%	being of importance ( <i>adj.</i> )	15.12%
			being able to show something clearly ( <i>adj.</i> )	62.76%		
solution	a way of dealing with difficult problems ( <i>n.</i> )	100%	a way of dealing with difficult problems ( <i>n.</i> )	51.93%	a way of dealing with difficult problems ( <i>n.</i> )	100%
			a kind of liquid ( <i>n.</i> )	48.07%		
solutions	a way of dealing with difficult problems ( <i>n.</i> )	100%	a way of dealing with difficult problems ( <i>n.</i> )	50.00%		
			a kind of liquid ( <i>n.</i> )	50.00%		
					used in names of companies ( <i>n.</i> )	100%
strategy	a general plan to achieve something ( <i>n.</i> )	100%	a general plan to achieve something ( <i>n.</i> )	100%	a general plan to achieve something ( <i>n.</i> )	100%
strategies	a general plan to achieve something ( <i>n.</i> )	100%	a general plan to achieve something ( <i>n.</i> )	100%	a general plan to achieve something ( <i>n.</i> )	100%
structure	something that consists of parts connected with each other ( <i>n.</i> )	100%	something that consists of parts connected with each other ( <i>n.</i> )	100%		
					the way of how something is organized or made ( <i>n.</i> )	100%
theoretical	based on abstract principles ( <i>adj.</i> )	100%	based on abstract principles ( <i>adj.</i> )	100%	based on abstract principles ( <i>adj.</i> )	100%
theory	formal ideas to explain something ( <i>n.</i> )	61.27%	formal ideas to explain something ( <i>n.</i> )	100%	formal ideas to explain something ( <i>n.</i> )	100%
	the set of rules or principles that form	30.73%				

	the basis of something ( <i>n.</i> )					
transportation	system of moving ( <i>n.</i> )	84.88%	system of moving ( <i>n.</i> )	96.00%	system of moving ( <i>n.</i> )	100%
	vehicle ( <i>n.</i> )	15.12%	vehicle ( <i>n.</i> )	4.00%		

Level 4	LAWEC		BAWE		BNC1994	
Item	Sense/Use	Occurrence (Percentage)	Sense/Use	Occurrence (Percentage)	Sense/Use	Occurrence (Percentage)
consolidate	to strengthen something (v.)	100%	to strengthen something (v.)	100%	to strengthen something (v.)	100%
empirical	relying on practical experience (adj.)	100%	relying on practical experience (adj.)	100%	relying on practical experience (adj.)	100%
empirically	relying on practical experience (adv.)	100%	relying on practical experience (adv.)	100%	relying on practical experience (adv.)	100%
inventory	supply or stock (n.)	93.44%	supply or stock (n.)	63.64%	supply or stock (n.)	33.33%
	in a fixed collocation “holding inventory” (tm.)	6.56%	in a fixed collocation “holding inventory” (tm.)	22.08%		
			a written report (n.)	9.09%	a written report (n.)	66.67%
			in a fixed collocation “phonetic inventory” (tm.) [Language]	5.19%		
logistics	skilful organization of something for efficiency (n.)	100%	skilful organization of something for efficiency (n.)	100%	skilful organization of something for efficiency (n.)	100%
optimal	the best level or state of something (adj.)	100%	the best level or state of something (adj.)	100%	the best level or state of something (adj.)	100%
remainder	rest of writing (n.)	100%	rest of writing (n.)	100%		
					something left (n.)	97.83%
					number (n.)	2.17%