

**TOWARD A TOOL FOR EVALUATING CORPUS-BASED WORD LIST FOR USE IN ENGLISH
LANGUAGE TEACHING CONTEXTS**

by

SARAH NASSER M ALZEER

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

Department of English Language and Linguistics

School of English, Drama and Creative Studies

College of Arts and Law

University of Birmingham

June 2021

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

ABSTRACT

Corpus linguistics has made possible the construction of lists of words that merit instructional focus based on the frequency of their use (Folse & Youngblood, 2017). As large corpora and more sophisticated corpus-analysis tools have become available, the number of corpus-based word lists targeting different types of vocabulary has rapidly increased during the last 20 years. This wide variety of lists has caused problems for practitioners, for whom it is not always easy to decide which list is most useful for their purpose and context (Schmitt, 2016). Given the paucity of systematic guidance on how to evaluate word lists, this study aimed to construct an evaluation tool that is based on Nation's (2016) framework of critiquing word lists, but is reformulated for a different purpose and for different target users. The tool is based on a thorough literature review, with data collected from practitioners regarding their views and uses of word lists and from consultations with ELT practitioners and word list experts.

The target users of the tool are ELT practitioners such as teachers, curriculum and assessment coordinators, and materials developers involved in directing vocabulary acquisition. The tool caters to practitioners with different levels of expertise and knowledge—especially those who are unfamiliar with the intricacies of developing corpus-based word lists. Specifically, the tool argues for a systematic procedure of assessing the suitability of a word list by starting with laying out the specification of the examined word list, the intended purpose of use, intended learners and educational context, the particular type of vocabulary targeted and then considering the unit of counting, the corpus, the approach to making word lists, and issues related to implementing word lists.

An initial explorative investigation on practices and views related to word lists was needed before designing the evaluation tool to better understand corpus-based word lists use in ELT. The few publications about how word lists are used were limited to EAP contexts and a specific practice base. Responses collected online from 74 practitioners, learners and researchers from a wide range of

domains in ELT revealed how word lists are exploited and perceived in ELT. Findings related to challenges to using word lists and issues related to implementing word lists (as well as other elements) informed the design of the initial version of the evaluation tool.

In its development through various iterations, the tool was revised by two panels of experts in corpus-based word list development and use (14 in total), and it was tested in real-life ELT scenarios by 11 practitioners who have been using word lists. The quantitative and qualitative data collected from the participants in three phases of the study suggested that the tool, and its supplementary materials, could help ELT practitioners assess the suitability of a word list for their purpose in a structured and thorough manner, although some reviewers raised issues regarding the tool's practicality and level of difficulty for some users. A secondary contribution of the study was to highlight the gap between corpus-based word list research and practice, and to raise awareness of word lists among ELT practitioners. The study aimed to bring the practitioner's voice in word lists research and evaluation. Finally, this study contributes to research related to the development of pedagogical word lists and improving their implementation.

ACKNOWLEDGEMENTS

The first and foremost word of thanks goes to Almighty Allah as with His blessings this journey with its joyful and difficult times was successfully completed. I would like to express my sincerest gratitude to my two supervisors, Dr Paul Thompson and Dr Amanda Patten, for their invaluable guidance, great patience and continuous support during this academic journey. They took me from being a novice to an accomplished researcher. It has been an honour and a privilege to have grown with both of you. A special thank you goes to Paul for his generous support when I needed it and for all the conversations we had. I also would like to express my deep appreciation to my examiners.

I would also like to thank King Saud University and the UK Saudi Arabian Cultural Bureau for my academic scholarship and the financial support to help me complete my PhD.

I would like to express my profound thanks to all the participants for their time and insights during my project, and for Emeritus Professor Paul Nation who chose not to participate in this project but was generous in his wise feedback to the tool that he gave in response to the couple of emails that I sent him. These comments supported what I did though they did not inform the research design.

It is my honour to express my deepest gratitude to my loving family: My parents, Dr Nasser Alzeer and Hind Bin Hamid, my sisters and my brothers for their unwavering love and encouragement throughout my life. I am grateful for my family for always being there for my kids and me. My warmest thanks are extended to my husband Muhannad, my companion at Birmingham, and my kids, Lulu and Saif, who waited for Mama and whose kisses and smiles sustained me in my work.

TABLE OF CONTENTS

Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Tables	xi
List of Figures	xii
List of Abbreviations	xiii
Glossary	xiv
1 Chapter 1. Introduction	1
1.1 Which Words Language Learners Need to Learn.....	1
1.2 Who Makes and Uses Word Lists.....	3
1.3 The Definition of Word Lists in the Context of English Language Teaching.....	5
1.4 Word Lists and the Language Practitioner	6
1.5 The Need for an Accessible Evaluation Tool	7
1.6 The Aims of This Study	9
1.7 Research Questions.....	10
1.8 Overview of This Study.....	12
1.9 Overview of This Thesis.....	13
2 Chapter 2. Literature Review	14
2.1 Introduction	14
2.2. A Historical Overview of Word Lists.....	14
2.2.1 Early Word List Development.....	14
2.2.2 Word Lists Based on Manual Analysis of Text Collections	16

2.2.3	A Breakthrough in Word List Development: Digital Corpora	20
2.2.4	Word Lists for Academic Vocabulary: A New Direction	22
2.3	An Overview of the Most Influential Corpus-based Word Lists Used in ELT	24
2.3.1	General English Word Lists.....	28
2.3.2	General Academic Word Lists	31
2.3.3	Specialised Word Lists	34
2.3.4	Multi-word Unit Word Lists.....	36
2.4	Issues in Corpus-based Word List Construction.....	38
2.4.1	Counting Words.....	39
2.4.2	Corpora and Word Lists.....	44
2.4.3	Approaches to Making Word Lists	45
2.5	Corpus-based Word Lists for Language Teaching	49
2.5.1	The Value of Word Lists in Language Teaching	51
2.5.2	Using Word Lists for Vocabulary Teaching and Learning.....	53
2.5.3	Issues with Using Word Lists for Vocabulary Teaching and Learning	57
2.6	Evaluating Corpus-based Word Lists.....	62
2.6.1	What Is Evaluation?.....	62
2.6.2	Evaluations of Corpus-based Word Lists.....	63
2.6.3	Lexical Coverage in a Corpus.....	64
2.6.4	ELT Practitioners' Perceptions of Usefulness	67
2.6.5	A Tool for Word List Evaluation.....	68

2.6.6	The Process of Developing Evaluation Checklists and Tools	72
2.7	The Research Gap and Rational of the Study.....	74
2.8	Summary	76
3	Chapter 3. Designing and Revising the Corpus-based Word List Evaluation Tool.....	78
3.1	Introduction	78
3.2	Research Design	78
3.3	Phase 1: Survey of the Uses of and Views on Word Lists in ELT	81
3.3.1	Aim	81
3.3.2	Participants.....	82
3.3.3	Data Collection and Analysis	85
3.3.4	Results of the Preliminary Survey	89
3.3.5	Themes Generated by the Survey.....	95
3.3.6	A Summary of How Word Lists are Used	103
3.4	Phase 2: Design of the Initial Version of the Proposed Evaluation Tool.....	105
3.4.1	The Adaptation Process.....	105
3.4.2	Literature Reviewed to Develop the Proposed Tool.....	108
3.4.3	Organisation of the Initial Version of the Proposed Tool (Appendix C).....	112
3.4.4	How is Nation's (2016) Framework Different from the Proposed Tool?	123
3.5	Phase 3: Initial Review of the Proposed Framework	124
3.5.1	Aim	124
3.5.2	Guiding Questions	124
3.5.3	Participants and Their Contexts	125

3.5.4	Data Collection	127
3.5.5	Procedure	129
3.5.6	Thematic Data Analysis	130
3.6	Achieving Trustworthiness	132
3.7	Research Ethics and the Confidentiality of the Participants	133
3.8	Summary of the Reviews – Results	137
3.9	Revisions to the Initial Tool	144
3.10	Summary	146
4	Chapter 4. Reviewing and Testing the Corpus-based Word List Evaluation Tool	148
4.1	Introduction	148
4.2	Aim	148
4.3	Research Questions.....	149
4.4	Study 1: Expert Review.....	149
4.4.1	Participants.....	149
4.4.2	Data Collection	151
4.4.3	The Critical Feedback Survey.....	151
4.4.4	Data Analysis	153
4.4.5	Results	154
4.5	Study 2: Practitioners Testing the Tool	163
4.5.1	Participants.....	164
4.5.2	Procedure	167
4.5.3	Data Analysis	167
4.5.4	Results	168

4.6	Research Ethics and the Confidentiality of the Participants	180
4.7	General Conclusion Based on Study 1 and 2	181
4.8	Final Adjustments to the Evaluation Tool	183
4.9	A Training Intervention for Corpus-based Word Lists: Plan A.....	184
5	Chapter 5. General Discussion	187
5.1	Practices and Views related to Word Lists in ELT.....	187
5.1.1	Word List Exploitation in ELT.....	187
5.1.2	Practitioners' Attitudes towards Word Lists	192
5.1.3	Evaluating Word Lists for Pedagogical Use	195
5.2	The Evaluation Tool of Corpus-based Word Lists.....	198
5.2.1	The Usefulness of the Tool in Achieving its Intended Purpose.....	199
5.2.2	The Content of the Tool: Accessibility, Comprehensiveness and Relevance	203
5.2.3	The Efficiency of the Tool	204
5.3	A Guide for Corpus-Based Word Lists in the Language Classroom (Appendix M).....	208
5.4	Summary	209
6	Chapter 6. Conclusion.....	210
6.1	Reflection on the Research Process and Results.....	210
6.2	Contributions of This Study	214
6.3	Implications for ELT Research and Practice	217
6.4	Limitations of This Study	218
6.5	Recommendations for Future Research	220
6.6	Conclusion	221
	References	223
	Appendices	243

Appendix A: Corpus-Based Word Lists for English Language Teaching (2000-2020).....	243
Appendix B: Survey of The Uses of Word Lists for Language Teaching and Learning (Phase 1).....	255
Appendix C: The Initial Version of the Evaluation Tool (Phase 2)	263
Appendix D: Interview Questions (Phase 3).....	277
Appendix E: Coding Scheme (Phase 3).....	278
Appendix F: Revisions to the Proposed Tool Based on the Interviews (Phase 3).....	280
Appendix G: Introduction to Corpus-Based Word Lists (Phase 3)	284
Appendix H: The First Iteration of the Evaluation Tool.....	285
Appendix I: Critical Feedback Survey (Phase 4)	311
Appendix J: Interview Protocol – Study 2 (Phase 4).....	313
Appendix K: Coding Scheme (Phase 4).....	314
Appendix L: The Final Iteration of the Evaluation Tool.....	315
Appendix M: A Guide for Corpus-Based Word Lists in the Language Classroom	341

LIST OF TABLES

Table 2.1 The Vocabulary Size and Text Coverage in the Brown Corpus.....	26
Table 2.2 Bauer and Nation's (1993) Word Family Levels	40
Table 2.3 Pedagogical Purposes of Word Lists for Language Teaching	55
Table 2.4 Nation's Questions for Critiquing a Word List.....	70
Table 3.1 The Purposes for Which Participants Used Word Lists	90
Table 3.2 Types of Corpus-Based Word Lists Participants Reported	90
Table 3.3 Names of Reported Word Lists.....	92
Table 3.4 Measures of Good Word Lists	94
Table 3.5 Literature Reviewed to Develop the Proposed Tool	108
Table 3.6 Changes to Nation's Questions Related to the Unit of Counting	117
Table 3.7 Changes to Nation's Questions Related to the Main List	119
Table 3.8 Changes to Nation's Questions Related to the Corpus	120
Table 3.9 Changes to Nation's Questions Related to Making the List	121
Table 3.10 Changes to Nation's Questions Related to Implementing the List.....	122
Table 3.11 Reviewers Profiles	126
Table 4.1 The Expert Panel Profile	151
Table 4.2 Practitioners Profiles	165
Table 4.3 Example Theme and Codes	168

LIST OF FIGURES

Figure 2.1 Using Word Lists for Vocabulary Instruction Employing the Four Strands Principle	54
Figure 3.1 The Process of the Research Design.....	80
Figure 3.2 Roles of Respondents.....	83
Figure 4.1 Criteria for the Critical Feedback Survey.....	152
Figure 4.2 Experts' Responses Regarding the Effectiveness of the Tool.....	155
Figure 4.3 Experts' Responses Related to the Content and Evaluation Questions.....	157
Figure 4.4 Experts' Responses Related to the Efficiency of the Tool	160

LIST OF ABBREVIATIONS

ELT	English Language Teaching
EGP	English for General Purposes
EAP	English for Academic Purposes
ESP	English for Specialised Purposes
EGAP	English for General Academic Purposes
ESAP	English for Specialised Academic Purposes
EFL	English as a Foreign Language
ESL	English as a Second Language
L2	Second Language
L1	First or Native Language

GLOSSARY

Corpus (plural: corpora): A collection of naturally occurring spoken or written text (of any size, even just one word) built to be representative of some aspect of language. These texts can include textbooks or textbook chapters, journal articles, literary texts, graded readers, teaching materials, audio-recorded conversations or lectures or movies.

Corpus analysis: A form of investigating a large body of written and spoken texts, i.e., corpus, using special computer programs to generate a huge amount of detail about language in use.

Dispersion: A criterion of word selection in a corpus. Dispersion is a number calculated by looking at the frequency of a word in each of the sub-corpora or text, i.e., how evenly a word occurs across different texts or sub-corpora.

Flemma: A unit of counting words in a corpus. A flemma is a more inclusive version of the lemma (see lemma's definition). A flemma includes inflected forms of different parts of speech. For example, the forms *developing* (*adj.*) and *developing* (*v*) belong to one *flemma*.

Frequency: A criterion of word selection in a corpus. Frequency is the number of occurrences of a word in the entire corpus. For example, 'the' is the most frequent word in English.

Lemma: A unit of counting words in a corpus. The lemma consists of a stem with inflected forms of the same part of speech. For example, the lemma *develop* includes *develops*, *developed*, and *developing*.

Lexical coverage: There are two senses of the term lexical coverage, both related to word lists:

- The percentage of words in a corpus that are covered by a word list; the higher the coverage, generally the better a list is.
- The percentage of words in a piece of discourse that language learners must know to comprehend it.

Range: A criterion of word selection in a corpus. Range is the number of different texts or sub-corpora that a word occurs in.

Stem: The basic form of a word (e.g. the verb 'go'). It is also called base or headword.

Token: Each occurrence of a word (every sequence of letters separated by space or punctuation in a spoken or written text) is counted as a token. For example, the sentence *counting₁ words₂ is₃ difficult₄ but₅ it₆ is₇ fun₈* contains eight tokens or as sometimes called **running words** or **word forms**.

Type: A unit of counting words in a corpus. A type is every different word in a corpus counted once. For example, the sentence *counting₁ words₂ is₃ difficult₄ but₅ it₆ is₇ fun₈* contains eight tokens but seven types, the word type (*is*) is counted once on its first appearance.

Word family: A unit of counting words in a corpus. A word family consists of a stem with its inflected and derived forms. For example, the word family *develop* includes *develops*, *developed*, *developing*, *undeveloped*, *underdeveloped*, *development*, *developments*, *developer*, and *developers*.

Word list: A sorted list of single or multi-word units grouped by their frequency counts in a corpus or other measures deemed to be important for language learning or an aspect of a language. It may be called vocabulary list or frequency list.

1 CHAPTER 1. INTRODUCTION

1.1 Which Words Language Learners Need to Learn

Knowledge of vocabulary is essential for acquiring a second language (L2) and for effective communication (Schmitt, 2010). One way to set vocabulary learning goals for L2 learners is to look at the size of the vocabulary of native speakers and consider that to be the target vocabulary size L2 learners need (Nation, 2006). Goulden et al. (1990) estimated the vocabulary of a highly educated native speaker of English to be approximately 20,000 word families, excluding proper nouns. Treffers-Daller and Milton (2013) found a lower average figure at 10,000–11,000 word families. Brysbaert et al. (2016) estimated the vocabulary of 20-year-old native English speakers at 11,100 word families. A general conclusion based on these studies—despite variations in methodological choices and limitations of vocabulary size tests—is that educated native speakers of English know approximately 10,000 to 13,000 word families (Schmitt & Schmitt, 2020). For non-native speakers to reach such numbers during the language learning process is a very ambitious goal, considering that L2 learners usually have limited exposure to authentic language input and limited time for language learning (Nation, 2006).

A more manageable approach to setting goals for learning vocabulary for L2 learners is to address the question of how much vocabulary they need to perform a particular activity in the target language, such as reading a book, watching movies or conversing (e.g. Nation & Waring, 1997; Schmitt et al., 2017). Laufer and Ravenhorst-Kalovski (2010), among others, have tried to answer this question by exploring lexical coverage, that is, the percentage of words in a text that must be known to comprehend it. They suggested that coverage of approximately 98% is needed for optimal reading comprehension and 95% to understand a written text adequately. To put this in more concrete terms, Nation (2006) estimated that 98% coverage equates to knowing around 8,000–9,000 word families for unassisted comprehension of written text and 6,000–7,000 word families for complex spoken language. Nation (2006) concluded that 3,000 word families plus proper nouns are needed

for 95% comprehension of spoken text and 5,000 word families to read authentic non-simplified texts. van Zeeland and Schmitt (2013) found that 95% lexical coverage is adequate for listening comprehension reducing the vocabulary size language learners must know to at least 2000- to 3000-word families. Although this is a far more practical estimate from a pedagogical standpoint, these calculations show that many words must be mastered in order to be a successful language user.

Another way to approach this lexical demand is to direct the learning of vocabulary towards important words (Nation, 2013). Corpus analysis, an approach to investigating a large body of written and spoken texts using special computer programs, has been fundamental for identifying lists of important words based on the frequency of their use (O’Keeffe et al., 2007). This approach assumes that it is more useful to learn the verb ‘to go’, for example, than to learn the verb ‘to increase’, because there is a higher probability of encountering the former in a general English text (according to frequency counts in Dang & Webb, 2016b). By prioritising certain vocabulary in this fashion, learners can optimise their efforts and derive maximum utility from their learning (Nation, 2013).

Frequency-based analysis can be traced back to the Vocabulary Control Movement in the early 20th century (Schmitt, 2000, 2010). In that movement, a group of researchers and professionals in English Language Teaching (ELT) endeavoured to systematize vocabulary acquisition in L2 learning by identifying a minimum core vocabulary to help language learners maximise their learning effort (Schmitt, 2000). Although frequency was emphasised most often, the movement represented three approaches to selecting and organising words into word lists. The first was the manual counting of the occurrences of words in a corpus, such as e.g. Thorndike’s (1921) *The Teacher’s Word Book* based on frequency and range. The second was the intuitive selection of important words in works like Ogden’s (1930) *Basic English*. The third was a mixed approach where corpus analysis information was adjusted subjectively to address the needs of the target learners and the purpose of the list (e.g. Faucett et al., 1936). The movement resulted in one of the most influential word lists in ELT, the

General Service List (GSL: West, 1953) which identified the most useful 2,000 words in English (Gilner, 2011).

From the 1950s through the 1980s, language descriptions based on corpus analyses were deemed invalid by Chomsky's criticism of corpus linguistics (Tognini-Bonelli, 2010). The dominant audio-lingual method and communicative approach during this period deemphasized vocabulary instruction, and word lists were disregarded in favour of incidental vocabulary learning (González-Fernández & Schmitt, 2017). However, as corpora became computerised and corpus analysis tools became available in the 1980s, corpus linguistics and word list research were revolutionised. An article by Paul Meara (1980) and a book by Paul Nation (2001) called *Learning Vocabulary in Another Language* helped to re-establish the value of vocabulary research and revived interest in word lists (González-Fernández & Schmitt, 2017). In the last two decades, many word lists have been compiled to identify the words that language learners should learn first. Prominent among them are the most frequent 2000 British National Corpus/Corpus of Contemporary American English word families (BNC/COCA2000: Nation, 2012) and the Essential Word List for Beginners (EWL) by Dang and Webb (2016b). In addition, researchers have compiled word lists of general academic English to address the needs of university L2 students, the Academic Word List (AWL; Coxhead, 2000) being the most famous example. More recently, researchers have shifted their focus towards specialised word lists that pertain to particular disciplines such as applied linguistics (e.g. Khani & Tazik, 2013), chemistry (e.g. Valipouri & Nassaji, 2013) or medicine (e.g. Lei & Liu, 2016). In addition, lists of multi-word units (more than a single word, e.g. phrases) have received increased attention. Of these, the Academic Formulas List (AFL; Simpson-Vlach & Ellis, 2010) is well known for English for Academic Purposes (EAP).

1.2 Who Makes and Uses Word Lists

The construction of word lists often involve frequency analysis that uses computational techniques or may be entirely driven by teacher intuition of what is helpful for their students. Word

lists have been used in several ways, depending on the user's interest(s). For example, researchers on authorship attribution might focus on the most frequent lexical items in some generated word frequency lists to determine whether the texts were written by the same author (Archer, 2009). Researchers in computational linguistics have used word frequency in the design of text processors and speech recognition and machine translation systems (Kilgariff et al., 2014). Psychologists (e.g. Aitchison, 2012; Tanaka-Ishii & Terada, 2011) interested in the effect of word frequency on language production, comprehension, and acquisition have used frequency lists (e.g. the MRC Psycholinguistic Database; Wilson, 1988) as a source for choosing words for psycholinguistic experiments. Lexicographers have relied on frequency data to contribute to the development of dictionaries. The *Collins COBUILD English Language Dictionary* (Sinclair, 1987) is the first English dictionary based on corpus evidence. Systems that generate and use word lists in areas relating to language learning include estimating the nature and size of vocabulary, exploring changes in vocabulary over time and assessing the readability of a text (Nation & Webb, 2011).

For educationalists, corpus-based lists have long been an established feature of vocabulary research and development. Researchers compiled word frequency lists based on native children's reading texts for use by researchers and teachers in L1 and L2 contexts (e.g. Stuart et al., 2003). Interestingly, *The Teacher's Word Book of 30,000 Words* (Thorndike & Lorge, 1944), one of the most influential lists in ESL, was prepared for native speakers of English.

In second language teaching, word frequency lists are used among other things for learning vocabulary. Nation (2016) elaborated: 'Word lists lie at the heart of good vocabulary course design, the development of graded materials for extensive listening and extensive reading, research on vocabulary load, and vocabulary test development' (p. x). Word lists can be used in different contexts and for different purposes. The current study focuses on using word lists to inform English vocabulary learning and it caters to ELT practitioners who lack expertise and background about word lists. The

next section defines word lists in the context of this study. That is followed by an introduction to the research problem and how it was addressed.

1.3 The Definition of Word Lists in the Context of English Language Teaching

Word lists, also known as vocabulary lists or frequency lists, are much-used resources in many disciplines. However, there is no consensus about what a word list is. ‘Word list’ has been used concurrently with ‘frequency’ to describe the process and the product: lists of words or phrases, identified and ranked according to their frequencies in a corpus (e.g. Nation, 2016; Palinkašević, 2017; Szudarski, 2018). While frequency is an obvious criterion, other subjective and statistical criteria (such as range and dispersion) are used for word selection in the context of language teaching. Word lists can be driven entirely by teacher intuition of what is important for the learner. An example is Ogden’s (1930) *Basic English* consisting of 850 essential words in English (Fries & Traver, 1950). Miller and Biber’s (2015) definition of word lists allows for frequency as one criterion, among others. They called them ‘lists of the words or lexical phrases found in a discourse domain, often including frequency information for the use of particular words’ (p. 30). Jones and Durrant (2010) emphasised the importance of human judgments to cater to the specific lexical needs of particular groups of learners when selecting words (Jones & Durrant, 2010). The definition by Webb and Nation (2017, Glossary section) highlighted the importance of the words for inclusion in a list. Their focus was on the ‘words that are deemed to be particularly important for learning a language or an aspect of a language’. Throughout this thesis, the following definition is used for the term ‘corpus-based word list’ — ‘word list’ for short. It is a sorted list of single or multi-word units grouped by their frequency counts in a given corpus, sometimes combined with other measures, deemed to be important for language learning.

For the purpose of this present research, the tool at this point focused on single-word lists, though one participant chose to use the tool to evaluate multi-unit word lists (see Section 4.5. Phase 4, Study 2). Evaluating multi-unit word lists require different criteria, and so it was decided not to

include them in this version of the tool. Word frequency is the most widely used criterion in deriving word lists from a corpus. All the word lists used to test the tool were informed by word frequency combined with other criteria. Word lists driven entirely by teachers' or other experts' intuition were eliminated in this study. These are limited in number, and assessing their suitability would be very context-specific. It would require the user to have extensive expertise in the target context and word list development while this tool aims to make the process of assessing and adapting word lists more accessible and efficient to ELT practitioners. The tool targets all three kinds of word lists: general English, general academic and specialised word lists (see Section 2.3. for more details of these kinds of word lists). In the testing Phase (Study 2, Phase 4), there were six general English word lists, three general academic lists and two specialised ones examined using the proposed tool. An operational definition of a word list in the proposed tool is a list of single words identified by a statistical analysis of their occurrence in a corpus, sometimes combined with other subjective measures, to serve the purpose of directing English language teaching/learning.

It is worth noting here that frequency, in this thesis, is discussed in relation to observed instances of words within corpora and not to the number of encounters with words that language learners have. In other words, the focus is on frequency as a textual phenomenon and not on second language acquisition.

Following this brief introduction to corpus-based word lists, this chapter traces the development of corpus-based word lists in the context of ELT. It describes what motivated this study, highlights the importance of the study, and summarises how the results of the study contribute to ELT. It states the research questions that guided each phase of this research. It finishes with a synopsis of the research design and the thesis structure.

1.4 Word Lists and the Language Practitioner

The rapid growth in pedagogical word lists indicates the value that people see in them. Research on making and using word lists, however, has not kept pace with the growing number of

word lists. Unfortunately, there are only a few publications on the applications of corpus-based word lists for language teaching and learning purposes targeted for the language practitioner; examples of such texts are Folse and Youngblood (2017), Lessard-Clouston (2013) and Nation (2016). Notable resources for English language classroom and research purposes in this area include: *The English Vocabulary Profile* (Capel, 2012) which offers information about which words and phrases are known and used by learners at each level of the Common European Framework (CEF); The Word List Highlighter (Haywood, 2007) which is a tool to highlight words from a list in a text, and vocabulary profilers such as LEX Tutor (Cobb, 2015) and Text Inspector (2021). These are used to check and identify the level of a text or to know which words in a text should receive more focus. However, the underlying assumption behind word lists and guidance on how they should be used by practitioners, particularly those with limited expertise and knowledge in corpus linguistics and access to published research, is not an area of attention yet.

As a novice teacher at a Saudi university, I became interested in word lists and how I could use them as an aid and resource for vocabulary development with my university EFL students majoring in English literature and applied linguistics. However, I faced difficulties in understanding the motives for developing word lists, and what decisions are related to selecting and using word lists. I thought about what a word list is and which word list to choose (I had several options) and how I could use them effectively. But I could not find good explanations. I became aware that a one-size-fits-all list does not exist. Reading the abstract of Schmitt's (2016) conference presentation, my interest turned towards evaluating word lists for the purpose of selecting and implementing them in ELT contexts.

1.5 The Need for an Accessible Evaluation Tool

The availability of electronic corpora and advances in corpus analysis tool has led to a proliferation of corpus-based word lists developed for different purposes. This wide variety of

available lists has created problems for practitioners and language learners, since it is not always easy to determine which list is most useful for a given purpose and context.

Researchers, acknowledging the pedagogical value of word lists, have attempted to evaluate many published lists (e.g. Dang & Webb, 2016a; Hartshorn & Hart, 2016; Nation, 2016; Schmitt, 2016). While these evaluations could be used as a second basis of evaluation by practitioners who are planning to use a word list, there were issues:

- Most evaluations focused on widely known published lists. For example, researchers like Hernandez (2017) explored the coverage of the AWL in various contexts and compare it against the Academic Vocabulary List (AVL; Gardner & Davies, 2014). However, teachers may choose to use unpublished lists or make their own.
- Most of these evaluations did not include certain elements to judge their usefulness (such as fit with learners' needs or proficiency level), which would be informative for specific pedagogical purposes (Brezina & Gablasova, 2017).
- The evaluations relied on calculating the lexical coverage of the list in a corpus. They, however, reported different results when different criteria were used (Dang & Webb, 2016a).

These issues are discussed in depth in the literature review in Section 2.6, but the main conclusion is that despite the value of such evaluative research, practitioners often do not have access to it, and even if they have it is not necessarily understandable to them. My key reference in dealing with word lists and word list evaluation at that point was Paul Nation's (2016) book about word lists, *Making and Using Word Lists for Language Learning and Testing*. In this book, Nation (2016, pp. 131-142) developed a framework to critique his word lists. It is a set of Yes/No questions that focus on making word lists. However, his framework does not appear until after the 12 chapters in his book, so the questions might be difficult to interpret for a practitioner with limited knowledge of word list construction and ability in corpus analysis. Using a word list or selecting one for a

pedagogical purpose raises other issues superficially addressed in the framework, though some are covered in Nation's (2016) book.

Schmitt and Schmitt (2020), in the second edition of *Vocabulary in Language Teaching*, highlighted the importance of 'judging' word lists for the purposes for which they will be used by examining the list's development and content. This is a good proposal that overlaps with the purpose of Nation's (2016) framework, but there are two problems. First, in addition to the daunting task of exploring the large number of lists, few published word lists clearly state the purpose(s) and target audience for which they were created (Schmitt, 2016). The other challenge is that making such a decision requires some level of skill and knowledge about the creation of corpus-based word lists (Schmitt & Schmitt, 2020).

Given this lack of accessible guidance on how to evaluate a list, this thesis aimed to construct an evaluation tool for use by ELT practitioners based on Nation's (2016) framework and consultations with ELT practitioners. The tool developed in this thesis addresses practitioners who are not expert in corpus-based word lists, and it focuses on *using* word lists for pedagogical purposes. Thus, it can also be seen as a response to a call to make existing lists more suitable to particular ELT contexts rather than publishing new word lists (Dang, 2020).

1.6 The Aims of This Study

The current study is motivated by the paucity of evaluative research in word lists for English language teaching and learning that considers the needs of ELT practitioners, and it aims to specify the factors that must be considered when selecting a word list and implementing it in the language classroom. It is proposed that the most important factors are the objective of using the list, the size of the list, the unit of counting, the characteristics of the learners and the type of vocabulary (Dang et al., 2020), but other major factors are substantial as well.

This study develops a tool to evaluate corpus-based word lists to help ELT practitioners assess the suitability of such word lists for pedagogical use in various contexts. This evaluation tool

would provide ELT practitioners with a formal structure from which to consider situational factors that are likely to differ across instruction scenarios and the ways they can be optimised. This evaluation tool is based on three major components:

- The framework developed by Nation (2016);
- A review of the relevant literature on the development of word lists and their use for pedagogical purposes;
- A survey of contemporary practices and views related to corpus-based word lists in ELT conducted for the current study.

A secondary aim was to find out how word lists are exploited and perceived by ELT practitioners. Word lists are fundamental in preparing L2 learners to be competent in the use of the target language. They have been used for so long, yet their value has an uneven past. While they were highly valued in the early 20th century, they were later neglected. Corpus analysis tools helped in re-establishing interest in word lists late in the 20th century. However, little is known about how word lists are perceived and used and the extent to which they are useful from the point of view of ELT practitioners and learners. The current research aimed to explore these practices and views and how they have influenced the usefulness of word lists in ELT.

Other aims which emerged during the study were to highlight the value of corpus-based word lists and to raise awareness of them and how they can be implemented effectively in ELT contexts. Exploring how word lists are used and perceived shed light on the gap between research on corpus-based word lists and ELT practice.

1.7 Research Questions

Motivated by the observation that a one-size-fits-all word list does not exist and that practitioners need to assess the suitability of word lists before using them, this study began with a broad aim: to develop a tool that ELT practitioners can use to make such evaluations. The nature of the study was exploratory and sequential. Therefore, the research questions were formulated for

each phase as the study evolved, guided by what was already known in the field, identifying and addressing real-world problems and building on what was found in each phase (Tashakkori & Creswell, 2007). The research questions were investigated using different data collection methods and different participants.

Before developing the tool, it was necessary first to understand the status of word lists in ELT contexts—who uses them, for what purposes and in what ways—and how those practices are perceived. For this purpose, an initial survey of contemporary practices and views related to corpus-based word lists in ELT was conducted. The following research questions guided this initial exploration:

- RQ1: What distinct uses are made of corpus-based word lists by practitioners, researchers and learners involved in L2 teaching and learning?
- RQ2: What methods are used to generate pedagogical corpus-based word lists?
- RQ3: Which measures can be used to evaluate a word list's utility for ELT practitioners, researchers and language learners?
- RQ4: How are word lists perceived by ELT practitioners, researchers and learners?

Based on the quantitative and qualitative results of the initial survey, the literature review on word lists and Nation's (2016) framework, I developed the first version of the tool. The next step was to subject the tool to an initial review. The tool in different iterations was examined by different participants in two phases. These phases addressed the following research questions:

- RQ5: To what extent does the proposed evaluation tool meet its intended purpose, according to expert users and evaluators of corpus-based word lists?
- RQ6: To what extent does the proposed evaluation tool meet its intended purpose, according to potential users of the tool (e.g. language teachers)?

In the context of this project, the extent to which the intended purpose is met was assessed by effectiveness, comprehensiveness and efficiency. 'Effectiveness' refers to the tool's capacity to

generate an intended result, in this case, an evaluation of one or more corpus-based word lists for pedagogical use. The ‘comprehensiveness’ and ‘relevance’ of the tool’s content are concerned with the questions and supporting materials: the extent to which the tool addresses all the important elements of corpus-based word lists (comprehensiveness) and the extent to which the tool relates to different ELT contexts (relevance). ‘Efficiency’ is concerned with the physical and practical aspects of the tool – its clarity, accessibility, structure, time spent on it and ease of use. Findings from these questions were then used to revise the tool according to the reviewers’ feedback in these areas in Phase 3.

The same research questions were addressed in the fourth phase. In that phase, another iteration of the tool was examined by different participants in two studies. In the first study, another expert panel reviewed the tool. In the second, the tool was used to evaluate the suitability of word lists from the perspective of ELT practitioners in real-world scenarios

1.8 Overview of This Study

This study does not follow a conventional research design. Data collection and methodological choices made during this study were driven by a real-world problem, how potential users could evaluate word lists. They followed an iterative process of surveying needs, developing the tool, seeking feedback and revising the tool, then testing it again. To my best knowledge, there was no existing model for developing and testing evaluation tools that produced a qualitative conclusion and involved practitioners. Thus, as a researcher, I had to adapt to the needs I discovered in developing the tool. In addition, as the tool was targeted to ELT practitioners, I endeavoured to answer the research questions from their perspective.

This study was conducted in four phases. The first was a preliminary survey of current practices and views related to the construction, exploitation and evaluation of corpus-based word lists for English language teaching and learning. An initial set of quantitative data was needed to establish how practitioners view and use word lists in ELT contexts before it was possible to develop

a soundly informed tool for ELT practitioners. The second phase entailed developing the proposed tool, which was based on and guided primarily by (a) the content of Nation's (2016) framework, (b) a review of the relevant literature on word list production and use for pedagogical purposes and (c) the results of the initial survey. In the third phase, the tool was reviewed by a panel of experts and practitioners using interviews. The researcher then revised the tool based on their feedback. The fourth testing phase included two studies: a review to obtain critical feedback on the tool from a second panel of experts and a study in which the tool was used to evaluate the suitability of a word list, from the perspective of ELT practitioners, in real-world scenarios. Informed by the feedback from these two studies, the researcher revised the tool and developed a guide to introduce corpus-based word lists for practitioners in an accessible language.

1.9 Overview of This Thesis

The thesis comprises six chapters. The chapters reflect the research process and how each phase informed the subsequent phase. Chapter 2, the literature review, gives an overview of how word lists have developed through the centuries. It also addresses the main issues related to word list construction, exploitation and evaluation in ELT. It reviews the process of developing and testing evaluation tools. Chapters 3 and 4 report the process of developing and testing the tool. They present the methodology and results of the project. Chapter 3 treats the process of developing the evaluation tool. It presents the methods used, the results of each phase and how those results fed into the iterative process. The heart of the thesis, Phase 4 (the testing phase) is presented in Chapter 4 to support the usefulness of the proposed tool. The findings of the study's phases are discussed thoroughly in Chapter 5. Chapter 6 concludes the thesis by reflecting on the research, making implications for practice and research and setting out recommendations for future directions.

2 CHAPTER 2. LITERATURE REVIEW

2.1 Introduction

The tradition of compiling lists of important words to direct the process of vocabulary teaching and learning can be traced back to the Vocabulary Control Movement in the early 20th century (Fries & Traver, 1950). Early word list development involved intuitive judgements of which words were important or manual tallying of words' frequencies in a corpus. With the invention of computerised corpora and advances in corpus-analysis tools, word list research was revolutionised, and in the last two decades, a large number of word lists with various objectives in mind have been published. These lists largely reflect counting words' frequency and range in a corpus (Nation, 2016; Schmitt, 2010; Webb & Nation, 2017), based on the assumption that a word's frequency is commensurate with its importance (Nation & Waring, 1997). This review focuses on the literature related to the construction, exploitation and evaluation of word lists in ELT. It begins with an overview of the history of word lists and proceeds to present the most influential word lists developed for ELT purposes. Next, it discusses concepts and issues related to word list construction. Then it reviews the limited research on word list exploitation for ELT, reflecting the uneven status of word lists in ELT over the years and how this influenced practices and views related to word lists. The final section details the literature on evaluating word lists and conceptualising a 'suitable' word list. To the best of this researcher's knowledge, the literature on word list use and evaluation does not include a systematic approach to word list evaluation, which motivated the present study. The chapter closes with a brief review of the literature on evaluation tool development and testing to justify the research design employed to develop and test the proposed word list evaluation tool.

2.2. A Historical Overview of Word Lists

2.2.1 *Early Word List Development*

Word list development is often said to have begun in the 1920s with the Vocabulary Control Movement (Schmitt, 2000). However, the practice of identifying vocabulary that merits instructional

focus can be traced back to the 15th century, at which time it is evidenced by language-teaching materials that grouped words according to topics such as body parts or animals (Fries & Traver, 1950). Examples resembling modern word lists appeared in the 16th century in textbooks such as Claudius Holyband's (1566) *The Schoolmaster* and *The Littleton*, which included lists of French vocabulary arranged by topic; in Richard Mulcaster's (1582) spelling word list, which attempted to standardise English spelling (Howatt & Widdowson, 2004); and in Timothe Bright's (1588) *Shorthand in English* (Fries & Traver, 1950).

In the early 17th century, John Amos Comenius (1631) promoted the idea of a limited vocabulary, presented in context. He grouped 8,000 Latin words into topics and presented them with visual aids. Later, Paul Festau (1672) adapted his French textbooks to English instruction dialogue and vocabulary list, and Guy Miege (1685) published *Nouvele Methode*, addressing the complexity of word knowledge and the constitution of a word (Howatt & Widdowson, 2004). The majority of early works on word list compilation were geared towards stenographers or aimed to teach reading or spelling to native speakers of English (or to teach a language other than English) (McArthur, 1998), yet they demonstrate researchers' awareness of the notion of limited vocabulary, which later developed in the early 20th century as part of the Vocabulary Control Movement (Schmitt, 2000).

A reference point in lexicography is Samuel Johnson's milestone *Dictionary of the English Language*, published in 1755, which took the form of a word list. By the end of the 18th century, English was taught across Europe, and John Miller (1797) published the first known non-European textbook for English as a foreign language (EFL), *The Tutor*. The book included an English word list with Bengali translation. The grammar-translation method (GTM), which relied extensively on bilingual word lists (Zimmerman, 1997), had become popular in Europe in the 19th century; the goal of this method was to prepare students to read and write in classical languages and develop their skills in analysing, rather than using, language. Vocabulary used in this method of instruction was selected according to the targeted grammar points. Objecting to the GTM, Thomas Prendergast

(1864) published *The Mastery of Languages*, which included a list of 214 common English words that together covered a wide range of communicative functions. Prendergast compiled the list according to his intuition, but it was later found to closely resemble lists produced later in the 20th century (Howatt & Widdowson, 2004).

These early attempts at instructional word lists were purely intuitive, but they reflected an awareness among educational writers and researchers that some words are more likely to be useful to learners than others. Even though the majority of the early lists were not developed for language-learning purposes, they paved the way for word list development for English as a second language learning (ESL).

2.2.2 Word Lists Based on Manual Analysis of Text Collections

Early in the 19th century, attempts were made to objectively compile lists according to how often words appeared in a collection of texts (i.e., a corpus). Early attempts at corpus-based word lists can be traced back to Freeman's (1820) 20,000 word corpus, which he used for his work in adult education, and Pitman's (1837) corpus of 10,000 words, used to develop his system of stenography (McArthur, 1998). An exceptional reference point in the history of corpus design and use is Käding's (1898, as cited in Fries & Traver, 1950) manual frequency count of an 11-million word corpus for the purpose of identifying words important to the training of stenographers in Germany. This work demarcates the greatest corpus counts to develop early word lists (Fries & Traver, 1950).

In the United Kingdom, Knowles (1904) assembled a corpus of 100,000 words for the development of a reading system for the blind. Of these 100,000 words, he extracted 353 that he claimed covered 75% of the texts he collected (McArthur, 1998). Eldridge (1911), in the United States, searched for the highest-frequency words among 43,000 word corpus of four American newspapers, attempting to compile a limited but universal vocabulary list. Cook and O'Shea (1914) created a corpus from the personal correspondences of 13 adults in the United States to investigate the spelling-education needs of American children. They noticed that a small number of very

frequent words—763, to be precise—composed 90–91% of the examined texts (McArthur, 1998).

Dewey (1932) attempted to develop another frequency list for stenography, *The Relative Frequency of English Speech Sounds*. Dewey felt that previous lists had not drawn upon sufficiently broad groups of texts, and the 100,000 word corpus he developed as a result could be considered the first English equivalent of Kåding's corpus (McArthur, 1998).

Although these corpora were small in size, compared with modern examples, it is fascinating to see that researchers were aware of the unequal frequencies with which words were used (McArthur, 1998). This was later explained by Zipf's Law (Zipf, 1949) which states that as there is a patterned decrease in the frequencies of words, this frequency figure is multiplied by the rank order of an item in a list ordered by frequency. Thus, the most frequent word is about twice as frequent as the second most frequent word, approximately three times as frequent as the third most frequent, and so on. This rule states that a small set of words covers a very large portion of any text or collection of texts.

In time, it became more attractive to produce objective lists by counting words in a corpus than to simply rely on one's intuition. Modern word lists for language teaching purposes owe a great deal to the Vocabulary Control Movement, in which a group of researchers and ELT teachers in the 1920s and 1930s undertook the mission to identify a minimum core vocabulary to help language learners maximise their learning (Schmitt, 2000). These researchers – Edward L. Thorndike, Ernest Horn, Itsu Maki, Lawrence Faucett, Harold Palmer and Michael West – and their contributions are described in the following section.

A point of reference in discussions of word lists is *The Teacher's Word Book*, a substantial, manually counted word list assembled by Thorndike (1921). The *Word Book* detailed the frequencies of 10,000 words in a 4.5 million-word corpus of 41 sources, including the Bible, children's storybooks, textbooks, trade manuals and periodicals. In 1932, Thorndike revised and expanded the list to 20,000 words, also accounting for range (occurrence across texts). This 1932 work earned Thorndike credit

as the first researcher to employ both range and frequency. In 1944, collaborating with Irving Lorge, Thorndike revised the existing list yet again and published *A Teacher's Word Book of 30,000 Words*. Although developed for native English-speaking students, Thorndike's work was influential on ELT purposes. For example, the first academic vocabulary list, developed by Campion and Elley (1971), and the original and widely used Vocabulary Level Test, developed by Pawley and Syder (1983), were both based on Thorndike's lists (Nation, 2016).

Following Thorndike's approach, Horn (1926) developed *A Basic Writing Vocabulary* by selecting the 10,000 most frequent words from a corpus of five million, collected from 65 sources that included business and personal letters, meeting minutes, newspapers, and magazines. Horn also used frequency and range to develop the list, but unlike Thorndike, he separately reported all inflections and derived forms. *The Teacher Word Book* (Thorndike, 1921, 1932) and *A Basic Writing Vocabulary* (Horn, 1926) were later combined by Faucett and Maki (1932) to create the Faucett–Maki List, which was further revised by Michael West at the Carnegie Conferences to become the General Service List (Gilner, 2011).

Taking an adjusted approach, Palmer (1931) created the *General List of 3000* according not only to frequency and range but also to qualitative criteria, that is, input from teachers. The list was divided into six bands of 500 words. One of the most innovative developments of Palmer's list was the grouping of common lexical derivations under a main word. Palmer was the first to use headwords for list organisation and item selection, which began to shift list production towards the use of word families (Gilner, 2011).

In the early 1920s, prior to Palmer's *General List*, West was serving as an officer of the Indian Educational Service. There, he began to create reading materials to help his students develop their language by controlling vocabulary in reading texts, i.e., carefully distributing known and unknown lexical items to increase texts' readability. After extensive research and experimentation, he

eventually produced the *New Methods Readers Series* (1927) and a list of about 1,500 words for expressing ordinary ideas (Howatt & Widdowson, 2004).

Setting out to create a word list by a quite different approach than Thorndike and Palmer and for a different purpose, Ogden (1930) attempted to simplify the English language into 850 essential words, calling his list *Basic English*. Following his intuition, he identified 200 names of objects that could be represented visually, 400 general names, 150 qualities and 100 words needed to express important ideas (Fries & Traver, 1950).

At this point in the review, the Carnegie Conferences of 1934 in New York and 1935 in London—which were pivotal in the history of word list construction—merit discussion. The first Carnegie Conference brought together Faucett, Palmer, Thorndike and West to draw up a tentative General-English word list for foreign learners. The process of word selection was based on the researchers' experience and involved both objective and subjective criteria. The objective selection element was primarily based on the Faucett–Maki frequency list (1932), while the subjective criteria were guided by the principles of Palmer and West. At the second conference in 1935, the list was revised, as was the *Interim Report on Vocabulary Selection* published in 1936 (Gilner, 2011).

In 1953, West published *A General Service List of English Words* (the GSL), a reissue of the 1936 *Interim Report*, drawn from a corpus of 2.5 million words. The GSL, which contained 2,000 word families, included highly frequent items as well as certain items that were less frequent but which, according to West, were necessary to express certain ideas and their higher-frequency equivalents were not easily expressible. West excluded words that were colloquial, emotional, or overlapped with other words in the list.

During the period between 1950s towards 1960s, the audio-lingual method and the communicative approach were dominant. It was thought that vocabulary could be learned incidentally without any principled lexical development. Thus, vocabulary instruction and word lists were neglected (González-Fernández & Schmitt, 2017).

This historical review on word lists traces word lists' evolution across the centuries. Word list development started with researchers' compilations of what they perceived as essential vocabulary which surprisingly corresponded to frequency information later in the 20th century used to develop lists of words derived from corpus analysis. It also highlights the role of the Vocabulary Control Movement (Schmitt, 2000) which resulted in the creation of one of the most influential word lists in ELT, the GSL. Although frequency was emphasised most often, the movement represented three approaches to word lists: the objective corpus-driven approach (e.g., Thorndike), the subjective approach (e.g., Ogden) and the mixed method approach—corpus-driven information adjusted subjectively (e.g., Palmer and West). To date, debates continue to be held about which approach is best, with Thorndike's and West's approaches the most serious contenders.

2.2.3 A Breakthrough in Word List Development: Digital Corpora

Advances in computers allowed for collecting, storing, and analysing texts of millions of words. Computerised corpora allowed the quick and reliable calculation of frequencies, the identification of patterns of vocabulary occurrence, and usage which had a great influence on vocabulary research and helped renew interest in vocabulary instruction and word lists (O'Keeffe et al., 2007). The computerisation of corpora truly revolutionised word list development. The first computerised corpus was the Brown Corpus compiled by Kučera and Francis (1967), containing one million words drawn from American English texts printed in 1960. Between 1970 and 1978, the British English counterpart to the Brown Corpus, the Lancaster–Oslo/Bergen Corpus (LOB) of one million words, was being developed in Europe. Another early computerised corpus compiled for a lexical purpose was Carroll et al.'s (1971) five million-word American Heritage Intermediate (AHI) Corpus, which contained words from school textbooks. The American Heritage Word Frequency Book was compiled according to frequency and range.

The Brown Corpus and LOB Corpus signal the first generation of digital corpora, serving as both a model for corpus compilation and a resource for research. The AHI notwithstanding, one

million words was considered a substantial size, as written texts had to be typed in manually for analysis. By the 1990s, corpora of 100 million words were available (Kennedy, 1998). The first computerised mega-corpus compiled for lexicographical purposes was part of the Collins Birmingham University International Language Database (COBUILD) project, which began in 1980. The corpus was compiled by the publishing company Collins and a group of researchers at the University of Birmingham under the direction of John Sinclair. In 1991, the Bank of English (BoE) was launched, with new texts added every month and the final release (in 1996) containing over 320 million words (Kennedy, 1998).

In 1995, another large corpus was released: The British National Corpus (BNC), which consists of 100 million—90 million written and 10 million spoken—British English words. The written-language samples were drawn from regional and national newspapers, specialist periodicals, journals, academic books, fiction, letters, and school and university essays. The spoken samples were transcribed from informal conversations, formal business and government meetings, radio shows and phone calls. Unlike the BoE, the BNC was (and is) finite—no new texts have been added to it since its completion (Kennedy, 1998). The spoken component of the BNC has been updated to include 11.5 million word gathered from recorded conversations from 672 speakers from different parts in the UK between 2012 and 2016 (Love et al., 2017). The written update of the BNC (BNC2014 written) is being compiled and will be made publically available when complete.

In 2008, Mark Davies released the Corpus of Contemporary American English (COCA: Davies, 2008), representing American English from 1990 to 2008 updated annually (20 million words are added). It is a large corpus of one billion words evenly divided across eight genres: spoken, fiction, popular magazines, newspapers, academic texts and in March 2020 TV and Movies subtitles, blogs and other web pages were added (Davies, 2021). COCA is probably the only large, genre-balanced corpus and it is widely used by researchers, teachers and learners. Among others, the preceding corpora are featured prominently in this literature review.

2.2.4 Word Lists for Academic Vocabulary: A New Direction

After the success of the GSL, a hunt began for the most frequent words in academic texts to address the lexical needs of L2 university students. Early lists of academic vocabulary were compiled either manually, by counting occurrences of words in small academic corpora (e.g., Campion & Elley, 1971; Praninskas, 1972), or by noting which words students annotated in their textbooks (e.g., Ghadessy, 1979; Lynn, 1973).

Campion and Elley (1971) developed the Academic Vocabulary List as an EFL entrance examination for New Zealand universities. The list included 500 common academic vocabulary words, based on a corpus of 301,800 words collected at New Zealand universities from 44 lectures, journals and examination papers across 19 disciplines; it excluded 5,000 words appearing in Thorndike and Lorge's (1944) *Teacher's Word Book of 30,000 Words*. Praninskas (1972) developed the American University Word list (AUWL) as part of an English course for native Arabic speakers at a university. He created a corpus of ten textbooks, one from each course completed by first-year students at his host university. He then sampled the tenth page of each textbook, excluding words from the GSL and anything excluded by West from inclusion in the GSL, such as proper nouns, abbreviations and foreign words. The final list, which accounted for both frequency and range, included 507 base words and 840 derived forms, which comprised ten percent of the most frequently occurring words in each book.

Taking a different approach, Lynn (1973) analysed 52 annotated textbooks from 50 business students (the teaching materials were in English but were taught in students' native language). Lynn's list included 120 word families, which accounted for 20% of students' hand-written definitions in the corpus. Similarly, Ghadessy (1979) assembled a corpus of almost half a million words from 20 textbooks used by non-native English-speaking freshman studying biology, chemistry or physics at a university. He recorded the words for which students wrote definitions in the textbook, the definitions that students produced, and the parts of speech in context. The list included 795 most

frequently glossed items in this sample of university textbooks. This list was based on the frequencies with which students felt compelled to look up the meaning of a word, not necessarily the frequencies of the words themselves.

Those early attempts were later combined into the University Word list (UWL) by Xue and Nation (1984). The UWL contained a total of 836 words identified according to frequency and range and divided into 11 sub-lists to make the list more manageable for teachers and students. The GSL was excluded from the new list as the latter was seen as a complement to the GSL for university language learners. According to Nation (1990), the UWL accounted for 8% of all words in a typical academic text. The UWL was a considerable improvement, but the fact that it was based on two existing subjectively developed lists meant it inherited their limitations and therefore lacked consistent selection principles (Coxhead, 2000).

Coxhead noted the limitations of the UWL and set out to create a new academic word list. Taking advantage of the developing field, she compiled the first computerised academic corpus, comprising approximately 3.5 million running words from university textbooks and materials from four different academic disciplines, namely law, arts, commerce and science. From this corpus, she developed the Academic Word List (AWL; Coxhead, 2000), which contained 570 word families not included in West's (1953) GSL, selected according to frequency and range (Coxhead, 2000). The list provides a lexical coverage of 10% of her Academic Corpus and 8.5% of another academic corpus assembled for the purpose of evaluating the AWL (Coxhead, 2000).

This historical survey highlights the role corpora have played in the reorientation of word lists and vocabulary instruction. The Vocabulary Control Movement early in the 20th century was a great turnaround in the history of word lists which led to influential manually compiled word lists. However, as the natural or communicative approach became dominant in the mid-20th century, controlled vocabulary instruction in the language classroom was disregarded in favour of incidental vocabulary learning and interest in word lists faded. In the 1990s computerised corpora became

available and provided useful insights on English vocabulary patterns and usage. Proponent researchers of explicit vocabulary learning stressed the role of word lists for intentional vocabulary learning (Folse, 2004; Meara, 1980; Nation, 2001; Schmit, 2000). Yet it is only in the past twenty years, as large corpora have become accessible and more sophisticated corpus-analysis tools have been developed and made available, that word list development has truly flourished (Vilkaitė-Lozdienė & Schmitt, 2020). The following section reviews the most influential corpus-based word lists published in the last two decades.

2.3 An Overview of the Most Influential Corpus-based Word Lists Used in ELT

There are more than 90 corpus-based word lists of single and multi-word units for one variety of English or another; these lists, which are based on frequency counts, are published to help ELT practitioners to organise vocabulary instruction for different pedagogical purposes. A table adapted from Burkett (2017) that provides details for corpus-based word lists published between 2000 and 2020 can be found in Appendix A. Burkett reviewed over 100 word lists for ELT, of which around 90 were corpus-based word lists. The word list table provided in this thesis in Appendix A reviews corpus-informed lists of single words for general, academic and specialised English. Information about the development of this table and its role in this research is discussed in Section 3.4.3.1.

Word lists are classified based on the vocabulary type or frequency band of the words included. A well-known classification of English vocabulary is the three-tier model developed by Beck et al. (2013), which classifies words according to their utility and role: Tier One words are basic, everyday words; Tier Two words include wide-ranging words with high utility across various contexts; and Tier Three is limited to domain-specific words. Although their classification is targeted at English vocabulary for native-speakers of English, Beck et al. (2013) noted that their model overlaps with Nation's (2001) four classifications of words necessary for L2 learners: high-frequency vocabulary

(corresponding to Tier One), academic and specialised vocabulary (both corresponding to Tier Two), and low-frequency vocabulary (overlapping with Tier Three).

Another classification of vocabulary is based on frequency thresholds. In the second edition of *Learning Vocabulary in Another Language*, Nation (2013), placed English vocabulary along a continuum of high-, medium- and low-frequency words. He explained that, in English, there seems to be 2,000 high-frequency word families for basic language uses that deserve explicit instruction. At the end of the continuum is low-frequency vocabulary that includes words that are rare, beyond the 9,000 words, and could be useful for domain-specific learning. In between, learners need to learn around 3,000–9,000 word families to be able to engage in L2 in different contexts without support (with high-frequency words). This type is called mid-frequency words and they often need to be learned intentionally in the classroom as they are moderately frequent (Schmitt & Schmitt, 2014).

Word frequency, that is how often a word occurs in a proportion of the language, is an important variable that has informed researchers' and practitioners' decisions regarding a word's usefulness (Nation & Waring, 1997). Focussing initially on the most frequent words and then moving on to less frequent items is a practical solution to address L2 learners' need to learn many words (see Section 1.1. for a discussion of the number of words that learners need to learn). The reason for using frequency as a criterion for identifying target learning vocabulary is to maximise text coverage; thus, the learner will know a large proportion of a spoken or written text by learning a set of words (Nation & Waring, 1997). This is illustrated in Table 2.1., which shows text coverage in the Brown corpus by certain numbers of high-frequency words.

Table 2.1

The Vocabulary Size and Text Coverage in the Brown Corpus (taken from Francis & Kučera, 1982)

Vocabulary size	Text coverage
1,000	72.0%
2,000	79.7%
3,000	84.0%
4,000	86.8%
5,000	88.7%
6,000	89.9%
15,851	97.8%

Note. Adapted from *Vocabulary Size, Text Coverage and Word Lists* (p. 9) by I. S. P. Nation and R. Waring, 1997, Cambridge University Press.

As can be seen in Table 2.1., the text coverage becomes greater as the vocabulary size increases, but the increase in coverage becomes smaller and smaller at each level. Table 2.1. also shows that there is a point at which learning the next 1,000 words would be inefficient and that learning vocabulary for an intended English context, such as academic English, would be more beneficial. Debates about the point at which learners should move from high-frequency words to a more specific vocabulary to address their learning needs and the potential usefulness of different categories of words are discussed in Section 2.3.1.

Word lists have traditionally been biased towards frequency counts in L1 discourse based on the cost/benefit principle. This principle suggests that learners receive the best return for their learning effort by learning very frequent words, as it is highly likely that they will use and/or encounter these words (Nation, 2013, 2016).

The calculation of frequency counts and text coverage is based on the assumption that vocabulary is learned in order of frequency; that is, the first 1,000 words are learned before the second 1,000 words, the second 1,000 words are learned before the third 1,000 words, and so on (Nation, 2013). Although frequency ranking works relatively well until about the 4,000 band, it then rapidly becomes arbitrary depending on the corpus that is used to determine word frequency

(Aizawa, 2007). Early stage L2 learners may know words such as ‘pencil’ and ‘orange’ despite the fact that they have relatively low frequency. Although it is a widely used measure, frequency has limitations as a criterion for determining a word’s usefulness. It has been criticised for its inability to extract low-frequency words with high information content (Richards, 1970) and for failing to reflect a word’s importance adequately if used as the sole criterion for word selection (Egbert, 2018).

Schmitt et al. (2021) found that, while frequency counts help to predict the words that learners need to learn to read and listen (that is, for receptive vocabulary knowledge), they fail to predict the words that learners need to know for language production. The authors called for a shift from frequency-based lists to knowledge-based word lists, which they found to be better predictors of the words that learners needed to know to produce the target language (Schmitt et al, 2021) and this could be a focus of future research. A discussion of other statistical and subjective criteria that are used in conjunction with frequency to construct pedagogical word lists is presented in Section 2.4.3.

Taking account of their learning purpose, Nation (2016) classified word lists as general English, general academic and specialised word lists, and noted that frequency cuts through all vocabulary types. This classification scheme is used in the current study because the main aim is to evaluate word lists in terms of how effectively they fulfil their purposes. Frequency is an important criterion when developing or evaluating pedagogical word lists, but it should not be used in isolation. In fact, corpus-driven information should be adjusted subjectively to make the list as useful as possible for the target use and context. Teachers’ perceptions and learners’ knowledge are key criteria that assist in determining words’ usefulness (see Section 2.4.3. for a discussion of the approaches to making word lists). White (1988) proposed seven criteria for selecting words for teaching, namely frequency, coverage, range, availability, learnability, opportunism (relevance to the learners’ situation) and interest. West’s (1953) efficiency criteria constitute another approach to producing pedagogical word lists by adjusting corpus-driven data subjectively.

2.3.1 *General English Word Lists*

General English word lists are designed to prioritise the basic vocabulary that language learners should learn first (Nation, 2016). Such lists reflect the Tier One of Beck et al. (2013), tend to be small, containing around 2,000–3,000 words, but cover around 70%–90% of texts of several sorts (Nation, 2013). Thus, learning those words is important to achieving ‘basic functionality with a language’ (Schmitt, 2010, p. 75). Gardner and Davies (2014) use the term ‘core vocabulary’ to represent a set of words needed for interaction in L2 inside and outside the classroom but that are not specific to context to culture, and not necessarily high-frequency words. Gardner and Davies (2014) note that there is a set of core vocabulary that are not stable and there are function and content words that are stable for a long period. Nation (2001) uses the term ‘high-frequency’, although as noted by Miller (2012), Schmitt’s (2010) term ‘general vocabulary’ is probably more appropriate in light of the fact that frequency is just one key feature of those words.

The most well-known and widely used general English word list is West’s (1953) General Service List of English Words (GSL), which has been extensively employed in vocabulary research and instruction. The GSL has been a reference point for many practitioners in the ELT community for many years (Dang & Webb, 2016a). The GSL reported high lexical coverage, around 70%–90%, of various text types, such as academic writing (Coxhead, 2000), academic speech (Dang & Webb, 2014) and novels. Despite its indispensable role in vocabulary pedagogy and research, there have been several attempts to revise the GSL (see below) on the grounds that it was outdated (Richards, 1974; Schmitt, 2010), biased towards written English (Carter & McCarthy, 1988) and based on intuitive judgements rather than empirical evidence (Nation, 2016). While some of these attempts did indeed result in updated lists characterised by high coverage (e.g., Nation, 2012) or which addressed the needs of particular learners (e.g., Dang & Webb, 2016b), the GSL remained a useful pedagogical list (Schmitt, 2010).

Half a century after the development of the GSL, Nation (2006) developed the BNC 2000 word list from the 100 million-word BNC corpus, resulting in a list of 1,000 word families divided into 14 sub-lists. Frequency, range and dispersion were used for word selection, which was adjusted by the subjective addition of less frequent words perceived as important. The list reported 81%–97% coverage within various corpora—higher coverage than the GSL—yet it may not be the best candidate to replace the GSL. The BNC 2000 reflects adult, formal, written British English represented by the BNC, which may not be suitable for young L2 learners (Dang & Webb, 2016a).

To improve his BNC 2000, Nation (2012) built a large corpus of six million words from spoken sources and four million from written sources, incorporating sources as various as children’s books and TV programmes. He used this corpus to derive 1,000 word families and construct the BNC/COCA 2000 (2012). As for his previous list, Nation approached the BNC/COCA 2000 by subjectively adjusting corpus-driven information to determine inclusion. The list reported high lexical coverage (80.1%–81.7%) in the source corpora, and it is thought by some researchers (Dang & Webb, 2016a; Dang et al., 2020) to be a useful pedagogical word list and the best candidate to replace the GSL. Dang and Webb (2016a) attributed the potential usefulness of the BNC/COCA 2000 for language learning to the balanced corpus of spoken and written texts from which it was drawn and the careful subjective revision of the list.

On the other hand, Brezina and Gablasova (2015) felt that the best way to replace the GSL would be to employ a quantitative approach. Their New General Service List (NGSL1) consists of 2,494 lemmas derived from four corpora: the LOB, the BNC, the British English 2006 corpus (BE06) and the English TenTen web corpus (EnTenTen12), consisting of 12 billion running words in total. Unlike the previously mentioned lists, the NGSL1 was developed by a purely quantitative approach (i.e., frequency, dispersion and distribution across corpora), and the lemma, not the word family, was used as the counting unit. Dang and Webb (2016a) point out that the strength of the NGSL1 lies in its basis in an enormous corpus and that it is divided into core and current vocabulary. However, they

note that the purely quantitative approach may have led to the exclusion of low-frequency words that are useful for L2 learning, and finally, they note that it is biased towards written British English.

An additional New General Service List 2 (NGSL2) was developed by Browne et al. (2013) in another attempt to update the GSL. This list consists of 2,818 lemmas (2368 word families) derived from the 273 million-word subsection of the Cambridge English Corpus (CEC). The researchers reported that they employed a combination of objective (frequency) and subjective measures (excluding proper nouns, abbreviations, etc., as well as comparing their list with the original GSL, BNC and COCA to include/exclude words as necessary and consulting Paul Nation on how to improve the list), mirroring West's criteria, to develop this list (Browne, 2013). However, the researchers did not clearly report their discussion with Nation, nor did they report examples of the words excluded or included based on comparison with other lists. The NGSL2 reported 92% coverage of most general English texts, higher than the original GSL. Even so, it may not be the ideal candidate to replace the GSL. It reported lower coverage per item than did other lists of its kind, no justification for the decisions made during its development has been published, and finally, its large size is simply not a practical learning goal (Dang & Webb, 2016a). In 2017, Browne et al. released a revised version to represent spoken general English, the New General Service List-Spoken. It consists of 721 words and provides up to 90% coverage of unscripted spoken English (Browne et al., 2017).

Dang and Webb (2016a) evaluated four of the above general English lists (the GSL, BNC 2000, BNC/COCA 2000 and NGSL1; the NGSL2 was excluded from analysis for the previously mentioned reasons) and found that while Nation's (2012) BNC/COCA 2000 is likely the most suitable candidate to replace the GSL if lexical coverage is the only criterion, their own list may be more suitable for L2 beginners. By combining the four lists, the pair developed The Essential Word List for Beginners (EWL: 2016b). The EWL is an 800-lemma list based on 18 corpora (nine spoken and nine written) of different sizes, discourse types and English varieties. The characteristics of the target users were also

taken into consideration, and in terms of its size, the list is an achievable learning goal. However, no justification was reported for using of the flemma rather than the lemma.

Nation (2016) pointed out that it is difficult to identify the best general English list for all L2 learners. The problem with general English lists is that L2 learners begin at different levels, belong to different age groups and learn the L2 in different contexts. Thus, Nation claimed that identifying a general English list of more than 1,000 words that would be suitable for all L2 learners was an optimistic goal. However, at the same time, Nation claimed that 1,000 words does not constitute a sufficient size for a general English list. The precise amount of general vocabulary that is essential for L2 beginners has been the subject of much argument. It has been assumed that a general English list should include 2,000 words because that is the number of items in West's (1953) famous list (e.g., Nation, 2001; 2013); Schmitt and Schmitt (2014), on the other hand, contended that high-frequency English vocabulary should include 3,000 words, while Dang and Webb (2016b) argued that a list of 2,000–3,000 words was an impractical learning goal unless it was broken into sub-lists of hundreds of words or fewer; hence, they compiled a list of 1,000 words for L2 beginners. Deciding on the number of high-frequency words after 1,000 should be based on the educational setting in which the list will be used and the characteristics of the learners.

2.3.2 General Academic Word Lists

General academic word lists focus on academic vocabulary across a wide range of academic disciplines targeted at students of English for Academic Purposes (EAP) programmes or first-year students at English-medium universities where both instruction and assessment are in English (Nation, 2016).

A number of terms have been used to refer to 'general academic' vocabulary, including 'semi-technical vocabulary' (Farrell, 1990) and 'sub-technical vocabulary' (Yang, 1986). Gardner and Davies (2014) used 'core academic vocabulary' to reflect their approach to a common core of academic vocabulary that exists across different disciplines, and Nagy and Townsend (2012) used the

term ‘general academic vocabulary’ in lieu of discipline-specific general academic words. The current study uses ‘general academic vocabulary’ to refer to words found more often in academic texts than in general ones (covering about 9–10% of academic texts) but which are not specific to a single subject area (Schmitt, 2010). As there is no intrinsic property that makes a word academic (Schmitt & Schmitt, 2020), this type of vocabulary is commonly identified statistically in relation to their occurrence in academic texts. Various general academic word lists have been compiled to highlight those words, which require attention during the process of language learning and teaching since they are important to understanding and producing academic discourse, but unlike General English words, they are not particularly frequent and unlike specialised terms, they are not emphasised by content teachers (Schmitt, 2010).

The first corpus-based—and most influential—general academic word list is Coxhead’s (2000) Academic Word List (AWL). The list has been influential for language teaching and learning as well as for research. The list has revolutionised EAP learning and has been cited by a great quantity of published research (see Section 2.3.4 for information about the development of the AWL).

The AWL was criticised for being based on a previous list, the GSL, and that the unit of word families was used to group words in the list. To address these problems, Gardner and Davies (2014) developed the Academic Vocabulary List (AVL). The AVL is a list of 3,015 lemmas derived from the 120 million-token academic sub-corpus of the COCA, further divided into nine disciplines (education, humanities, history, social sciences, philosophy/religion/psychology, law/political science, science and technology, medicine/health, and business/finance). The AVL covers about 14% of the academic sub-corpora in the COCA and the BNC (Gardner & Davies, 2014). Unlike the AWL, the AVL was not based on a previous list. The principle of ratio was used to eliminate non-academic words, while range, dispersion and discipline measure (the word could not occur more than three times the expected frequency in any of the nine disciplines) were used to eliminate specialised or technical terms in the academic corpus. The corpus-comparison approach was used to identify general

academic words. Under this metric, a word is academic if it is 50% more frequent in the academic corpus than in a general corpus (Gardner & Davies, 2014). A word-family version containing 1,991 word families was developed for the purpose of comparing the top 570 AVL families with the 570 AWL word families (Gardner & Davies, 2014).

Another addition to general academic word lists is the New Academic Word List (NAWL) of Browne et al. (2013). The list was developed as an update to Coxhead's AWL; while the AWL was based on West's GSL, the NAWL was designed to fit with an updated GSL, the NGSL2, developed by the same researchers. The list consists of 963 words drawn from a corpus of 283 million words, which were collected from academic texts largely in the Cambridge English Corpus of academic journals, non-fiction, student essays and academic discourse; the Michigan Corpus of Academic Spoken English (MICASE) and the British Academic Spoken English (BASE) corpora; and an academic textbook corpus (Browne et al., 2013). The NAWL, in combination with the NGSL2, provided about 5% more coverage of most academic texts than the AWL and the original GSL (Browne, 2016). The list is available in several formats: headwords, lemmas, and also English and Japanese (Coxhead, 2018). However, neither word-selection criteria nor a rationale for the unit of counting were discussed. This might be attributed to the fact that Browne et al.'s self-reported focus was on creating pedagogical resources that would serve L2 teacher and learners, rather than on publishing academic papers for research purposes (Browne, 2016).

With a focus on spoken academic discourse, Dang et al. (2017) developed The Academic Spoken Word List (ASWL) to help L2 learners understand academic speech in English-medium universities, regardless of their discipline and proficiency level. The ASWL consists of 1,741 word families chosen according to their frequency, range and dispersion in an academic spoken corpus. The corpus comprises 13 million words divided into four equally sized disciplinary sub-corpora. Nation's BNC/COCA 2000 list was used as a general service word list to divide the ASWL into these

four levels. Dang et al. (2017) claim that L2 learners who master the ASWL may reach 92–96% coverage of academic speech.

While the developers of such lists endorse a generic approach to academic word lists, other researchers (e.g., Durrant, 2014, 2016; Hyland & Tse, 2007) question whether such lists can ever be equally useful to all L2 students, irrespective of discipline. Having analysed the distribution of the 570 word families of the AWL in an academic corpus of different disciplines and genres, Hyland and Tse (2007) concluded that academic vocabulary is not equally distributed across different disciplines, and even if it were, it would vary in meaning and collocation. Similar findings were reported regarding the AWL's behaviour in different subject-specific corpora (e.g., Martínez et al., 2009). Furthermore, analysing students' academic writings, Durrant (2014, 2016) found that academic vocabulary use varied across different disciplines.

Despite this criticism, Durrant (2016) concluded that general academic word lists would still be valuable because some disciplines are similar and that a relatively small general academic vocabulary can be identified. A list of this sort would be valuable to first-year students, who are often undecided as to their discipline and may be required to write papers in general academic courses (Dang et al., 2017). Moreover, universities today are more inter-disciplinary, and hence, a shared general academic vocabulary might be of value (de Chazal, 2013).

2.3.3 Specialised Word Lists

In response to criticism of the generic approach to academic word lists and a call for learning vocabularies related to specific subject areas (Hyland & Tse, 2007), several specialised word lists have been developed. For example, Vongpumivitch et al. (2009) compiled a list of specialised terms in applied linguistics, Valipouri and Nassaji (2013) in chemistry, Martinez and Schmitt (2012) in agriculture, Yang (2015) in nursing and Lei and Liu (2016) in medicine.

These lists include "words and phrases that are closely related to the ideas covered in a particular subject area" (Nation et al., 2016, p. 146). Specialised vocabulary could include terms that

occur exclusively in a certain discipline and have a very specific meaning, such as *morpheme* in applied linguistics; terms occurring across various contexts and having different specialised meanings (*token*) or a specialised meaning intelligible from the common meaning (*range*); and common terms that are frequent in the discipline (*word*) but which would have precise meaning to a specialist (Nation, 2001).

Specialised vocabulary has been known by different terms in the field, including ‘technical vocabulary’ (Chung & Nation, 2004; Hyland & Tse, 2007; Nation, 2013, 2016; Schmitt, 2010; Webb & Nation, 2017) and ‘discipline-specific’ vocabulary (Nagy & Townsend, 2012); some researchers (e.g., Coxhead, 2018; Nation, Coxhead, et al., 2016) use the term ‘specialised vocabulary’ to refer to both general academic (Tier Two) and technical (Tier Three) vocabulary. English for Specialised Purposes (ESP) is viewed as an umbrella term for all specialised English purposes, including EAP, professional and occupational English, and English in the trades (Coxhead, 2018). In the current study, ‘specialised vocabulary’ refers to words and phrases used in a particular subject area reflecting Tier Three (Beck et al., 2013), occurring within and outside of academia, (i.e., also covering professional and occupational English and English in the trades).

The primary characteristics of vocabulary of this type could be said to be infrequency and narrow specialised meanings (Ha & Hyland, 2017). However, identifying specialised vocabulary has been problematic (Chung and Nation, 2004), as it includes both high-, mid- and low-frequency vocabulary (Nation, 2013, 2016). Coxhead (2018) reviews the most common approaches to developing specialised word lists, one is to consult an expert in the field, which is a time-consuming and difficult process, not to mention heavily influenced by what terms the expert considers technical or special. Referring to technical dictionaries is another way of identifying technical vocabulary, but it is influenced by the quality of the dictionary and the decisions made as to its contents. The corpus-comparison method is the most efficient means of identifying specialised vocabulary.

Examples of specialised word lists include a list developed by Vongpumivitch et al. (2009), which contained 128 specialised terms in applied linguistics, drawn from a 1.5 million-word corpus of research articles. Martínez et al. (2009) compiled a list of 92 word families from the AWL based on AgroCorpus, a corpus specialising in agriculture (AgroCorpus consists of 800,000 words from 218 agriculture research articles). Valipouri and Nassaji (2013) developed the Chemistry Academic Word List (CAWL), containing 1,400 word families derived from a 4 million-word chemistry research articles (1185 articles) corpus to help EFL chemistry students. Yang (2015) developed a Nursing Academic Word List (NAWL) of 676 word families drawn from a 1,006,934-word corpus of nursing research articles and found 475 AWL and 128 non-AWL word forms of word families. It is important to note that these studies focused on the coverage of the AWL in the researcher's discipline-specific corpora or excluded the GSL and AWL from their analysis. Most of these lists are derived from journal article corpora, which may not be suitable reference corpora for students who are mostly exposed to textbooks. Also, word families are used as the counting unit is used in most of the published lists despite that it may not be the best unit for specialised vocabulary.

Recently, Lei and Liu (2016) developed the Medical Academic Vocabulary List (MAVL), employing a refined set of criteria that synthesised the AWL and AVL criteria (minimum frequency, ratio frequency, range, dispersion, discipline measure and special-meaning criteria for general high-frequency and corpus comparison). The MAVL consists of 891 lemmas derived from a corpus comprising 2.7 million running words found in medical academic journal articles and 3.5 million words found in medical textbooks. MAVL lemmas were selected to help medical students, as well as non-native professionals and researchers, read research articles.

2.3.4 Multi-word Unit Word Lists

Through corpus analysis, researchers were able to look into not only the frequency of word occurrences but also at the most frequent combinations of words referred to as multi-word units. It was found that these combinations make up a large portion of English texts, approximately one third

and one half of discourse (Conklin & Schmitt, 2012). The functions and importance of multi-word units in the process of L2 learning have been long studied and emphasised (e.g., Biber et al., 2004; Hyland & Tse, 2007; Sinclair, 1991). Knowledge of these sequences have been linked to successful communication and native-like fluency (Pawley & Syder, 1983).

This type of vocabulary is known by various terms, including ‘collocations’, ‘lexical bundles’, ‘multi-word constructions’, ‘formulaic language’ and ‘word combinations’ (as reported in Wray, 2009). The lack of a consistent terminology and definition reflects the difficulty of developing a clear operational definition of what counts as a multi-word unit. This is because of the different functions and categories of these units; and the different methods of identifying them (Nation et al., 2016). The term multi-word unit is used in the current study to refer to any combination of words of any length for any function whether adjacent or not.

A recent trend in building pedagogical word lists has been developing multi-word lists. Multi-word lists are limited in number relative to single-word lists, but two recent examples are worth the mention; the Phrasal Expressions List or the PHRASE List of Martinez and Schmitt (2012) and the Academic Formulas List (AFL) of Simpson-Vlach and Ellis (2010). The former is a general phrase list, while the latter is restricted to academic language. To date, however, discipline-specific pedagogical phrase lists are still limited; examples include a multi-word unit list in pharmacology (Grabowski, 2015), law (Breeze, 2013), EU documents (Jablonkai, 2010), and applied linguistics (Qin, 2014).

The PHRASE List (Martinez & Schmitt, 2012), consists of 505 multi-word units (ranging from two to four words in length) derived from the 100 million BNC corpus. The researchers relied on frequency and a qualitative criterion (expression that may cause decoding problems if read word-for-word) to determine which repeated combinations of words would improve L2 learners' receptive comprehension. Interestingly, almost all of the sequences on the Phrase List are made up of the 2,000 most frequent single-words of the BNC (Martinez & Schmitt, 2012).

The Academic Formulas List (AFL) includes 200 multi-word units in academic spoken discourse, 200 in written academic discourse, and 200 common to both registers (Simpson-Vlach & Ellis, 2010). The researchers identified three-, four-, and five-word sequences from a two corpora each totalling 2.1 million words that represent academic speaking and writing. To identify sequences for the AFL, the researchers used a combination of measurements for frequency, range, and strength of association between the words in a formula. Finally, the researchers asked an independent group of English teachers to rate the value of each sequence.

This section reviewed various kinds of published corpus-based word lists. It is clear that there are different types of vocabulary, different functions and characteristics of word lists and a number of ways to organise them. Thus, it could be said that as word lists can play a part in structuring and guiding pedagogy, the right list needs to be used. To be able to select and use them effectively (the purpose of this study), practitioners need the right list and to make such a judgment they need to understand how word lists are made (Schmitt & Schmitt, 2020). In the next section, the most common criteria and issues in word list construction—units of counting, corpus construction and selection criteria—are discussed.

2.4 Issues in Corpus-based Word List Construction

Approaches to creating word lists employed by the Vocabulary Control Movement in the early 20th century are still employed: the intuitive approach (e.g., Ogden, 1930), the corpus-driven approach (e.g., Thorndike, 1921), and the mixed method approach (e.g., West, 1953). The previous review shows that most word lists today are derived from corpus analysis either exclusively or by adjusting the corpus analysis data subjectively. It also shows that most word lists have been organised around word families, and that different types of corpora have been used. These choices depend on the purpose, and other factors that are discussed next.

2.4.1 Counting Words

Defining what counts as a 'word' is a complicated decision that affects the quality, size and usefulness of a word list (Gardner, 2007; Nation, 2016). The conceptualisation of the 'word' construct can range from the treatment of every sequence of letters separated by space or punctuation as a distinct word to the treatment of all words of the same stem, regardless of additional morphemes, as members of one word family. The word family **type** recognises each form as a different word. For example, *predict* is counted upon its first appearance in a corpus as one type, and *predicts* is counted as another type. The type has been the unit of counting for a few mostly specialised word lists such as the Applied Linguistics Academic Word List (Vongpumivitch et al., 2009) and the OPEC (Organisation of Petroleum Exporting Countries) Word List (Aluthman, 2017). The word family type is placed on Level 1 of Bauer and Nation's (1993) word family levels. These levels were created to systematically approach vocabulary in language pedagogy and research. They are ordered according to the frequency of the affix, productivity of the affix—how likely it will be used to form new words—, predictability of the meaning of the affix and regularity of English affixes—the amount of changes to the base when the affix is added—, divided into seven levels (see Table 2.2 for a summary of affixes included at each level).

Table 2.2*Bauer and Nation's (1993) Word Family Levels*

Word family level	Affixes
Level 1	A different form is a different word
Level 2	Inflectional suffixes: plural, third person singular present tense, past tense, past participle, <i>-ing</i> , comparative, superlative and possessive (8 affixes)
Level 3	Most frequent and regular derivational affixes: <i>-able, -er, -ish, -less, -ly, -ness, -th, -y, non-, un-</i> (all with restricted use) (10 affixes)
Level 4	Frequent, orthographically regular affixes: <i>-al, -ation, -ess, -ful, -ism, -ist, -ity, -ize, -ment, -ous, in-</i> (all with restricted use) (11 affixes)
Level 5	Regular but infrequent affixes: <i>-age, -al, -ally, -an, -ance, -ant, -ary, atory, -dom, -eer, -en, -en, -ence, -ent, -ery, -ese, -esque, -ette, -hood, -i, -ian, -ite, -let, -ling, -ly, -most, -ory, -ship, -ward, -ways, -wise, ante-, anti-, arch-, bi-, circum-, counter-, en-, ex-, fore-, hyper-, inter-, mid-, mis-, neo-, post-, pro-, semi-, sub-, un-</i> (50 affixes)
Level 6	Frequent but irregular affixes: <i>-able, -ee, -ic, -ify, -ion, -ist, -ition, -ive (ative), -th, -y, pre-, re-</i> (12 affixes)
Level 7	Classical roots and affixes

Note. Adapted from *Making and Using Word Lists for Language Learning and Testing* (p. 27) by I. S. P.

Nation, 2016, John Benjamins.

The **lemma** (e.g., *PREDICT*) consists of a stem *predict*, together with its inflected forms of the same part of speech: *predicts, predicted* and *predicting* (Dang et al., 2017). The lemma has been chosen as the unit of counting for some recent word lists (Brezina & Gablasova, 2015; Gardner & Davies, 2014; Lei & Liu, 2016). The choice to employ the lemma is based on the assumption that if learners know the base form of a word, they may recognise its inflected forms, thus decreasing the learning burden (Nation, 2016).

A pure lemma includes only inflected members of the same part of speech. This led Pinchbeck (2014) to use a more inclusive version of the lemma, the '**flemma**' (family lemma), which adds different parts of speech to the original lemma. For example, the noun *walks* and the verb *walks* belong to the same flemma (Nation, 2016). Bauer and Nation (1993) originally placed lemmas at Level 2 of the word family levels (see Table 2.2), but these levels could be sub-divided to situate the flemma at Level 2.5, leaning towards Level 3 (Nation, 2016). The flemma was chosen as the unit of

counting in designing the EWL for L2 beginners (Dang & Webb, 2016b) and a version of the AVL (Gardner & Davies, 2014).

The Level 6 word family, called the **word family**, includes closely related derived forms of a stem alongside its inflectional forms. For example, the word family predict includes *predicts*, *predicted*, *predicting*, *prediction*, *predictor*, *predictability*, *unpredictability*, *predictably*, *predictable*, *unpredictable* and *predictive*. The Level 6 word family of Bauer and Nation's (1993) levels (see Table 2.2) reflects this conceptualisation, incorporating up to 50 derived and inflected forms (Dang et al., 2017; Nation, 2013). Many of the influential word lists have been developed using the Level 6 word family as the counting unit (e.g., Coxhead, 2000; West, 1953; Xue & Nation, 1984), as well as some specialised word lists (e.g., Yang, 2015). Grouping words into this unit follows from the assumption that if a learner knows the stem and has some knowledge of morphology, the inflected and derived forms can be inferred (Nation, 2016).

A much-debated question has been whether counting types, lemmas or word families (Level 1, Level 2 and Level 6, respectively) is best, but the debate has recently shifted to asking which level is most suitable for a list's potential uses and users (Gardner, 2007). To address this question, three elements must be considered: (a) the purpose of using the list (in terms of productive or receptive vocabulary development), (b) the target vocabulary type and (c) the target learners and context (Dang, 2020).

Lists based on word families can be suitable for receptive vocabulary development (i.e., listening and reading comprehension), according to the assumption that the meanings of inflected forms (e.g., *predictable*) can be guessed if the meaning of the base (in this case *predict*) is known, thus lowering the learning burden (Nation, 2016). Word family groupings have been widely used for word list development and evaluation; for developing tests, course designs, and graded readers; and in corpus analysis programs such as Range (Heatley et al., 2002) and AntWord Profiler (Anthony, 2014). The widespread use of the word family has been criticised (e.g., Gardner, 2007; Schmitt &

Zimmerman, 2002). This grouping assumes that learners have mastered English morphology, which is commonly acquired later in the learning phase, and thus may not be suitable for beginners (McLean, 2018). Furthermore, learners from certain first-language backgrounds may have difficulty acquiring affix knowledge as it does not exist in their L1 (Ward, 2009).

Gardner and Davies (2014) criticised the use of word families in creating the AWL for the following reasons:

1. Not all members of a word family are closely related in meaning; for example, *react* (respond), *reactionary* (strongly opposed to social or political change), *reactivation* (the making of something to happen again) and *reactor* (a device or apparatus) have different meanings despite belonging to the same word family.
2. Word families do not distinguish between parts of speech.
3. Word family-based lists require more complex morphological knowledge that is usually acquired at a later stage of L2 learning.

Nation (2016) refuted these arguments by pointing out that to be included in a word family, according to Bauer and Nation's (1993) scale, the meaning of the base in the derived word must be closely related to the meaning of the base when it stands alone or is combined with other derived forms. However, making such an assessment requires subjective judgement, which may lead to inconsistent results. There has been a growing body of evidence that for many L2 learners, the word family may not be appropriate for purposes of pedagogy and assessment (Gardner, 2007; McLean, 2018; Mochizuki & Aizawa, 2000; Schmitt & Meara, 1997; Ward & Chuenjundaeng, 2009). Notably, McLean (2018) found that most advanced learners had limited knowledge of derived forms.

In light of the difficulties associated with using the word family, the debate has shifted towards using the lemma as an alternative. Lemma-based lists are said to be suitable for (a) productive vocabulary development and (b) assessment purposes aimed at beginner-to-intermediate learners (Gardner & Davies, 2014; Kremmel, 2016; McLean, 2018). It is assumed that those learners

have not mastered the derivations of English but would usually be aware of the inflections of a known stem. On the other hand, it has been argued that even though lemmas distinguish parts of speech of words of the same form, lemmas, like word families, do not effectively distinguish homoforms, words that have the same form but different meanings (Nation, 2016). In fact, distinguishing parts of speech increases the amount of effort needed to acquire a word, which directly contradicts the aim of choosing the lemma in the first place. Additional problems with lemmas are whether irregular forms such as *mice* (plural of *mouse*) ought to be included as members of the lemma and whether the base form or the most frequent form should be the headword (Nation, 2013). Finally, lemma-based lists may be restrictive, as most English derivations are not difficult. Word family-based lists, on the other hand, are shorter and devoid of word repetition. An additional advantage of word families for counting is ease of use; most people are familiar with headwords used in dictionaries (Schmitt, 2010).

Gardner and Davies (2014), Kremmel (2016) and McLean (2018) argue for the use of the lemma or flemma over the word family. However, further investigation is required as to the suitability of flemma- over lemma-based lists for intermediately proficient learners. The use of the flemma as the unit of counting assumes that learners can make ‘conceptual links between the use of flemma constituents to express, for example, a nominal entity, a verbal process, or an adjectival attribute’ (Stoeckel et al., 2018, pp. 1-2), thus lowering the learning burden (Dang & Webb, 2016b). The number of word lists based on the flemma is still limited, for example, the NAWL (Browne et al., 2013), the NGSL2 (Browne, 2013) and EWL (Dang & Webb, 2016b) are available in the flemma format, yet the superiority of the flemma to the lemma is uncertain.

For productive vocabulary development (i.e., writing or speaking), the type or the lemma are claimed to be more suitable than the word family because knowledge of the appropriate form of the word is needed to produce that word. Schmitt and Zimmerman (2002) found that knowledge of one form in a word family does not necessarily imply that other forms in that family can be successfully

produced. Type-based word lists are also suitable for technical vocabulary learning, in which one form of a word may be technical (such as *patient* in the nursing discipline) while related forms (e.g., *patience*) might not be (Coxhead, 2018). They are also suitable for spelling lists, where they help to ensure that the correct form is learned (Nation, 2016). However, type-based lists are usually quite large. Dang and Webb (2016b) showed that type-based word lists are six times larger than word family-based lists, which may limit their practicality.

Another challenging issue in word counting for word lists is related to variations of meaning among words with the same or similar forms. For instance, a given sequence of letters counted as a word can have multiple related meanings (e.g., *break* as in *break a leg* vs *break a cup*, termed polysemy) or unrelated meanings (e.g., *bank*, referring to the side of the river or to the financial institution, termed homonymy) treated as two different words (Gardner, 2007).

2.4.2 Corpora and Word Lists

The historical overview of word lists shows that analysing corpora to investigate the occurrence and behaviour of words in a text has been a well-established practice for developing word lists based on empirical evidence, rather than relying on a native speaker's intuition (Szudarski, 2018). Therefore, the nature of the list would depend on the texts from which the words were selected—that is, the corpus (Nation & Sorell, 2016).

Even though corpus analysis has revolutionised word list development, it does not necessarily mean that corpus-based word lists are superior to intuitively designed ones (Nation, 2016). For this to be true, the source corpora must be well-designed. A word list can be said to be good if it is based on a corpus that accurately represents the actual or potential language that the target audience will likely encounter. However, it is simply not possible to collect *all* instances of the target language (e.g., academic English). Consequently, we must rely on a sample that represents, as closely as possible, the language used by the target population. For example, if a list is intended to be

used for young children, then the list should be based on a corpus of texts that those children are likely to encounter.

When collected for lexical investigations, it has been strongly recommended that corpora are large (e.g., Sinclair, 1991), but the size of the corpus depends very much on the purpose of the resulting list, as well as other practical considerations. Small specialised corpora, such as the 826,416-word corpus from which the Agricultural Word List (AgroCorpus List) was derived by Martínez et al. (2009), have been compiled for specific research or pedagogical purposes. Thus, the purpose of corpus design could be seen as an answer to the question of how large a corpus should be. In fact, the more domain-specific the research interest is, the smaller the corpus can be (Hunston, 2002; Weisser, 2016). Another example of target corpus size arises when developing a list based on the teaching materials of a particular teaching context. In this case, a small pedagogic corpus of less than one million words, built on the teaching material, would result in a reliable and useful list (Coxhead, 2018).

Other practical issues that may constrain the size of the corpus include the following:

- Some commercial corpus software sets size limits (Nation, 2016).
- Spoken corpora tend to be smaller than written corpora because it is time-consuming to collect and transcribe spoken texts. According to O'Keeffe et al. (2007), corpora of over one million spoken words are considered large, whereas corpora of fewer than five million words of written text are quite small.
- The use of published texts may be limited by the permissions one can obtain from the copyright holder (Nation, 2016).

2.4.3 Approaches to Making Word Lists

Two word lists, the GSL and the AWL, stand out as having the most influential methodological approaches to word list development (Palinkašević, 2017). To create the AWL, Coxhead relied on

statistical measures—namely frequency and range—for word selection and ranking, while West used frequency, but adjusted the results subjectively, to ensure that his GSL was efficient (suitable to the target learners and purpose).

Since the 1920s—and indeed represented by the three approaches of the Vocabulary Control Movement of that era—the argument has been whether the objective corpus-driven information approach, the subjective approach or the mixed-approaches (subjectively-adjusted corpus-driven information) is best. Instead of developing new lists, some researchers and ELT practitioners prefer to combine more than one ready-made published word list to come up with one that is suited to their use. Each approach has its advantages and disadvantages, but the criteria used for list development must be evaluated against the list's purpose and context. The following paragraphs present a discussions of the most common approaches to compiling corpus-based word lists in ELT.

The objective corpus-driven approach exclusively uses one or more statistical criteria to identify and order words into a list. The common statistical criteria are **frequency**, **range** and **dispersion**.

Frequency, the most widely used statistic in corpus analysis, refers to the number of times a word occurs across an entire corpus (Gries, 2008). For a long time, word importance has been biased towards frequency, pursuant to the assumption that high-frequency words are more likely to be encountered by students and are thus more important. Frequency has also been used to categorise words into high-, mid- and low-frequency groups (Szudarski, 2018). Despite being a widely used measure, frequency has been criticised for its inability to extract low-frequency words that have high information content (Richards, 1970), and for failing to adequately reflect word importance if used as the sole criterion (Egbert, 2018). Also, some words may appear many times in a small number of text, confounding the measure. For example, a corpus of 100 texts might contain 1,000 occurrences of word *a* and only 900 occurrences of word *b*. However, if word *a* appears only in five texts, most learners are less likely to encounter the word, and therefore learn the word. If word *b*, in contrast,

appears in half of the corpus texts, then it is much more likely that the learners will encounter the word and learn it (Hashimoto & Egbert, 2019). In such a case, the use of range and frequency counts is important to avoid this issue (Dang et al., 2017). **Range** refers to the number of different corpora or sub-corpora in which a word occurs. Nation (2016) suggests that range should be the first criterion for selection because items that are truly useful will be more important to learn. However, range is affected by the size of the sub-corpora; low-frequency items are more likely to occur in large sub-corpora than in smaller sub-corpora. Moreover, range does not ensure even distributions within multiple corpora or sub-corpora (Dang & Webb, 2016a). For example, when developing the ASWL (Dang et al., 2017), the word *predator* occurred 47.38 times per million in the Management sub-corpus but fewer than eight times in the other sub-corpora. **Dispersion**, ‘a measure of the word’s distribution or spread across the parts of a corpus’ (Burch et al., 2017), is another measure of the likelihood that learners have been exposed to a word. However, dispersion’s relationship with word difficulty has yet to be researched. Dispersion combines frequency and range and examines how evenly a word is distributed across a corpus (Dang & Webb, 2016a). These three measures are widely used in combination with one another, as in the AVL (Gardner & Davies, 2014).

The major advantage of the objective corpus-driven or the statistical approach is that it is a replicable procedure. This makes it easier to sequence lists, and it allows for comparison between lists that use different corpora or criteria (Coxhead, 2018; Nation, 2016). Word lists based on statistical measures are claimed to be more reliable and accurate than those based on intuition (Szudarski, 2018). These measures, however, do not provide information about variation in the meanings or functions of a word across discourse types (Dang & Webb, 2016a). Finally, the corpus from which words are drawn greatly affects the resulting word list (Nation, 2016). Nation notes that a large, truly representative corpus is often impossible and that most corpus software tools are still limited. Thus, a statistics-based word list might be useful for a particular group of learners but not for others.

Accordingly, Nation (2016), like others, claims that it might be necessary to combine subjective criteria with statistical measures to produce useful pedagogical word lists that account for the target audience, context and purpose. In fact, many corpus linguists (e.g., O’Keeffe et al., 2007) highlight the equal importance of approaches that are both quantitative and qualitative to corpus analysis (Szudarski, 2018) especially if the purpose is to construct a word list that is useful to learners. The GSL is the most common example of subjectively adjusting statistical results in an attempt to make the list as efficient and complete as possible for the target users. Despite its age, the GSL still achieves high lexical coverage in various corpora (Dang & Webb, 2016a). Nation’s BNC/COCA 2000 lists are another example of the mixed-criteria approach.

A combination of subjective data (e.g., teacher perception of usefulness) and corpus-driven data has been used in the development and validation of academic vocabulary lists (e.g., He & Godfroid, 2019; Simpson-Vlach & Ellis, 2010). Evaluating two high-frequency lists based on lexical coverage and asking language teachers to determine their usefulness proved that pedagogical lists, tailored subjectively, are more useful (Dang et al., 2020). Those lists are claimed to be more pedagogically useful because they take into account the teaching context, type of vocabulary and purpose of use. Finally, it has been argued that even the statistical approach is not purely objective since corpus selection and development involve subjective decisions (Sorell, 2013).

The purely qualitative approach authorises an expert native speaker to judge a word’s importance (Brezina & Gablasova, 2017). Nevertheless, corpora could also be used qualitatively to inform word list development (Coxhead, 2018). Word selection could be made by analysing students’ annotations in their textbooks and reading materials, proceeding from the assumption that students annotate words that they perceive as important or problematic (e.g., Ghadessy, 1979). Selection could also be made by consulting concordance lines to examine meaning and collocation (e.g., Hafner & Candlin, 2007) or by using surveys, interviews and questionnaires to gather information about the importance of words from the perspective of teachers (e.g., Coxhead, 2012). For examples of lists

following the qualitative approach, the reader is referred to Coxhead (2018). The number of published word lists following the expert-consultation approach is limited, although it is important to note that lists developed by teachers who draw on their experience and teaching materials are included under this approach. Though those methods account for practitioners' experience, they may lead to unreliable judgments and inconsistent results influenced by teachers' years of experience, students' proficiency levels or simple differences in personal preference (Schmitt, 2010).

It is important to note that the identification of different types of vocabulary may require different criteria. For example, specialised word lists may benefit from unique metrics such as the judgement of a specialist. Most important is that the criteria are formalised and consistently applied. It is also important to recall that while advances in corpus software tools have led to a majority of word lists being corpus-based (Coxhead, 2018), corpora need not be involved in word list development. Some lists have been developed by combining several previous lists, benefiting from their strengths but also absorbing their weaknesses, as in the cases of The University Word List (Xue & Nation, 1984) and, more recently, the EWL for L2 beginners (Dang & Webb, 2016b).

This section, 2.4., has summarised the main issues and methods of compiling corpus-based word lists. The main conclusion is that there is no perfect word list and no perfect criteria. Decisions related to what to count as a word, compiling the corpus and the criteria of selecting words depend on the purpose of the list, type of vocabulary and the target audience. These elements should guide decisions related to word lists. Furthermore, Schmitt (2016) argues that those decisions must be justified by the list compilers. The following section discusses applications of corpus-based word lists to language teaching.

2.5 Corpus-based Word Lists for Language Teaching

The Vocabulary Control Movement in the early twentieth century was a starting point in the history of word lists which, in the course of time, led to the creation of manually compiled word lists that have had considerable influence on vocabulary research and instruction (see Section 2.2.2.

above). The GSL (West, 1953), which was a product of this movement, is a famous example.

However, as the natural or communicative approach became dominant in the 1970s, controlled vocabulary instruction in the ELT syllabus was disregarded in favour of incidental vocabulary learning and interest in developing learners' communicative skills; consequently, word lists use received little attention (O'Dell, 1997). Using and making word lists for vocabulary acquisition was neglected for many years because the trend was to emphasise learning vocabulary incidentally from the context (González-Fernández & Schmitt, 2017).

Computerised corpora became available in the 1990s, and provided useful insights into English vocabulary patterns and usage. They have had a tremendous influence on reviving interest in vocabulary instruction, and have allowed for the compilation of better word frequency lists that have resulted in more confidence when deciding which vocabulary to include in a syllabus. Nevertheless, word lists have been subject to many myths and misconceptions (Folse, 2004; Sorell, 2013). In the minds of both learners and teachers, word lists have often been associated with older teaching methodologies, such as the grammar translation school, and for some, word lists trigger memories of unsuccessful experiences with memorising lists of words (Sorell, 2013). Section 2.5.3.2. explores these negative perceptions.

It is worth mentioning that a group of researchers (Folse, 2004; Meara, 1980; Nation, 2001; Schmitt, 2000) promoted explicit vocabulary learning and thus stressed the role of word lists in intentional vocabulary learning. However, word list development has only truly flourished in the past 20 years as large corpora have become accessible on a much broader scale, and more sophisticated corpus-analysis tools have been developed and made available (Vilkaitė-Lozdienė & Schmitt, 2020). Although publications presenting and discussing word lists have begun to increase, how word lists are used and viewed remains unclear. This section summarises the arguments in favour of word lists for language teaching. It highlights the vital role corpus-based word lists have played in informing language teaching and learning. In light of teacher cognition, it discusses the reasons that word lists

have been neglected, namely a lack of awareness of word lists, negative perceptions of them and the misuses and the challenges associated with them. This section can be seen as an attempt to dispel the myths about word lists.

2.5.1 *The Value of Word Lists in Language Teaching*

Language learners must master a large number of words to become successful language users, and one way to manage this task is to use word lists (See Chapter 1 for research on L2 learners' lexical needs). The most obvious purpose of a word list is to direct the acquisition of vocabulary in a principled manner by prioritising words that merit instructional focus based on their frequency of use (Nation, 2016). This approach assumes that the more frequent a word is the more important it is because it is more likely L2 learners need these words across different contexts. The aim is to ensure that the efforts made are rewarded (Nation, 2016). This assumption is called the cost-benefit principle where the cost is the effort on learning vocabulary and the benefit is using or understanding the learned vocabulary when it is encountered (Nation, 2013).

Nation and Webb (2011) discuss the value of word lists beyond selecting words for teaching. In his book devoted to word lists, Nation (2016) provides an overview of their pedagogical uses. These are summarised into the following points:

- Word lists provide course and programme designers with data about what words should be learned and at what stage of proficiency that learning should take place. They provide the basis for setting short- and long-term learning goals (Nation & Webb, 2011).
- Word lists can be used to analyse the vocabulary load of a text to ensure that it includes the target vocabulary and that it reaches the target proficiency level of the learners. For example, the Range program (Heatley et al., 2002) with either the GSL/AWL lists or with the BNC lists can be used to analyse the vocabulary load of a text to tell how much and what vocabulary occurs in it. Such information can then be used to pre-teach those

words, provide a glossary of unknown words with the assignment and remove or simplify them.

- Word lists guide the creation of graded readers (simplified texts) to ensure that the words introduced correspond with the current target level and the preceding level. Nation (2016) notes that publishers make and use their own word lists but do not share them. This makes organising graded reading and evaluating word lists difficult.
- Word lists are used to inform the development of language-learning materials. A famous example of using word lists for material development is Coxhead's (2000) AWL, which has been the basis of many books, such as *Inside Reading 1 Student Book Pack: The Academic Word List in Context*, by Burgmeier and Zimmerman (2007).
- More recently, word lists have inspired software designers to create programs to help L2 learners develop their lexical knowledge (Wadden et al., 2016).
- Word lists are excellent resources for creating vocabulary-focused activities for classroom and independent-study use. The next section contains a detailed explanation of Nation's (2007) balanced 'four strands' approach to vocabulary instruction and how word lists can be used.
- Word lists provide a basis for constructing tests to inform teaching and learning: this includes testing the size of vocabulary or proficiency. One example is the Vocabulary Levels Test by Schmitt et al. (2001), which was based on Coxhead's (2000) AWL (Chen, 2011; Nation, 2016).

Schmitt and Schmitt (2020) added an element of practicality in using word lists. Word lists provide ELT practitioners (even experienced ones) with a resource to consult based on scientific work. This saves their time and effort, and it reduces human error and inconsistency when making choices.

2.5.2 *Using Word Lists for Vocabulary Teaching and Learning*

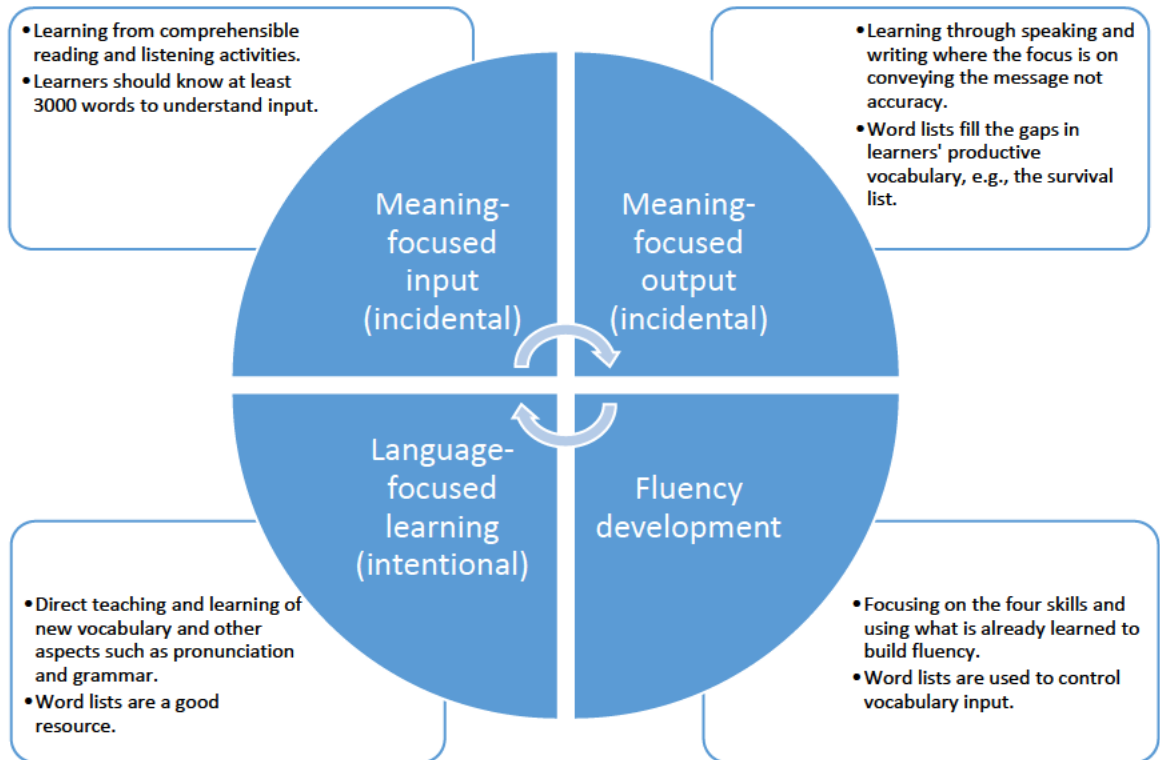
The previous section reviewed the reasons for using word lists as reported in the literature. It is important to note that presenting students with a list of target vocabulary items and asking them to learn these words was not mentioned in the previous section because teachers should not simply hand students a word list. The content of the list is the target that should be used by the teacher (or decision maker/practitioner) to guide the process of vocabulary acquisition (Folse, 2004). This section reports on the limited explorations of the extent to which word lists are used for vocabulary acquisition.

Acquiring vocabulary is a complex process that involves ‘learning lexical items (i.e. single words and formulaic language) in sufficient depth to use them both productively and receptively, by means of multiple incidental and intentional encounters with these items in varied contexts’ (González-Fernández & Schmitt, 2017, p. 280). Nation (2007) suggested the principle of the four strands, a balanced approach to vocabulary instruction. This principle states that a good language learning program (including courses to learn vocabulary) must provide balanced opportunities for meaning-focused input, meaning-focused output, language-focused learning and fluency development. Meaning-focused input involves learning through reading and listening; meaning-focused output involves writing and speaking; language-focused learning involves intentional vocabulary learning and fluency development focuses on acquiring and processing what was learned easily and quickly (Nation, 2007, 2013). Nation and Yamamoto (2012) described how the four strands principle was applied to autonomous learning of Spanish by the second author. They concluded that there is not a right way to learn a language and that there must be a balance of learning activities of the four strands. Researchers have compared the strands (e.g. intentional versus incidental learning), and they have compared activities within the strands (e.g. looking up words in a paper versus electronic dictionary). However, the balance between the four strands in principle has not been

explored, to the researcher's best knowledge. For vocabulary instruction, Nation (2016) showcases how word lists can inform each of these strands. His approach is summarised in Figure 2.1.

Figure 2.1

Using Word Lists for Vocabulary Instruction Employing the Four Strands Principle



Not much attention has been given to the role word lists play in directing the process of vocabulary acquisition in a balanced approach and how they are used in different ELT contexts. There are two researchers, however, who have investigated the use of word lists to teach English vocabulary in general and how such lists are viewed by practitioners.

In 2015, Burkett conducted an exploratory study of the uses of word lists in English-intensive university programmes. An online survey was completed by 164 respondents (teachers, administrators, curriculum and assessment developers and coordinators) from 95 programmes around the world, and a follow-up interview was conducted with four of them. The survey results showed that even though 90% of the respondents believed that word lists are useful, only 60% formally used them. Burkett explained that factors such as the limited duration of the programme,

the suggestion but not imposition of word list use, and the lack of teaching experience may have constrained the use of word lists in the study context. It was also reported that classroom teachers and students used word lists more often than course designers and testers. However, details about how word lists are used were not explored.

Another interesting finding is the predominance of the AWL in these institutions. This confirms Nation's (2016) observation about the powerful status of the AWL among other lists. This, however, may have changed after the publication of the Academic Vocabulary List (AVL; Gardner & Davies, 2014), which was relatively new at the time of Burkett's first article. Finally, the distinct word lists used reflected the diverse nature of the programmes examined, and they reinforced the view that an all-purpose word list does not exist.

In his PhD research, Burkett (2017) further investigated the uses of word lists by interviewing practitioners in five English-intensive university programmes in the United Arab Emirates (UAE). Basing his work on Nation's (2016) book and his analysis of the interviews, Burkett established four categories of word list use in pedagogical contexts, which are shown in Table 2.3.

Table 2.3

Pedagogical Purposes of Word Lists for Language Teaching

Purpose	Description
Course design	Setting short-term (course) learning goals Setting long-term (programme) learning goals
Teaching and learning	Selecting material that includes the target vocabulary Selecting words for intentional vocabulary-learning activities Selecting vocabulary from texts to focus on Analysing the vocabulary load of a text
Testing	Modifying the level of vocabulary in tests Selecting words for tests of vocabulary size and proficiency Testing vocabulary in context Using results to inform teaching and learning
Materials development	Setting the vocabulary level of materials Creating vocabulary-focused class materials Creating vocabulary-focused independent-study resources

Note. Adapted from *An Investigation into the Use of Word Lists in University Foundation Programs in the United Arab Emirates* (p. 195) by T. H. Burkett, 2018, University of Exeter.

The most common purpose for using word lists is to establish vocabulary-learning goals for both language programmes and specific courses. Burkett (2017) found that four of the five programmes surveyed used the AWL or a list adapted from the AWL to set learning goals for the programme. According to Burkett (2017), all five programmes used word lists for teaching and learning. Lists used for this objective mainly involved day-to-day practical activities, designing material for intentional learning and checking the level of reading texts and other materials. Three programmes used word lists to develop vocabulary-focused teaching material, while only two used lists to set the level of the teaching material and create materials for independent learning. It seems that three of Nation's (2007) four strands—meaning-focused input, meaning-focused output and language-focused learning—were used in Burkett's (2017) study context (fluency development could not be identified, and thus could be a focus of future research). Four of the five programmes surveyed used word lists for vocabulary assessment.

Banister (2016) examined how the AWL (Coxhead, 2000) was used and perceived by 193 EAP teachers. He found that 72% of the participants ($n = 84$ out of 116 who responded to the question about using the AWL) used the AWL. Generally, both those who did and did not use word lists expressed positive attitudes towards the AWL. Those who used the AWL appreciated that this resource was based on scientific research and that it provided them with structure, especially in EAP contexts in which learners belong to different disciplines. On the other hand, they reported challenges related to the size of the list and the absence of multi-word units. However, the most cited defect was the decontextualised presentation of words.

In a survey of the vocabulary teaching practices of 30 ELT teachers, Rossiter et al. (2016) discovered that 80% were aware of word lists and about 60% were aware of corpora and concordances that were freely available. Varying percentages of teachers reported that they sometimes or often use the following resources to guide their instruction: frequency lists (66%), formulaic sequences (50%), corpora (43%) and concordances (27%). Only 28% used a computer

program such as LEX Tutor (Cobb, 2015) to verify the reading levels of supplementary materials they used in class.

It is important to note that these investigations were explorative and the reports of practices were based on the memory and views of the participants in the survey. They were based on a small-scale study and a sample of specific contexts.

Even though research on the use of word lists in EAP contexts has finally gained attention, the exploration of how word lists are employed in the language classroom is still in its infancy. To the best of the researcher's knowledge, no studies have explored how word lists are used and viewed outside of academia. While there has been some discussion about the specificity of word lists for particular subject areas within academic contexts (e.g. Durrant, 2014, 2016; Hyland & Tse, 2007, see Section 2.3.3.), few empirical investigations have examined how word lists are used and perceived in these contexts. The following section discusses the issues with using word lists in ELT.

2.5.3 Issues with Using Word Lists for Vocabulary Teaching and Learning

The previous section showed the value of word lists in vocabulary acquisition, but the question is why they have been neglected if they are valuable. One answer to this question can relate to the gap between corpus linguistics research and practice, the challenges with using word lists for ELT purposes and teacher cognition.

2.5.3.1 Lack of Awareness of Corpus-based Word Lists. The rapid growth in pedagogical word lists indicates the value that researchers and corpus linguists see in them. While there is an indication that practitioners are using word lists, awareness of corpus-based word lists among ELT practitioners is still not as widespread as it should be. This is based on the small numbers reported in the literature and personal communication; see Section 2.5.2. Furthermore, the extent to which word lists are useful and the process whereby word lists have been chosen remain understudied.

The omission of guidance on corpus linguistics in general and on word lists in particular in ELT teacher training can be seen as one of the main reasons for this lack of awareness. There is a gap

between corpus-based word lists research and practice that is like the gap between corpus linguistics research and teaching practices (e.g. Chambers, 2019; Römer, 2006). In this section, corpus-based word lists (one application of corpus analysis) are discussed with respect to applied corpus linguistics. Nesi et al. (2021) noted that, for instance, *The Competency Framework for Teachers of EAP* and the *TEAP Accreditation Scheme Handbook* (British Association of Lecturers in English for Academic Purposes [BALEAP], 2008, 2014) do not introduce corpora or corpus linguistic approaches, and the *BALEAP Can Do Framework for EAP Syllabus Design and Assessment* (BALEAP, 2013) does not explain how corpora can be used in EAP.

To close this gap, researchers (e.g. Farr, 2008; Römer, 2006) have called for a focus on teacher training. For example, they propose that teachers must understand that the use of concordances should be based on selecting a suitable corpus. Similarly, to use word lists effectively beyond giving them to students to memorise new words, practitioners should understand how word lists are compiled and how they should use them.

Several publications showing how corpus linguistics can be used in teaching have appeared recently. They include Eric Friginal's (2018) *Corpus Linguistics for English Teachers: New Tools, Online Resources, and Classroom Activities* and Robert Poole's (2018) *A Guide to Using Corpora for English Language Learners*. Dana Gablasova launched the *Corpus for Schools* project in 2017, Tatyana Karpenko-Seccombe's *Academic Writing with Corpora* (2021) and Elen Le Foll's (2021) *Creating Corpus-Informed Materials for the English as a Foreign Language Classroom. A step-by-step guide for (trainee) teachers using online resources*. However, this is not the case with corpus-based word lists.

2.5.3.2 Negative Perceptions of Word Lists for Language Teaching. Another issue that may have hindered the use of word lists for language teaching and learning is related to how they are viewed. Word lists have been negatively associated with rote memorisation and unsuccessful decontextualised learning (Schmitt & Schmitt, 2020).

For many teachers, word lists are used to prepare for tests, only to be forgotten later. Many L2 teachers learned vocabulary this way when the grammar translation method predominated. They may have learned the vocabulary, but were unable to use the words or forgot them after the examinations. Thus, they believe that word lists are useless.

The status of word lists has also been influenced by misconceptions and misuses that led to this negative attitude towards them. Folse (2004) refuted these negative arguments against list learning. He acknowledged that list learning may not be interesting/fun for all learners, but it is certainly effective for some. He cited studies in which the use of word lists was found to be effective, and emphasised that preferences for using word lists could be related to learners' educational backgrounds that encourage rote learning or to their different learning styles. Even though the experiments he cited were quite old, the observation that word lists still exist and are used (mostly as a self-study tool) supports Folse's arguments.

Webb, Yanagisawa, and Uchihara (2020) performed a meta-analysis to summarise the extent to which L2 vocabulary is learned via the most extensively researched word-focused activities, flashcards, word lists, writing and fill-in-the-blanks tasks, in 22 experiments. The purpose of this meta-analysis was not to argue for the effectiveness of intentional learning for L2 lexical development, as L2 lexical development is a lengthy and complicated process that frequently involves learning words via a range of intentional and meaning-focused activities. The researchers found that the increase in vocabulary development via intentional vocabulary-learning activities was relatively large in immediate post-tests, but was much smaller in delayed post-tests than had been suggested in reviews of the literature on vocabulary learning (e.g. Nation, 2013; Schmitt, 2000).

Furthermore, the findings indicated that activities contributed to learning in varying degrees. The results showed that both flashcards and word lists led to relatively large gains in knowledge of the form–meaning connection, while writing and fill-in-the-blanks tasks led to relatively small gains. In fact, the finding for flashcards and word lists were in line with the perception that intentional learning is effective and efficient.

Practitioners of the communicative approach have neglected word lists and intentional vocabulary teaching, and have expected students to learn vocabulary incidentally within contexts instead (Lessard-Clouston, 2013). Some practitioners were aware of word lists and had used them. They were aware that word lists should not be given to learners, and some thought that learners lacked engagement with the lists (Banister, 2016).

These two factors, (a) lack of awareness or knowledge and (b) negative perceptions, are interrelated. In other words, if teachers are aware of word lists, they can use them effectively, so they would have positive attitudes towards word lists. While this has been documented in a few studies in EAP, the views of teachers outside academia have not been explored, to the researcher's best knowledge.

2.5.3.3 Challenges of Using Word Lists. Using word lists to teach languages is not without problems. Nation (2016) identified challenges based on his experience with word list development and ELT. Sorell (2013) reviewed these challenges in his PhD research on general English vocabulary. The practitioners that Burkett (2015, 2017) and Banister (2016) surveyed perceived similar issues. These challenges can be summarised as follows:

1. The large size of word lists: Word lists are usually over a thousand words, and it is difficult to implement such a large number of words in a course and ensure that they are learned. To address this problem, Dang (2020) suggested making smaller lists or dividing word lists into sub-lists that could be implemented in courses within language programmes. She cited published lists (e.g. EWL) as examples of lists of manageable size.

2. Word lists present words in isolation: Even when words are presented in context, however, word lists are restricted to declarative knowledge. Foley (2009) and Hirsh and Coxhead (2009) have shown how words from the AWL have been practised in meaningful ways. It is important to note that including context for general academic word lists is more controlled than it is for general English word lists.
3. Lack of materials based on word lists: The number of teaching materials and resources based on published lists is limited (Nation, 2016). Nation attributes this to a lack of awareness among materials developers and course designers and how frequency may inform vocabulary acquisition.
4. Lack of access to potentially useful lists, particularly discipline-specific ones: Many useful word lists are not freely available for practitioners, particularly outside academia.
5. Lack of support and guidance: More research is needed into practitioner-targeted word lists. Schmitt (2016) called on list compilers to provide guidance for practitioners. The PHRASE list (Martinez & Schmitt, 2012) is an example of a list with a guide.

Teachers bring to the classroom their beliefs, attitudes, identity constructions, perceptual frames, and knowledge, all of which inform what they do in the classroom (Borg, 2006). Studying the relationship between these 'mental constructs' and teachers' practice in the language classroom is known as teacher cognition (Borg, 2003). Borg (2001) explained that teacher cognition, 'what teachers think, know, and believe' (p. 81), is influenced by teachers' schooling, professional coursework and classroom experience as well as contextual factors. It is important to note that the terms 'teacher' and 'practitioner' are used interchangeably while practitioner refer to individuals involved in ELT but not necessarily teaching.

Teacher cognition and studies of how vocabulary is taught have focused on what teachers know about the structure of the English lexicon, the relationship between how teachers teach

vocabulary and their beliefs about effective strategies to teach vocabulary, and the contextual factors that inform their decision-making (Schmitt & Schmitt, 2020).

Very little attention has been paid to how practitioners use corpus tools and resources for vocabulary teaching and teacher cognition. The available research (e.g. Boulton, 2009; Mukherjee, 2004; Römer, 2009) has focused on exploring teachers' awareness of corpus linguistics, training them on corpus-based tools for language teaching, and understanding how this informs their attitudes and practices. Other studies have explored teachers' knowledge and opinions about the use of corpus tools in vocabulary instruction, and they investigated instructors' perceptions of using corpus-based materials after they receive training. Results have shown that teachers lack awareness of corpus tools and how to implement them in language teaching. After training, most teachers regarded corpus data as useful for their teaching practices; they stated that corpus-based language pedagogy might have a place in the agenda for language teaching. Although participants' general perception of using language corpus was positive, negative experiences such as technological issues, difficulties in classroom applications and time concerns were also identified. Those studies pinpointed a need to train teachers how to incorporate corpus-based data into their teaching experiences.

To address this gap in research, the current study aims to investigate the knowledge, thoughts and beliefs of practitioners regarding word lists and how they influence their practices.

2.6 Evaluating Corpus-based Word Lists

2.6.1 *What Is Evaluation?*

Evaluation is the systematic review of the merits, value or significance of an object, person or effort according to defined criteria (Yarbrough et al., 2010). In education, evaluating pedagogical products (i.e. teaching materials or resources) refers to measuring their value (or potential value) and judging their effect on potential users (Tomlinson, 2003, p. 15). Hubbard (2006) identified three purposes of software evaluation for computer-assisted language learning (CALL): evaluation for use

in the class, evaluation for self-access or other pedagogical uses and evaluation for publishing reviews or development. These descriptions of evaluation—evaluation in general, evaluation of teaching materials and evaluation of CALL materials—share the idea of judging an object or process and identifying its effectiveness and utility to its users.

Word lists are important pedagogical resources for vocabulary instruction, and their value depends on whether the list used is suitable (Nation, 2016). The purpose of evaluation in the current study is to support decision-making, that is, to decide whether the word list is appropriate for a given setting and to provide insights into how the list may be effectively implemented. The definition of *word list evaluation* in this study reflects the one Kashoob (2018) used for defining the evaluation of ELT materials. It is the process of investigating a corpus-based word list with respect to a given setting by the potential users (with varying levels of expertise) to assess the suitability of the list leading to impressionistic or subjective conclusions about the list and how to adapt it. There are four components in this definition: the purpose of the evaluation (assessing the suitability of the list), the procedure (investigating a list using a tool with respect to a given setting), who conducts the evaluation (the list user, the practitioner) and the results of the evaluation.

2.6.2 Evaluations of Corpus-based Word Lists

In response to the rapid growth of corpus-based word lists in ELT, researchers have begun to evaluate word lists in relation to their intended purposes, to validate in-progress word lists, and to compare or critique existing lists. Many word lists have been evaluated by calculating their lexical coverage, the percentage of the words on the list that are in a corpus (Nation & Waring, 1997). While lexical coverage provides valuable implications about the usefulness of a list, additional factors also must be considered when L2 learners are involved (Dang, 2018).

In the second edition of a key publication in the field of vocabulary acquisition and research, *Vocabulary in Language Teaching*, Schmitt and Schmitt (2020) highlighted the importance of assessing word lists for the purpose they will be used for. They conceptualised word list evaluation as

testing validity; that is, word lists need to be validated against what they are claimed to do just as tests must be validated against what they claim to measure. According to Schmitt and Schmitt (2020), practitioners must assess a word list in terms of how it is compiled, the content of the list and how the list could be used in context. This echoes Nation's (2016) call to examine these elements when assessing word lists, and it resembles my statement in this study regarding the evaluation of word lists. However, while this is a good proposal, there are two obstacles. Many practitioners do not have access to such information about the examined word list. A self-tutoring teacher or a teacher in a public school, for example, does not have access to published papers that present the details about the list. Furthermore, the ability to assess word lists requires knowledge of the complex relationship between understanding how word lists are constructed, the ability to judge them, and use them effectively. In light of this, the current study aimed to propose an evaluation tool and supporting documents that are approachable and accessible to the practitioner who lacks information about word lists. The following section reviews the existing research on word list evaluation, and it highlights the absence of a systematic approach to evaluating word lists that involves the practitioner, the word lists' end users, and that accounts for contextual factors such as learners' characteristics and the purpose of the list. It briefly reviews the literature on the process of developing evaluation tools and checklists to feed into developing the proposed evaluation tool.

2.6.3 Lexical Coverage in a Corpus

Lexical coverage can be calculated for various purposes: overall coverage, average coverage and the coverage of the most frequent items (Dang & Webb, 2016).

Overall coverage calculates the percentage of the items in the list that are found in a corpus (Dang & Webb, 2016). Brezina and Gablasova (2015), when they developed their NGSL1, compared its overall coverage with that of the original GSL in four corpora. The comparison revealed that both lists achieved similar coverage, but the NGSL1 did so with 40% fewer items. Assessment using this

type of coverage could be biased towards long lists and lists with large counting items, such as word families (Dang & Webb, 2016a).

Gardner and Davies (2014) used coverage of the most frequent 570 items of their list, the AVL, to validate it in comparison with Coxhead's AWL (2000). They excluded infrequent items from their AVL. They found that the 570 items of the AVL had almost twice the coverage of the 570 items of the AWL, suggesting the increased suitability of the AVL, relative to the AWL, for university L2 students. The researchers attributed the higher coverage of the AVL to advances in technology that allowed them to derive their list from a large corpus, as well as differences in methodological decisions. Unlike the AWL, which excludes terms from the GSL, the AVL was not compiled with regard to other lists. This may have led to the higher frequency in the AVL. This method shows the value of items in a list. However, it could be biased towards larger lists, as infrequent items are excluded, but it does not evaluate the list as a whole (Browne et al., 2013).

Newman (2016), Qi (2016) and Hernandez (2017) compared the lexical coverage of the AWL and AVL in three independent academic corpora to evaluate these lists for L2 university students. They found that both lists were useful for university L2 students, but the AVL outperformed the AWL in all studies. Qi (2016) concluded that the value of a word list could differ according to the applied evaluation criteria or when L2 learners with different proficiency levels were considered. For example, when the number of items in each list was the same in the three studies (coverage by the most frequent items), the 570 AVL words were found to be more useful than the 570 AWL words for learners with limited English vocabulary. In contrast, the AWL provided more support for learners who had already mastered the most frequent 1,000–5,000 BNC/COCA word families. Moreover, when the lists were compared to discipline-specific sub-corpora, it was found that both lists would be useful for students majoring in business, education or engineering but not as beneficial to students majoring in the arts or humanities or music (Qi, 2016).

To develop their Essential Word List (EWL), Dang and Webb (2016) used both average coverage and coverage by the most frequent items to compare four high-frequency lists—the GSL, the BNC2000, the BNC/COCA 2000 and the NGSL1—in 18 corpora of various text types, sizes and varieties of English. The average coverage was calculated by dividing the overall coverage by the number of items in the list. This method is useful both when comparing lists of different sizes and when evaluating lists as a whole. The comparison revealed that, with regard to coverage by the most frequent items, the NGSL1 performed best, while in terms of average coverage, the BNC/COCA 2000 produced the highest coverage. This means that the BNC/COCA 2000 might be the most useful for learners if used as a whole, whereas if portions of the lists are used, the NGSL1 might be more useful.

Those studies indicate that there is no single perfect approach to evaluating word lists when coverage is used as the criterion. Therefore, different approaches to evaluation should be used to assess word lists (Dang & Webb, 2016; Qi, 2016). Additional issues in examining lexical coverage of word lists in a corpus are related to unifying the unit of counting to make the evaluation accurate during a comparison (Dang & Webb, 2016a). Further, the lengths of the lists and whether it is fair to curtail the length of the list also are important issues to address when the type of coverage is selected (Dang & Webb, 2016). Furthermore, there is likely to be considerable variation between the language of the source corpus and the language encountered by learners in other contexts. Thus, lists with greater lexical coverage may be relevant to learners in a particular context, but not in another (Milton, 2009). For example, Stein (2017) pointed out that some items in the NGSL1 may not be relevant to EFL beginners, and that some necessary vocabulary was missing. She raised the concern of the immediate usefulness for teaching and learning corpus-based word list that is based on objective criteria. Stein's position was supported by researchers like Dang et al. (2020), who examined the relationship between lexical coverage, learner knowledge and teacher perceptions of the usefulness of high-frequency words. They found small correlations between lexical coverage and learner knowledge and teacher perceptions. He and Godfroid (2019) found a moderate correlation

between the frequency of academic words in the COCA and COCA Academic corpora and teacher perceptions of the usefulness of those words. While the studies of lexical coverage cited in this section may have provided valuable observations about the usefulness of the particular lists examined, there are still very few evaluations of word lists that consider the knowledge of the end users—teachers and learners.

2.6.4 *ELT Practitioners' Perceptions of Usefulness*

While words with high lexical coverage are likely to be useful to learners, other contextual factors such as target vocabulary, learning purpose, and learners' characteristics (e.g. age, proficiency level and discipline) must be considered when determining the usefulness of a list (Gerami & Noordin, 2013). Another criterion for evaluating word lists is teachers' perceptions of word usefulness. As end-users of word lists and being closely involved in teaching languages, teachers' judgements of which words to teach are a good indicator of usefulness (Dang et al., 2020). Research in other languages (i.e. French, Italian and Turkish) shows that teacher perceptions (Treffers-Daller & Tidball, 2008), or frequency counts combined with teacher perceptions (Bardel et al., 2012), were better predictors of lexical sophistication than coverage value alone.

Similarly, Dang et al. (2020) used teacher perceptions and the learners' knowledge of vocabulary to evaluate 973 non-overlapping words from the BNC/COCA2000 and the NGSL1. The researchers emphasised the use of criteria other than lexical coverage to evaluate word lists. They asked 78 ESL teachers to rate the usefulness of the selected items, and it examined 135 L2 learners' knowledge of the 973 words. Teachers found the BNC/COCA2000 more useful, and learners' knowledge of items from that list also was higher. This suggests that the BNC/COCA2000 is most useful as a high-frequency list. The researchers attributed this to the selection criteria; the BNC/COCA is based on objective corpus-driven information that has been adjusted subjectively. The second reason for this result may be related to the corpus from which the list was derived. It contained a greater balance of spoken and written texts, representing different varieties of English.

Teacher perception has been used in evaluating published word lists. For instance, Dang (2020) and Omidian et al. (2017) evaluated the potential usefulness of the items in the list. Their purpose was to evaluate published word lists for pedagogical use by considering different aspects related to the context of use. Omidian et al. (2017) asked L2 teachers and advanced learners to judge the pedagogical value of the top 15 multi-word units from three lists developed by Biber et al. (2004), Martinez and Schmitt (2012) and Simpson-Vlach and Ellis (2010). Items reported as valuable by both the teachers and the learners came largely from Martinez and Schmitt's (2012) PHRASE List. When the teachers' and learners' evaluations were analysed separately, it was found that teachers favoured the list developed by Biber et al. (2004), which was frequency-based. Teachers thought the phrases in that list were particularly problematic for students and thus merited increased attention. On the other hand, students selected items from Martinez and Schmitt's (2012) list because they felt these items could be employed in a variety of text types. It is important to note that this latter list relied on expert judgements during development. Evaluating two high-frequency lists based on lexical coverage and asking language teachers to determine their usefulness proved that pedagogical lists, tailored subjectively, are more useful (Dang et al., 2020). Those lists appear to be more pedagogically useful because they consider the teaching context, type of vocabulary and purpose of use.

Omidian et al. (2017) noted that evaluating word lists based on teachers' and learners' judgements may be problematic, as evaluations may be influenced by teachers' experience with the target language. Still, such a study is likely to be informative.

2.6.5 A Tool for Word List Evaluation

This review shows that researchers have become aware of the importance of evaluating word lists to inform research and to validate newly developed lists. These evaluations could be used as a second basis of evaluation by practitioners who are planning to use a word list. However, a problem with these evaluations is that they do not include ordinary ELT practitioners with little or no

background in corpus-based word lists. Despite the value of such evaluative research, practitioners often do not have access to it and, even if they have, they do not necessarily understand it. Dang (2018) proposed a model that teachers could use to select word lists based on the learners' target academic subject areas then determine the learning goals and sequence of learning based on the proficiency levels and learning purposes of the learners. She used the ASWL (Dang et al., 2017) and HSWL (Dang, 2018) to illustrate her model.

Another issue is that these evaluations are focused on widely known published lists, whereas teachers may choose to use unpublished lists or to construct their own. Furthermore, word list evaluation, like word list development, raises issues such as the unit of counting, the list size, practicality, context, the intended use of the list, the individuals involved and the evaluation criteria. These elements have been overlooked in the reviewed evaluations. Moreover, as demonstrated earlier in the calculation of lexical coverage, evaluations have yielded different results when different criteria were used.

One approach to assess word lists is to use evaluation checklists or tools. In reviewing the literature on the design of such checklists or tools, Kashoob (2018, p. 6) expressed dissatisfaction with the available tools. There is a lack of tools 'that can be used by ordinary teachers, who have little or no background in educational research and its academic jargon that take into account teachers' and learners' needs'. A systematic approach to evaluating word lists that considers these factors and is accessible to the practitioner is still needed.

At this point in the review, it is worth highlighting the one evaluation framework that has been developed—the one from Nation (2016, pp. 131-142). That framework for evaluating word lists comprises 26 Yes/No questions that focus on eight aspects (see Table 2.4) related to constructing word lists; the framework is presented in Section IV, *Making the Lists* in Nation's book about word lists and is based on his extensive experience with corpus-based word lists, vocabulary instruction and research. Nation used the framework to critique his own BNC/COCA lists.

Table 2.4*Nation's Questions for Critiquing a Word List*

Focus	Question
Purpose	<ul style="list-style-type: none"> • Was the target population for the word list clearly described? • Was the purpose of the list clearly described?
Unit of counting	<ul style="list-style-type: none"> • Was the unit of counting suitable to the purpose? • Was the unit of counting clearly defined, including issues such as UK and US spelling, alternative spellings, parts of speech, abbreviations and numbers? • Was the unit of counting explicitly well-justified?
Corpus	<ul style="list-style-type: none"> • Was the content of the corpus suitable to the purpose of the list? • Was the corpus large enough to get a reliable result? • Was the corpus divided into sub-corpora so range and dispersion could be measured? • Were the sub-corpora large enough, of equal size and coherent? • Was the corpus checked for errors?
Main word list	<ul style="list-style-type: none"> • Was there an explicit description of what would be counted as words and what would not be included? • Were homographs dealt with? • Were proper names dealt with, including proper name homoforms? • Were content-bearing proper names distinguished? • Were hyphenated words dealt with? • Were transparent compounds dealt with in a way consistent with hyphenated words? • Were acronyms dealt with, including acronym homoforms? • Were the lists of proper names and other lists revised on the basis of initial output?
Other lists	<ul style="list-style-type: none"> • Were marginal words dealt with? • Were any supplementary lists used?
Making the lists	<ul style="list-style-type: none"> • Were the criteria for inclusion and ordering in the list (frequency, range, dispersion or some composite measure) clearly described and justified? • Were the criteria for making sub-list(s) clearly described and justified? • Were any subjective criteria used? Were they described and justified? • Were the lists checked against competing lists not just for coverage but also for overlapping and non-overlapping words?
Self-criticism	<ul style="list-style-type: none"> • Are the weaknesses of the list clearly acknowledged?
Availability	<ul style="list-style-type: none"> • Are the lists readily available in electronic form for evaluation?

Note. From *Making and Using Word Lists for Language Learning and Testing* (pp.131-132) by I. S. P.

Nation, 2016, John Benjamins.

The first part of the framework focuses on whether the purpose of the list and the target audience have been described. The second is concerned with the unit of counting used to group

words into a list. The suitability of the unit of counting in word lists is one of the most important issues that research has addressed recently, as it has a huge influence on the usefulness of the list and its size; hence, on the practicality of implementing it. For example, type-based word lists are six times larger than are word family-based lists, which may limit their practicality (Dang & Webb, 2016b); see Section 2.4.1. for more information about the influence of the unit of counting on the size of the list. As the word family has become the popular unit of counting, recent list developers have begun to examine the different units of counting, and Nation (2016) devoted a chapter of his book to deciding on the unit of counting when making word lists. The second set of questions pertains to the corpus and sub-corpus on which a list is based. Again, Nation devoted a chapter to selecting and preparing the corpus when making word lists. This is followed by questions about the items to be added to the main list and the items to be included in a separate list. The focus of the next questions is on other lists, namely, marginal and supplementary lists, followed by questions about the criteria for list preparation. The last two parts are about any limitations of the list and its availability in electronic form.

Nation's (2016) framework is a useful tool for organising all the important aspects to consider when evaluating word lists. He used the framework to critique and revise his 'long-time tested and validated BNC/COCA word lists' (Ding & Reynolds, 2018, p. 2). However, his framework is based on the 12 preceding chapters in his book; apart from the first one, these chapters pertain to the construction of corpus-based word lists. The framework appears to focus on the issues that should be considered when compiling word lists. Using word lists or selecting a word list for a pedagogical context raises other issues that are not considered in Nation's (2016) list of questions, although some of them are addressed in his book. Furthermore, the questions might be difficult for a practitioner with limited experience and knowledge of word list construction and the tools used for corpus analysis to interpret. In fact, the book is targeted at researchers and those who have experience of vocabulary teaching and learning.

The adaptation proposed in this thesis addresses the non-expert user of word lists. It aims to consider the measures for *using* word lists for pedagogical purposes rather than generating word lists. The tool is expected to help language teachers and learners, as well as syllabus designers and developers of materials, select the best word lists for their purposes. A detailed discussion of how the proposed tool differs from Nation's (2016) work and the aspects that were adapted from his work is presented in Chapter 3 (see Section 3.4.4.).

2.6.6 The Process of Developing Evaluation Checklists and Tools

Different people have used checklists widely to evaluate different products. This section summarises the literature on developing evaluation checklists and tools. The terms tool, checklist and framework are used interchangeably in the current study. While various definitions of the term checklist have been suggested (e.g. Scriven, 2007; Stufflebeam, 2000), the one from Hales et al. (2008) best reflects what is meant by a tool in this thesis:

A checklist is an organized tool that outlines criteria of consideration for a particular process. It functions as a support resource by delineating and categorizing items as a list—a format that simplifies conceptualization and recall of information (p. 22).

When carefully created, validated and implemented, checklists are considered valuable and useful for assessment (Stufflebeam, 2000). Scriven (2005, 2007) summarises the advantages of checklists in the following points:

- Checklists reduce the chances of overlooking some important aspect.
- Checklists in general are easier than most theories or statistical analyses for a non-expert user to understand and validate. In the context of the current study, most practitioners for instance cannot conduct lexical coverage analysis to evaluate word lists, as they might not be able to interpret the calculations.

- Checklists reduce the influence of the halo effect (i.e. the tendency of a positive judgment of an evaluation based on a particular aspect of it); they force the evaluator to consider each dimension that might be relevant. Thus, popular lists such as Coxhead's AWL is assessed for the target context and adaption is planned.
- Checklists reduce the influence of the Rorschach effect (i.e. the tendency to see what one wants to see); they force a separate judgment on each dimension and a conclusion based on those judgments.
- Checklists frequently incorporate a significant amount of information about the particular examined object. Specifically, checklists are a form of knowledge about a domain. Generally, evaluative checklists can be developed more easily than theories of the evaluated object.

Scriven (2007) has suggested certain requirements for checklists: the list should be complete or very close to complete (i.e. avoid significant omissions) and concise, the items should be contiguous (non-overlapping) and the criteria should be clear. Scriven (2007) also warns the developers of checklists to keep in mind the balance between ease of use and length, to consider brevity, which is desirable, and clarity which he described as essential. Specifically, 'The design of good checklists should be related to ease of recall and understanding as well as comprehensiveness and ease of implementation' (p. 80). An evaluation checklist should identify relevant dimensions of value, clarify the criteria for measuring the activity or performance, help the evaluator remember important criteria and enhance the objectivity, credibility and reproducibility of the assessment.

Stufflebeam (2000) offered eight criteria for evaluating a checklist: (1) applicability to its intended purpose, (2) clarity, (3) comprehensiveness, (4) concreteness, (5) ease of use, (6) fairness, (7) parsimony and (8) relevance to the content area. Although these criteria were designed for evaluating checklists, the proposed tool shares the purpose of a checklist, i.e. assessing the suitability of word lists for use, so these criteria are used in the current study.

There is a growing body of literature aimed at developing and validating tools and checklists to evaluate teaching materials, CALL software and educational or medical applications. Kashoob (2018) developed a checklist for teaching materials from design-based research. Gladman et al. (2020) developed and tested a tool for rating the value of a mobile app to support the training of healthcare professionals. Llorens-Vernet and Miró (2020) used the Delphi method to create and validate a guide for stakeholders to develop and assess mobile health apps. Schroeter (2008) used a qualitative non-experimental and exploratory method. Guidy-Oulai (2009) employed a three-phase data-collection process, but she had no clear specification of the methodology, Walker-Egea (2014) used a mixed research method with four phases. These evaluation tools return a numeric value or a percentage in relation to the examined product. Accordingly, they have been validated based on critical judgements (that is, empirically).

2.7 The Research Gap and Rational of the Study

The question of how much vocabulary an L2 learner needs in order to use the language effectively has been of central importance to the global ELT community (e.g. Nation & Waring, 1997; Schmitt et al., 2017). Researchers and practitioners have been aware of the relative value of certain vocabulary items above others for L2 learners as early as the 16th century (Fries & Traver, 1950). One of the approaches to systematising vocabulary instruction has been compiling lists of important words based on how frequently they are used which can be traced back to the Vocabulary Control Movement in the early 20th century.

This chapter observed the uneven past of word lists in English language teaching, which was affected by the status of vocabulary instruction in the field and the advances of corpus-analysis tools. An initial goal of this study was to explore the current status of corpus-based word lists as they are used and perceived in ELT today. The recent explosion of corpus-based word lists indicates an acknowledgement of the value of word lists for ELT. However, little is known about how word lists are used to inform vocabulary instruction inside and outside the classroom, who is using them and

how valuable they are from the perspective of those who use them. Research reviewed in this chapter has focussed on the theoretical and technical issues of compiling word lists and a few studies on the uses of word lists in EAP (Banister, 2016; Burkett, 2015, 2017). For this purpose, the initial explorative phase surveyed a wide range of domains in ELT to get a clear understanding of the status and views related to word lists in ELT. In the first phase, I was aiming to understand the range of uses of word lists and practitioners' practices before moving to the main study.

The primary aim of this study was to propose an evaluation tool for assessing the suitability of a given corpus-based word list for use in the ELT context. The appearance of a large number of word lists in the last two decades indicates an acknowledgement of the value of word lists (which could be a good sign), yet it puts the onus of the challenge on the practitioner to select the right list. Nation (2016) stresses the importance of using a suitable word list. Research on word list evaluation have focussed on calculating coverage, which is a good indicator of word list usefulness, but the problem with this is that it might not be accessible to regular ELT practitioners and factors other than lexical coverage need to be considered in regard to the educational context to assess the suitability of a word list for use in ELT. The absence of a systematic approach to word list evaluation is the main problem that motivated the present study. The study aimed to develop and test an evaluation tool for ELT practitioners, particularly those with little corpus-based word list construction background. This tool aims to guide the practitioner through an analytical process of analysing their contexts and learners' needs and, with this in mind, assess their word lists and plan implementation before putting them into use. The tool is based on Nation's (2016) critique framework which has been reformulated by a thorough literature review and findings generated by the initial survey. The tool focuses on assessing corpus-informed word lists for ELT purposes. The development of the tool can be seen as mediation of research into a document accessible to the regular teacher and other practitioners. In particular, the tool aims to help the practitioner analyse their context when assessing the unit of counting, the corpus from which the list was derived, how it was compiled and how they can

implement and adapt the list for maximal usefulness. An emerging aim was to raise awareness of word lists and bridge the gap between corpus-based word lists research and ELT practice. Corpus linguistics research and word lists development have relied heavily on statistical data driven from corpus-analysis and have ignored the voice of the end-user, the practitioner. This study calls for involving the practitioner's perspective in corpus-based word list research to maximize the usefulness of such research.

Evaluation tools are one of the most valuable means for evaluating practices and resources. Most of the literature on developing evaluation checklists and tools proposes using numeric values to validate these tools. However, these systems ignore the regular practitioner (Kashoob, 2018) and those factors that do not return a numeric value. This study has filled a gap in the methodology of developing and testing evaluation tools by using an approach that involves practitioners in their design by considering the target users and context and by involving experts in corpus-based word lists development and use.

The study contributes a tool for the ELT community and highlights the gap that exists between corpus-based word lists research and production and practice. Proposing the introductory document could be seen as an attempt to bring research and practice closer together in a document that is accessible to regular practitioners. Furthermore, it draws the attention of list developers to the needs of the learners and the practitioners who are using word lists, and to the voices of the end-users of word lists, practitioners, in research related to word lists which would improve the applicability and usefulness of word list research.

2.8 Summary

This chapter outlined the history of word lists from the early 16th century to their use today. Word lists are the fruit of the Vocabulary Control Movement in the early 20th century. That movement produced one of the most influential lists, the GSL. Yet the systematic approach to vocabulary instruction was ignored between the 1940s and 1980s in favour of incidental vocabulary

learning. When large corpora and more sophisticated corpus-analysis tools became available late in the 20th century, the two competing approaches of the Vocabulary Control Movement—statistical analysis and subjective selection based on corpus analysis —came back into vogue.

This review then moved to present the most influential corpus-based word lists for language teaching and learning published in the last 20 years. The diversity of those lists reflects the different functions of word lists, different opinions of what a word is and the different ways to organising word lists. The discussion concludes that issues of word list construction are controversial and arguments over the counting units, corpus and approach depend on the purpose of the list and the learners with whom it will be used. The next section discussed why word lists are useful and how they can inform vocabulary instruction and research. The aim was to dispel the many myths and misconceptions that surround word lists. In light of teacher cognition, the factors that led to misconceptions about and neglect of word lists also are discussed.

The final section details the literature on evaluating word lists and conceptualising a ‘suitable’ word list. Research on word list evaluation has focussed on calculating coverage, which is a good indicator of word list usefulness, but it might not be accessible to regular ELT practitioners. The absence of a systematic approach to word list evaluation is the main problem that motivated the present study. The study aimed to develop and test an evaluation tool for use by ELT practitioners. Most of the literature on the process of developing evaluation checklists and tools deals with numeric values. The next chapter presents the emerging research design that led to the four phases of developing and testing the tool. The final section described the research gap that has been exposed in the review of existing research. It stated the purpose of the study and the objective of each phase. It also highlighted the contributions of the study to pedagogy and research.

3 CHAPTER 3. DESIGNING AND REVISING THE CORPUS-BASED WORD LIST EVALUATION TOOL

3.1 Introduction

This chapter presents the process of developing and revising the corpus-based word list evaluation tool during this study, beginning with an overview of the research design. It is organised chronologically: the first three phases of this project—the preliminary survey, the development of the evaluation tool and reviewing the tool—are presented in this chapter. Each section pertains to a particular phase of the project. Then, it reports the design and results of Phase 1: the initial exploration of the state and general perceptions of word lists in ELT. Next is Phase 2 where the initial version of the proposed tool is designed. The final section of this chapter details Phase 3: a review of the tool by experts and potential users and the revisions of the tool based on their feedback. The design and the results of the fourth phase, the testing phase, are laid out in detail in Chapter 4. Considering that phase's content, length and significance, a whole chapter is dedicated to it.

3.2 Research Design

This study employed a mixed-methods design. This type of design involves collecting or analysing both quantitative and qualitative data in a single study (Creswell & Clark, 2017). The combination of quantitative and qualitative data in this study provided a better understanding of the focus of this study—word lists in ELT how word lists are used and perceived. It also allowed for different groups of participants and different types of data to contribute to the construction and validation of the tool. This coincides with the premise of mixing methods as argued by Creswell and Clark (2017).

The underlying epistemological position of the researcher in this mixed-method study was to be pragmatic. This paradigm makes the purpose of the inquiry and the research design central, and it gives the researcher flexibility to choose the most suitable methods (Creamer, 2017). This study is a response to two problems in ELT practice: the absence of tools for evaluating corpus-based word lists for pedagogical use and the lack of manuals to introduce word lists to the non-expert practitioner.

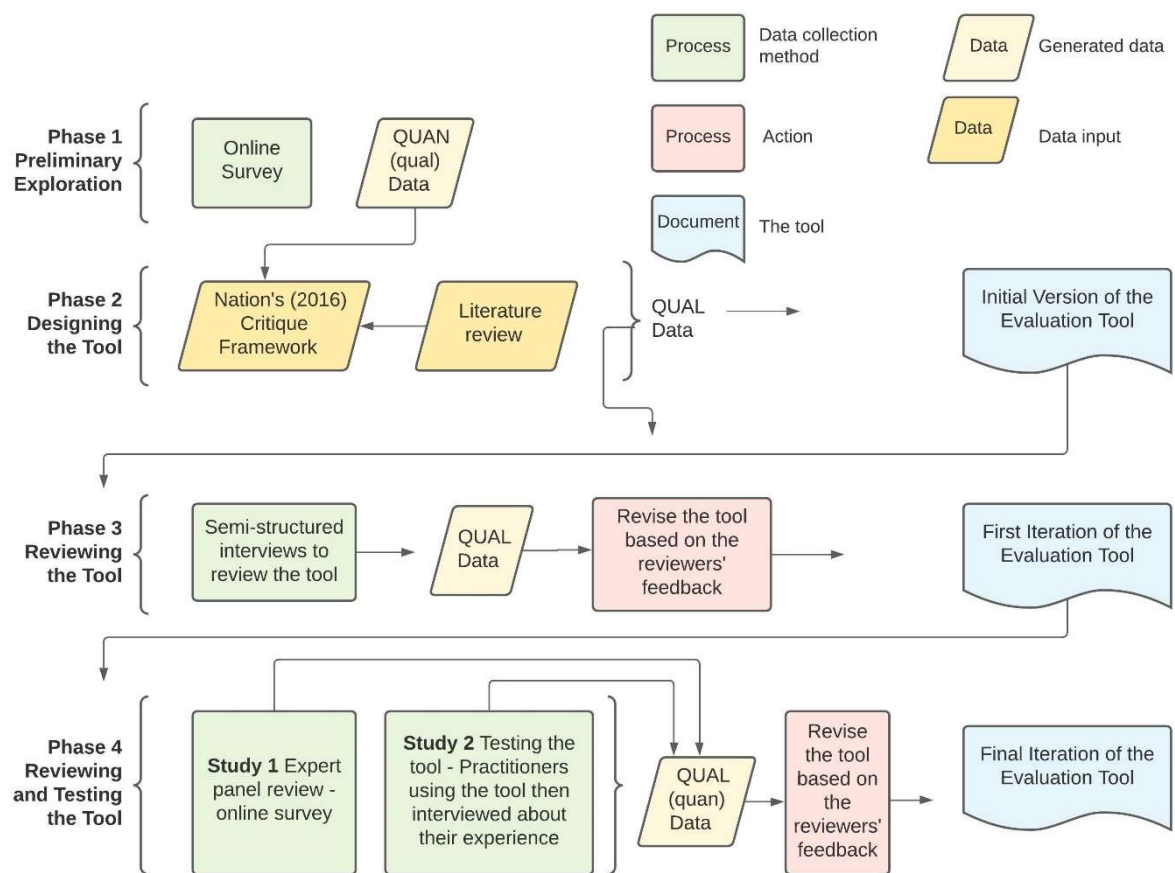
Researchers of the process of developing evaluation tools like Gladman et al. (2020) and Walker-Egea (2014) focus on numeric values, which cannot be employed to evaluating word lists, at least in the tool's current format (see Section 2.6.6). Therefore, developing and testing a tool that leads to a subjective conclusion was guided by what was known and by involving potential users. This paradigm has several features. First, it supports a relational epistemology, that is, relationships in research are best determined by what the researcher deems appropriate to a particular study. Second, it uses a non-singular reality ontology where the emphasis is on what works; theories about the nature of truth and reality are avoided (knowledge is context-specific). Third, it selects an appropriate methodology. In this study, it is a combination of quantitative and qualitative research methods. Fourth, it has a value-laden axiology: it is research that benefits people (Creamer, 2017).

Creswell and Clark (2017) identified three core mixed-methods designs: the convergent, the explanatory sequential and the exploratory sequential. In the convergent design, the researcher compares or combines the qualitative and quantitative data to obtain a broad understanding of the research problem or to validate one type of data. The explanatory sequential design has two phases: collecting and analysing quantitative data, then collecting qualitative data to expand or explain the quantitative findings from the first phase. The exploratory sequential design comprises three steps: a qualitative exploratory step followed by the development of an instrument or approach that is then tested quantitatively. In this design, the development of the instrument is informed by the initial qualitative exploration. The mixed-methods design of the present study emerged during the process of conducting the study. Figure 3.1 outlines the process of developing and testing the proposed evaluation tool. The process follows the order of explanatory design, but the purpose is more like that of an exploratory sequential design (to explore, design and test). The purpose of the explanatory sequential design (quantitative data first) is to explore quantitative data first, followed by developing a qualitative explanation (Creswell & Creswell, 2017).

It was difficult to follow a fixed mixed-method design because of the nature of the inquiry in this study. The data collection and analysis were formulated during the process of the study and based on the findings of the preceding step. Complex designs, though difficult to explain and justify, are acceptable (Creswell & Clark, 2017).

Figure 3.1

The Process of the Research Design



The study began with a preliminary survey of current practices and views related to the construction, exploitation and evaluation of corpus-based word lists for English-language teaching and learning. An initial collection of data (primarily quantitative) was needed to determine how ELT practitioners viewed and used word lists in language-learning contexts before it would be possible to develop a soundly informed tool for ELT practitioners. An exploratory phase was needed because of

the limited literature on the exploitation of word lists in ELT at the time of the study. Collecting qualitative data would have been time consuming. The second step involved developing the proposed tool, which was adapted, according to the literature on word lists and the themes generated by the initial exploration, from Nation's (2016) framework. The third step entailed conducting interviews to review the tool in terms of its effectiveness, content and efficiency and then revising it accordingly. Collecting quantitative data at this step would have led to artificial findings. The final phase involved testing the tool from two perspectives; the expert and the ELT practitioner. In that phase, quantitative and qualitative data were converged (during interpretation). This final iteration of the tool is qualitative, so it could not be tested quantitatively. Each phase was used to support and inform the development of the tool and to plan the next step.

The following sections present the aims and detailed description of the methods employed in each phase along with the findings that informed the next step and the development of the tool.

3.3 Phase 1: Survey of the Uses of and Views on Word Lists in ELT

3.3.1 Aim

As shown in the literature review in Chapter 2, research on word lists has focused on the technical and theoretical aspects of their construction, while little attention has been paid to questions regarding the use and perceptions of word lists in pedagogical contexts. Accordingly, a preliminary survey was conducted to explore the following questions:

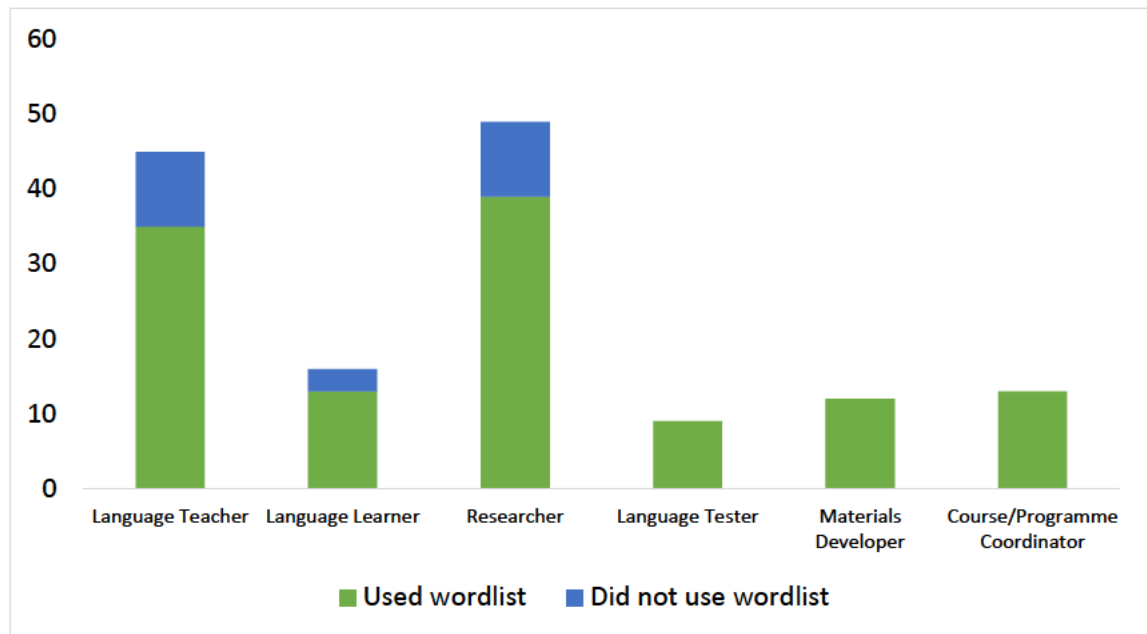
- RQ1: What distinct uses, if any, are made of corpus-based word lists by practitioners, researchers and learners involved in L2 teaching and learning?
- RQ2: What methods are used to generate pedagogical corpus-based word lists?
- RQ3: Which measures can be used to evaluate a word list's utility for ELT practitioners, researchers and language learners?
- RQ4: How are word lists perceived by ELT practitioners, researchers, and learners?

The data collected in this preliminary survey was needed to understand word lists in ELT contexts: who uses them, for what purposes, in what ways, and how ELT practitioners, researchers and learners perceive their pedagogic value. The results of this survey were published in a paper (Thompson & Alzeer, 2019).

3.3.2 Participants

Seventy-four individuals involved in English language teaching and learning—including practitioners, researchers and learners—completed the online survey. The following inclusion criteria were defined before the survey was launched: respondents should be aged 18 or older, and they should be involved in teaching or learning ESL, including EFL. The second inclusion criterion was the use of word lists for language teaching and learning. Of the 74 respondents who completed the questionnaire, 59 (79.7%) reported that they had used word lists for language teaching and learning, while 15 (20.3%) reported that they had not. Those who reported not using word lists were asked why they had not done so, after which the survey ended for those respondents. Respondents were asked to report all their roles. Within each role, respondents were divided into two groups, namely word list users and non-word list users. Not restricting participants to one reported role assisted in reflecting on the roles of practitioners who used and who did not use word lists. As Figure 3.2 shows, no responding course/programme coordinators, material developers or testers did not use word lists.

Figure 3.2.

Roles of Respondents

The reason for recruiting participants with different roles and from a wide range of domains was that the researcher needed to better understand the status of word lists in ELT and how they were perceived from the points of view of those who used and who did not use word lists. The few publications that have explored how word lists are used and viewed (Banister, 2016; Burkett, 2015, 2017) have focussed on the use of word lists in EAP contexts; the targeted practitioners had been using word lists and the responses were reported only from the points of view of the practitioners. While this small sample size ($n = 74$) may not be representative of the ELT community as a whole, its range of representation was adequate for this preliminary exploratory step.

Almost one-third of the 74 respondents who responded to the online survey in the current study were involved in teaching and/or learning general academic (78.4%) and/or specialised (40.5%) English at the university level. More than 40% of the respondents were focussed on general English for adults and/or school students. Two respondents reported that they were teaching English for professional training. This may indicate that word lists were more popular amongst practitioners in

academia than they were in general or professional English language teaching contexts. Caution must be exercised when considering these numbers, as they may reflect a potential bias due to accessibility to the researcher and the use of the snowball-sampling technique (see below) towards practitioners in academic contexts.

Unlike previous research on word list use (Burkett, 2015), which consisted of reports on students' uses and views by members of staff on English language programmes, this explorative survey examined the use of word lists directly from the learners' point of view. Furthermore, the current survey did not insist on the prerequisite of the use of word lists to participate; it included practitioners, learners and researchers who were not using wordlists to determine why they did not use word lists and what they thought about them. However, a limit on the generalisability of the findings from this group might be the insufficient number of collected responses. The survey design did not collect sufficient information from non-users of word lists; furthermore, their lack of interest in (or awareness of) word lists might have led to their under-representation. For example, one respondent reported that they had not used word lists for language teaching although they reported using a list of vocabulary items from the novels they taught. This respondent did not realise that their list of vocabulary was a word list even though it was not clear whether the list was frequency based.

Another reason for recruiting from different domains in ELT was that the initial target users of the proposed tool included learners and researchers, as well as practitioners. It was believed that the data collected from a single group of practitioners might not have addressed the needs of the broader ELT community. However, in Phase 2, the target users of the proposed tool were limited to practitioners, and learners were eliminated because the researcher considered the process of evaluating word lists to be too complicated; moreover, asking learners to evaluate word lists is challenging, and may not even be an option in some cases. It is important to note that the results that informed the development of the tool were obtained primarily from the participants who had

used word lists. These results included challenges in using word lists, how they were used and how they were selected.

The researcher used a non-probability snowball-sampling technique to identify potential participants. In this technique, members of non-probability samples are selected according to the resources available to the researcher; they are not randomly selected. This technique was used because it was not possible to employ systematic purposive sampling online (Dörnyei, 2007) as it was difficult to obtain a complete list of members of the large ELT community from which participants could be randomly selected. Snowball sampling is the process of asking participants who fit the study-inclusion criteria to identify additional eligible candidates (Dörnyei, 2007). Snowball sampling can lead to over-representation of a particular viewpoint or context (Cohen et al., 2018). In the present study, participants interested in the use of word lists in academic contexts may over-represented the ELT community. Notably, however, the survey aimed to explore the current status and views of word lists in ELT, not to generalise the results (Cohen et al., 2018). Therefore, the challenges typically posed by snowballing sampling were not at play in this study. I replicated these sampling procedures in the subsequent phases of the study.

3.3.3 Data Collection and Analysis

The preliminary survey was administered online to efficiently collect data about how word lists are used (Ball, 2019). Conducting the survey online allowed access to a specialised international population that would otherwise be disparate and remote to reach (Dörnyei, 2007). Other advantages of the online survey included the convenience of automatically processing responses once a questionnaire was completed; a flexible design that changed according to respondents' answers, smoothing the flow of the questionnaire; and the ability to permit respondents to complete the survey at their leisure (Cohen et al., 2018). However, survey design, particularly closed-ended questions, can sometimes influence respondents' answers too strongly, leading to non-representative and potentially superficial research results (Babbie, 2010). Therefore, optional open-

ended comment spaces were added so as to allow the respondents to elaborate on their answers.

The survey was self-administered, meaning respondents decided whether they wanted to participate. This may have led to sample bias because it is possible that only respondents interested in word lists elected to participate (Ball, 2019).

3.3.3.1 The Questionnaire. An online questionnaire consisting of 14 questions (10 closed- and four open-ended questions) was designed and administered via QualtricsXM (2018). Qualtrics is General Data Protection Regulation compliant, and the subscription to the service was provided by the University of Birmingham. To the best of the researcher's knowledge, no published questionnaires were suitable for the current study. Burkett (2015) did employ a questionnaire to survey the uses of word lists in intensive English foundational programmes; however, that questionnaire was aimed at language institutions. Therefore, I designed a new questionnaire for the current study (see Appendix B).

This questionnaire contained both open- and closed-ended questions. Several questions were mixed: they allowed participants to provide specific answers and to follow up with additional comments. The optional comments were included to create opportunities for respondents to elaborate on their responses and provide examples, which the researcher used to derive themes and offer illustrations. The closed-ended questions required participants to choose from a set of given options, which led to quantitative data (Dörnyei, 2007). Some were multi-select multiple choices questions—that is, more than one answer could be selected—thereby allowing the respondents to convey their situations and opinions more accurately. The questionnaire contained routing questions, the responses to which resulted in the appearance of relevant adaptive questions as the survey progressed. For example, if a respondent reported that they have used self-made word lists, the resultant adaptive questions would ask how and why they created the lists, but questions about ready-made published word lists would not appear. Adaptive questioning reduces the complexity of questionnaires (Ball, 2019).

The survey consisted of items classified by Dörnyei (2007) as factual, behavioural and attitudinal, and it generated nominal, ordinal and word-based data. Factual questions were used to explore the characteristics of word lists users; behavioural questions were used to explore what the respondents did or have done with word lists while attitudinal questions were used to find out what the respondents thought about word lists. The first part of the survey aimed to identify the types of respondents surveyed (i.e. their ELT roles and contexts) and whether the respondents used word lists. It also aimed to identify why non-users elected not to use word lists.

The second part of the survey was administered to those who did report using word lists. Because there is little agreement on what constitutes a word list, the first question of this portion asked respondents to define 'word list'. It then investigated the types of word lists used and the purposes of such use. In response to each word list type listed in an answer, the respondent was presented with adaptive questions querying their motivations for using that type of word list and how it was prepared and used. The final section of the survey concerned respondents' evaluations of word lists—e.g., the measures used to select words, any problems related to the use of word lists, and respondents' perceptions of the effectiveness of word list use. Word lists' names, as well as the names of other resources used in conjunction with word lists, were also collected.

The respondents were identified and recruited via emails accessible to the researcher and direct email contacts (searching profiles online for, e.g., developers of corpus-based word lists). A recruitment email that explained the purpose of the study and provided a link to the questionnaire page at QualtricsXM was sent, containing a consent form, a respondent information sheet (approved by the university's Research Ethics Committee) and detailed information about the research (see Appendix B). Participants were required to indicate their consent before they could begin the questionnaire. Additionally, participants were informed of their right to withdraw from the project any time until two weeks after the survey-completion deadline. Participants would face no consequences for withdrawing from the study. No participant withdrew.

Email addresses were collected to identify the participants in case they agreed to be contacted for participation in subsequent steps of the study, requested a summary of the findings or decided to withdraw their data from the study. The collection of email addresses also allowed the researcher to screen out multiple submissions by the same respondent. Participants were informed that their email addresses would not be analysed or published as part of the research and would be deleted from the data files once the research was complete.

The recruitment email also invited recipients to circulate the invitation to other potential respondents who fit the inclusion criteria and might be interested in the study. The questionnaire was open to any participant who had the survey URL address. Compensation for participation was not offered, but a summary of the findings was sent to the participants who requested one.

The questionnaire was open for two months. Forty-five invitations were sent, and the response rate was checked weekly. Two weeks after the initial recruitment email, one follow-up email was sent to each non-respondent and in response to each incomplete survey submission. At least 117 respondents opened the survey link, but only 74 completed it, representing a response rate of 63%. Partial responses were excluded from the analysis. Thirty respondents finished only the first section, while 12 stopped in the middle but did not provide an email address. These partially completed questionnaires without associated email addresses were excluded from the study because those respondents might have opened the URL on another occasion and completed the questionnaire. One invited respondent did not consent to participate.

QualtricsXM tools were used to perform simple statistical analysis and identify the frequencies and percentages associated with the closed-ended questions. Most multiple-choice questions allowed respondents to select multiple answers and provide open-ended responses. Therefore, percentages were calculated as a share of the total respondents, not the total responses. Open-ended questions and optional comments were manually analysed for themes before statistical analysis took place, as suggested by Sue and Ritter (2012).

3.3.4 Results of the Preliminary Survey

Respondents who reported using word lists were asked to provide their own definition if they did not agree with the three proposed definitions. Fourteen respondents each selected the first option, which emphasised frequency, or the second, which emphasised the importance of the words, while 25 respondents selected the third option, a more comprehensive definition of lists of words selected/ranked according to certain criteria and for a specific purpose. One respondent elaborated by saying that a word includes both single- and multi-word units and noted that the criteria include frequency. Two respondents wanted to select all three options, and one wanted to select the first two but in consideration of context and purpose. Most curiously, one respondent, a language learner, defined a word list as a list of words learned in class.

3.3.4.1 Practices Related to Word Lists in the English Language Context. A decisive item on the questionnaire asked if the respondents have used a word list for a second language teaching/learning purpose. Almost 80% of the respondents used word lists, while the 20% ($n = 15$) who did not use word lists explained that they had not considered doing so ($n = 8$). Some felt that word lists were useless ($n = 3$), while others reported that they did not know what word lists were or were unaware of the existence of ready-made lists such as the AWL ($n = 2$). Two respondents did not provide justification for not using word lists.

Those who reported using word lists were asked about the purposes of that use. Table 3.1 presents the responses to that question, which clearly shows that language teaching (50.8%) and learning (40.6%) are the most common uses of word lists. Interestingly, eight language teachers who used word lists noted that their use of those lists was unrelated to classwork; students were expected to learn words from the list—in some cases independently outside of class time—and were then tested on them. Research purposes included comparing existing lists, investigating the coverage of lists in a corpus, developing lists for various purposes, measuring proficiency and examining the order of vocabulary acquisition.

Table 3.1*The Purposes for Which Participants Used Word Lists*

Purpose	<i>N</i>	%
Language teaching	30	50.8
Language learning	24	40.6
Language testing	16	27.1
Materials development	16	27.1
Course design	12	20.3
Research purposes	12	20.3
<i>Total</i>	59	100

The survey also explored the types of word lists used. Those details are presented in Table 3.2, which reveals that a plurality of respondents (47.5%) used word lists synthesised from the combination of published and self-made lists to achieve their goals. Others (32.2%) preferred to use ready-made lists, and in some cases (20.3%) self-made word lists were used. It is important to note that five respondents added notes to state that they interpreted “combining” published and self-made lists as using both types in different contexts and for various purposes.

Table 3.2*Types of Corpus-Based Word Lists Participants Reported*

Type	<i>N</i>	%
A combination of published and self-made word lists	32	47.5
Ready-made published word lists	19	32.2
Self-made word lists	12	20.3
<i>Total</i>	59	100

Those who combined published and self-made lists felt that different word lists were available for their purposes and situations, and hence, they combined two or more lists to create single, more efficient lists. This was done mostly in response to dissatisfaction with published word lists ($n = 16$), as summarised and exemplified by the following reported observations:

- Most lists target receptive vocabulary development.
- Lists with both single-word and multi-word units are limited.

- Words in the list do not match the students' needs and proficiency levels.
- Most lists are organised based on frequency, not based on themes or topics.

Among those who decided to use ready-made lists, the primary motives of that approach were:

- The suitability of the word list to the purpose and context (84.2%).
- The popularity of the list (52.6%).
- Availability in electronic form (21%).
- The requirement of the list by an authority (15.7%) and/or its suggestion by a colleague (10.5%).

Interestingly, two respondents explained that their list selections had been based on the unit of counting that was suitable for their target learners. For example, one respondent reported that they selected a lemma-based list, as it was suitable for their beginner level students. In terms of how ready-made word list users employed those lists, the majority used them without modification (84.2%), an unexpected result as there is no fits-all-sizes word list (Nation, 2016). The only exceptions were three respondents who decided either to delete ($n = 1$) or add ($n = 2$) items to make the list better conform to their purpose and context.

Respondents indicated in the optional comments that they assembled their own lists when published lists were unsuitable for their purposes and contexts. Other reasons to create one's own word lists were related to lack of validity and reliability on the part of the published lists, a lack of documentation of how some lists were made and a lack of access to published lists. One respondent further added that they created their list by updating an older published list. Two respondents commented that creating a word list was a good learning activity.

To further investigate word list-related practices, respondents were asked to name the word lists they used and, if they were unpublished, to briefly describe them; 53 out of the 59 respondents answered this question. Table 3.3 presents the names of the lists mentioned more than once; some

respondents used more than one list. The AWL ($n = 19$) was highly cited, and about half as many respondents used the AVL ($n = 9$). The BNC/COCA 2000, GSL and NGSL seem to be used at equivalent rates. Brezina and Gablasova's (2015) NGSL1 was referred to once, Browne's (2013) NGSL2 was named thrice, and on two occasions, it was unclear which new GSL was being referred to.

Table 3.3

Names of Reported Word Lists

Name of word list	<i>N</i>	%
Academic Word list (AWL)	19	35.8
Academic Vocabulary List (AVL)	8	15
BNC/COCA 2000	6	11.3
The General Service List (GSL)	6	11.3
New General Service List (NGSL)	6	11.3
New Academic Word list (NAWL)	3	5.6
<i>Total</i>	53	100

Other lists seem to fall into one of four categories:

- Word lists for English for General Academic Purposes (EGAP)—29 (58%). Some of these lists are specific to an institution and correspond to its teaching materials, while others were designed to develop the academic vocabulary necessary for writing.
- Word lists for English for Specific Academic Purposes (ESAP)—12 (24%). These are lists developed for a specific discipline, such as engineering or accounting and management. Some of the lists mentioned in this category were published lists reported only once thus not include in Table 3.3.
- Word lists for General English Purposes ($n = 9$).
- Three respondents reported using lists of other languages or for translation purposes.

There were lists self-made for specific purposes that were not published 18 (36%). Those include lists for particular courses, lists of reporting verbs, and lists for personal use such as communication purposes or new vocabulary. Unfortunately, respondents did not provide information about the development of those lists.

Word lists are among a number of tools used in the ELT context. Therefore, respondents were asked to report the resources used in conjunction with word lists. Of the 59 respondents, only 34 reported using resources other than word lists to achieve their goals. Eight did not respond to the questions about other resources, while 13 reported not using other resources at all. A variety of resources were reported, but three stand out:

- Websites or online exercises for practice.
- Concordances or teaching texts to identify the context of the target word.
- Other concordancing software and vocabulary-profiling software.

3.3.4.2 Making Word Lists. RQ2 '*what methods are used to generate pedagogical corpus-based word lists?*' aimed to explore approaches to making word lists and how they may affect attitudes towards word lists. Respondents who reported making word lists were asked to rank the importance of six criteria commonly used in word list compilation (namely frequency, range, dispersion, intuitive judgment, qualitative methods and duplication of previous lists) and were allowed space for further explanation.

The difference between the criteria's ranks was insignificant, which could be a result of the small number of respondents ($n = 12$). In general, however, statistical measures (frequency, range and dispersion) were considered more important, although comments throughout the survey stressed the importance of expert consultation in conjunction with statistics. Alphabetical and necessity-based organisation schemes were reported as criteria for the selection and ranking of words in lists.

When asked about the justification for their approach to making word lists, respondents generally reported using the criteria of a published researcher—sometimes with modification—or the criteria provided by the software in use. However, one respondent reported developing their own set of criteria.

3.3.4.3 Measures of a Good Word List. As for selecting word lists, Table 3.4 shows the measures thought important by list users. Topic relatedness, i.e., cohesion in the subjects to which words in a list pertain appears to be the most important measure ($n = 38$). Just over half of the respondents ($n = 30$) indicated that word lists should be based on corpus evidence. Roughly an equivalent number felt that subjective judgements ($n = 22$) and statistical measures ($n = 21$) were important bases for word selection and ranking when designing or choosing word lists, as confirmed in respondents' extended comments.

Table 3.4

Measures of Good Word Lists

Measures	<i>N</i>	%
Topic-related	38	64.4
Based on corpus evidence	30	50.9
Based on subjective judgements for selection and ranking	22	37.3
Based on statistical measures for selection and ranking	21	35.6
Small in size and divided into sub-lists	19	32.2
<i>Total</i>	59	100

Some additional measures were identified in the open-ended comments:

- The suitability of the list to the purpose and situation was emphasised and elaborated on, such as discipline-relatedness, students' needs, the proficiency level of the words and pedagogical relevance.
- The availability of the context of the words in the list emerged as a theme throughout the survey.

3.3.4.4 Perceptions of Word Lists. One of the main goals of this exploratory phase was to shed light on attitudes towards word lists for ELT purposes. Though the majority of respondents thought of word lists as useful ($n = 45$ out of the 59 respondents who used word lists), a sizeable number were negatively disposed towards word lists: 11 doubted the usefulness of word lists, while three affirmatively stated that word lists were useless.

When asked about the problems associated with using word lists, 11 reported that there were none, two did not respond, and all others agreed that there were problems. The large size of word lists was the most challenging aspect of dealing with them ($n = 32$), followed by a lack of explanation on how to use word lists or how they are made ($n = 22$). Other problems included format of ($n = 6$) and access to ($n = 5$) word lists.

In the optional comments, respondents mentioned other problems as well. These can be summarised by the following points:

- Word lists lack context. They are lists of isolated words with no context nor definition.
- Word lists encourage rote memorisation, and students are unmotivated to learn from them; words are not retained after learning.
- There is a lack of published supplementary materials such as exercise books or tests based on word lists; word lists are not related to classwork.
- There are validity and inconsistency issues in word lists construction.
- There is insufficient time and resources to make word lists.

3.3.5 Themes Generated by the Survey

The findings in this study were informed by the different voices of practitioners, researchers and learners from a wide range of ELT domains. The initial aim of this exploration was to obtain an idea about the status of word lists and to understand how they were used and perceived in the field of ELT before designing a tool to evaluate them. This was motivated by the paucity of research on views and practices related to word lists. This section summarises and discusses the main themes generated by the survey. Its purpose is to highlight the themes and issues that informed the development of the proposed tool.

3.3.5.1 Using Word Lists in ELT Pedagogical Contexts. It is evident from the responses that corpus-based word lists are used for various pedagogical purposes across various contexts, albeit mostly to inform language teaching and learning in academic contexts. Validating this finding, Burkett

(2015) observed that language teachers and learners use word lists more than programme coordinators and materials developers. It was found that using word lists for language teaching and learning meant that teachers assigned words for students to learn or work on in the classroom, or to learn independently and then be tested on these words.

Word lists provide a basis for course and programme planning (Nation & Webb, 2011), yet it seems that the course designers are not using word lists to inform course planning (Nation, 2016). Nation (2016) explains this could be due to the large number of words that need to be incorporated into courses. Another factor is related to lack of knowledge of vocabulary acquisition process, frequency and vocabulary control and lack of awareness of existing word lists and how they are developed and how they can be utilised. The current study confirmed Nation's observation that abandoning word lists or misusing them could be due to a lack of awareness and knowledge about them. On the other hand, neither programme coordinators, material developers nor testers who responded to the survey reported not using word lists. This could be attributed to the small sample size, but it might also indicate that word lists are essential for these purposes. However, the extent to which word lists are used and how they are used requires further investigation.

Even though the survey targeted the broader ELT community, the results indicated that corpus-based word lists were mainly used to focus on general academic (50%) and specialised (40%) English at the university level. This could be attributed to the great importance of academic English and the increased flexibility in designing language teaching programmes, relative to school education (where students are taught other subjects and have limited time and short-term learning goals). This finding must be taken with caution as the sample may have been biased towards academics, to whom the researcher had easier access.

The dominance of the AWL is, thus, anticipated. The AVL or other unnamed lists for general academic English were also mentioned, surpassing the number of other types of word lists. It is important to note that the AVL was not mentioned in Burkett (2015), perhaps because the list had

been released relatively recently at the time of his survey. In his 2017 study, Burkett attributed the absence of the AVL in the UAE context to the fact that many published materials, software programmes and tests are already based on the AVL, which was a breakthrough in EGAP (Nation, 2016). Thus, the common use of the AVL could be driven by the factors of what is available and what has been practiced in the past. This also emphasises the importance of the accessibility of resources and materials based on word lists. Accessibility of the word lists itself and availability of resources based on it were mentioned as reasons for selecting word lists.

One of the practices related to word lists reported in the survey was that the practitioners tended to combine lists (mainly published and self-generated lists) to produce a single efficient list. The survey revealed that this was a common practice that reflected the practitioners' dissatisfaction with published word lists. A frequently cited reason was that the participants found that the published word lists were not suitable for learners' needs and proficiency levels. A few responses reported that combining word lists meant deleting words from or adding words to these lists to produce a list that served their purposes and suited the learners. This practice requires further investigation with regard to how word lists are combined and whether the practice is effective. On the other hand, those who used published word lists alone did not modify them despite being dissatisfied with said lists.

The most important finding was that word lists were being used. Due to advances in corpus analysis tools, the field has become flooded with a large number of corpus-driven word lists that serve different pedagogical purposes. However, apart from a few publications that have explored word list use in EAP contexts (Banister, 2016; Burkett, 2015, 2017), little is known about whether word lists are used and, if so, by whom. The findings of this survey revealed that word lists were used widely, although mainly in the field of EAP, which confirmed the importance of designing the evaluation tool.

Derived by the abovementioned observations, the following points were formulated and taken into consideration when developing the proposed tool (see Appendix C):

1. The uses reported in the survey and the literature were taken into consideration. Burkett's (2017) table of available word lists was adapted (see Appendix A).
2. The contexts in which word lists were used as reported by the respondents and the elements that must be considered when evaluating word lists were incorporated into the tool. Those included the type of vocabulary, the characteristics of the target learners and other contextual factors.
3. The fact that practitioners tend to develop their own lists or adapt published lists for the purposes meant that they need to organise information about the target list. The tool includes a table for organising the examined word list's attributes (See Table 2 in Appendix C). It also meant they need guidance on how to make a word list or adapt a published one.
4. Most importantly, the fact that practitioners need to analyse their context of teaching and purpose of using word lists before evaluating and using a word list was reflected in sections 1 and 2 in the tool (Appendix C).
5. Some of the questions in Nation's (2016) framework were deleted or paraphrased to suit the purpose of the proposed tool and to improve accessibility. It was found that some practitioners were not aware of the sources of the resources they used. For example, they were not familiar with concepts such as the unit of counting or corpus design; thus, information about these important concepts was included in accessible language in the tool.

3.3.5.2 Perceptions Related to Using Word Lists in ELT. One of the aims of the current study was to explore practitioners' and learners' attitudes towards word lists and their pedagogic value. Perceptions related to word lists in ELT have been understudied, and the survey revealed positive as well as negative attitudes towards word lists. Exploring these perceptions also revealed the challenges of using word lists and how these could be addressed in this study or in future

research. Generally, practitioners thought that word lists are useful (76.2%) for the various language teaching and learning purposes included in the survey. Even so, they reported a number of problems with word lists.

- Impractically long word lists. Word lists commonly include 1,000 words or more that need to be covered over a period of time
- Lack of word context and definition in word lists
- Lack of supplementary materials based on word lists
- Word lists' lack of relation to classwork
- Encouraging rote memorisation; lack of motivation to learn from lists; lack of retention after learning
- Validity and inconsistency issues associated with word list generation
- Time and resources required to generate or edit word lists

The open-ended comments suggested that respondents found word lists useful for some purposes but not for others. Word lists were reported to be a valuable resource for book writers, lexical-development researchers, or teachers without a clear lexical syllabus on which to focus. Areas in which word lists were seen as being less useful included the independent vocabulary learning of words not covered in the teaching materials. Using word lists in this way was reported to lead to short-term learning; neither the teachers nor the students saw the value of word lists because they were not used or encountered. Nevertheless, the respondents noted some challenges when using word lists. Those points, as detailed below, are in line with those raised by Burkett's (2015; 2017) participants in both of his studies.

- Word lists present isolated items.
- Words in word lists are unrelated to classwork.

- Word lists lack user guides for practitioners and learners and fail to detail the research that they are based on.

Most importantly, the survey captured evidence that word lists were sometimes misused. Teachers misused word lists by giving them to students and asking them to learn the list independently. Nation (2016) warned that word lists are not to be handed to learners but should be used to direct practitioners towards the most important words instead. This practice led to unsuccessful or short-term learning, which would influence how word lists were viewed and valued. The results of the study indicated that many teachers and learners still associate word lists with unsuccessful rote memorisation. This might be attributable to two factors that arose in the comments, namely that words are presented in isolation and that they are not related to the teaching materials.

These issues are similar to those raised by Burkett (2015) and by Sorell (2013). The respondents in Burkett's (2015) study reported that word lists were associated with rote memorisation, while Sorell (2013) mentioned this myth as being one of the problems that may have hindered the widespread use of word lists; however, neither study explored the factors that may have led to this association. Despite Folse's (2004) argument against the myth that learning vocabulary from word lists is useless, this belief persists today. The use of the wrong list or simply differences in learning styles may have contributed to this negative perspective (Folse, 2004). This could be attributed to list-learning, a style of learning a second language from lists by which many teachers were instructed as L2 learners (Schmitt & Schmitt, 2020). However, the growing number of word lists being developed and published is an indication of an interest in word lists.

Bearing this in mind, a secondary aim of the proposed tool was to raise awareness about word lists and how they should be used in ELT. The discovery that many practitioners were not aware of word lists and lacked knowledge about the resources they used led to the adaptation of the wording in the first version of the tool. This also meant that responding to Nation's (2016) questions was

challenging, and confirmed the need to adapt Nation's framework to evaluate the tool based on the purpose for which it is used. Such guidance was provided in the supplementary notes in the first version of the tool (see Appendix C). The problems raised by the respondents and those discussed in the literature, such as the size of word lists, the lack of context and of supplementary materials, were taken into consideration as aspects that affected the usefulness and utilisation of word lists and were addressed in the tool (see Section 4 Implementing Word Lists in Appendix C). These informed the last set of questions pertaining to the implementation of word lists in the tool.

3.3.5.3 Word List Generation. Respondents' reported approaches and opinions to making word lists (see Section 3.3.4.2) align with Nation's (2016) suggestion that to construct a good word list, it is important for experts' judgements to account for the purpose, context and characteristics of the learners with whom the list will be used.

This is an interesting finding that calls for the subjective adjustment of statistically informed lists to ensure that they address the needs of the pedagogical context. In recent years, many published lists have relied heavily on corpus data in the belief that nothing further needed to be considered. This perspective highlights the importance of the teachers' perspective in corpus-based word-list research, a gap that was identified in this phase and in later phases of this study. Dang et al. (2020) published a study after this phase in which they incorporated teachers' perceptions and the learners' knowledge of vocabulary to evaluate two high-frequency word lists. The authors suggested that teachers should trust their intuitive perceptions regarding what was useful for their learners when they selected important words for teaching. The proposed tool aims to encourage practitioners to adapt published lists for their purposes. However, to accomplish this successfully, they must first understand how word lists are compiled and what informs their revisions. Unfortunately, most of the respondents who used published word lists reported using them without modifications. This might have been because the respondents lacked experience in adapting resources to their purposes and contexts, or because they encountered difficulty in managing lists of words. The respondents

generally used the criteria provided by a published researcher, sometimes with suitable modifications, or the criteria provided in their software of choice.

Reviewing the literature, it was found that despite acknowledging the importance of accounting for the purpose and target audience of a word list, few list developers made such considerations in practice (Schmitt, 2016). For example, word-selection criteria and rationales for the unit of counting were not discussed by Browne et al. (2013), possibly because those researchers' professed focus was on developing pedagogical tools that would serve L2 teachers and learners, rather than publishing academic papers for research purposes (Browne, 2016).

In light of this, the tool encouraged practitioners to trust their intuition in what is useful for their students, and adapt their lists accordingly. The tool also takes into account the methodological approach to the examined word list and how this may affect the usefulness of the list for the target purpose and learners.

3.3.5.4 Measures for Selecting a Word List for Pedagogical Uses. The measures practitioner, researchers or learners take to select word lists for use in ELT has not been explored in previous studies. The survey results show that there seems to be an agreement that the applicability of the list to the purpose is strongly associated with the selection and development of word lists. Other measures include:

- discipline-relatedness, students' needs, difficulty level of the words and pedagogical relevance;
- the 'context of the words in a list', a theme that emerged throughout the survey;
- the popularity of the list;
- the availability of the list in electronic form;
- topic-relatedness, which appears to be the most important measure;
- being based on corpus evidence; and

- the suitability of the unit of counting for the target learners.

However, due to dissatisfaction with the available lists, most practitioners prefer to synthesise multiple word lists to achieve their goals. One possible explanation for this is related to the complex nature of pedagogical contexts. As observed by Brezina and Gablasova (2017), learners' characteristics and needs, programmes' goals and the purposes of using a list are unlikely to be the same across different pedagogical contexts, and hence, any list of words purporting to be suitable to all learners would be making a bold claim indeed. The most important motivation for the development of the proposed tool was to guide practitioners through a process of analysing their contexts and purposes for using word lists to enable them to assess the applicability and suitability of the list. Practitioners are encouraged to think about their purposes and contexts when evaluating and implementing a word list using the tool.

It is important to note that many of the findings in the Phase 1 survey regarding the challenges encountered when using word lists and how they could be used, overlapped with the findings reported by Burkett (2017), who investigated the use of frequency lists in EAP in the UAE; however, the researcher was only able to access Burkett's PhD thesis after having collected the data in Phase 1. Burkett's (2017) thesis was only made accessible in the summer of 2018.

3.3.6 A Summary of How Word Lists are Used

In summary, the responses to the survey were preliminary to the understanding of word lists and were used in addition to Nation's (2016) discussion to develop the initial version of the tool. The survey revealed the following:

1. Word lists were used; however, word lists that were not suitable for the target learners or purpose were sometimes used because they had been prescribed or because they were accessible. This indicated that there was a need for an evaluation tool and for improving the implementation of word lists. There was evidence that word lists were used by researchers for

pedagogical research purposes but, at the time of the survey, it was not clear how word lists were used in ELT or by whom.

2. There was a lack of awareness of word lists, how they should be used and their sources. Support for practitioners who lack background knowledge about corpus-based word lists is needed. This gap was not highlighted in the studies that explored the use of word lists, although the gap between corpus linguistics and ELT practice was mentioned.
3. Implementing word lists in educational settings is challenging because there is a lack of support regarding how to use word lists, word lists are long, the words in the lists lack context, and word lists are not related to the classwork or to the students' needs.
4. In terms of how word lists are evaluated or selected, practitioners considered whether supplementary materials were available, the reputation of the publisher of the list, the suitability of the list for their purpose and their students, and access to the word list itself.

Participants from a wide range of domains were recruited to represent the broader ELT community to determine how word lists were used and perceived. Empirical investigations of how different practitioner groups use or select word lists in the literature are limited; furthermore, the influence of word lists on practitioners' applications has been understudied or is not reported in the literature. The data collected in this survey reflected the different uses of word lists and the different criteria used to select word lists. However, the collected data were insufficient to characterise different practitioners in terms of the distinctively different purposes of word lists or the different ways in which word lists were selected.

The limited scope of the survey and the fact that it did not report on the individual roles of the participants means that the findings are not suitable for characterising different groups of practitioners. The aim was to sample respondents to ensure diversity; the researcher did not attempt to control the sample of respondents to represent different practitioner groups to profile practitioners in terms of their approaches to using word lists, but to obtain different voices. This was

confirmed in Phase 3 when the researcher attempted to ask the participants to think of only one reason for using a word list, to which they replied that it was difficult to distinguish the uses of a word list (see Section 3.8.). Burkett (2017) highlighted this point in his PhD work, which, as noted above, was only accessible after Phase 1.

3.4 Phase 2: Design of the Initial Version of the Proposed Evaluation Tool

The researcher developed the proposed evaluation tool based on and guided primarily by (a) the content of Nation's (2016) critique (see Table 2.4 in Section 2.6.5). It was reformulated as a new tool for a different purpose, namely as a tool for assessing the suitability of a word list for use in ELT. This reformulation was informed by (b) the results generated by the initial survey (see the themes in Section 3.3.5), which led to the addition of a set of questions related to the implementation of word lists; and (c) a review of the relevant literature concerning word lists for pedagogical purposes (see Section 3.4.2. Table 3.5 below).

3.4.1 The Adaptation Process

The development of the tool employed Stufflebeam's (2000) guidelines for developing evaluation checklists; Stufflebeam suggested the following 12 steps in checklist development:

1. **Focus on the checklist task.** The first step aimed to define the purpose and the target users of the tool and to study the relevant literature. In this step, the parameters for which the tool will be used were set; that is, the type of word list. The purpose and the target users are described in detail below.
2. **Make a candidate list of checkpoints.** In this step, the tool's questions were drafted largely derived from Nation's framework. More questions and explanations were added derived from the literature and the generated themes from the survey. The main question E in Section 3 *How can the list be implemented in context?* was added (see Appendix C). This question included a set of sub-questions which did not appear in Nation's framework.

3. **Classify and sort the checkpoints.** Sections were numbered and classified. Tables in this thesis are referred to with two digits, and one digit for tables in the tool.
4. **Define and flesh out the categories.** Each section's purpose and how to approach the questions were defined.
5. **Determine the order of categories.** Questions about the corpus from Nation's framework was reordered.
6. **Obtain initial reviews of the checklist.** Outlined in Phase 3 (Section 3.5): Potential users were recruited and interviewed to provide critical feedback on the initial version.
7. **Revise the checklist content.** Following the feedback obtained from the reviewers in Phase 3, the researchers reflected on the issues raised in the interviews and decided on the actions to be taken.
8. **Delineate and format the checklist to serve the intended uses.**
9. **Evaluate the checklist.** Outlined in Phase 4 (Chapter 4): critical feedback survey and field study with practitioners.
10. **Finalise the checklist.**
11. Apply and disseminate the checklist.
12. Periodically review and revise the checklist.

As the proposed tool was based on Nation's framework, the first five steps for establishing evaluation items entailed using the themes from the literature review and the survey to adapt Nation's questions (see Table 3.5 and the section that follows). Supporting notes were added as well. Special emphasis was given to the wording and supportive notes to serve the intended purposes and users. The initial review was obtained in Phase 3 of this study, and the tool's formatting and content were revised accordingly. Then the tool was revised in Phase 4 via two studies and then it was finalised. The last two points 11 and 12 are beyond the scope of the study.

The purpose of the proposed tool is to evaluate the suitability of corpus-based word lists to use for pedagogical purposes in ELT contexts. It is based on the observation that a one-size-fits-all word list does not exist (Nation, 2016). Contextual factors are likely to differ from case to case (Brezina & Gablasova, 2015) and the tool considers these factors in determining the usefulness of a word list and how it can be adapted for maximum utility.

The tool was designed to be accessible to ELT practitioners and researchers with different levels of expertise and knowledge, particularly those who are unfamiliar with the intricacies of developing corpus-based word lists. Broadly speaking, the target users of the tool are ELT practitioners (such as teachers and curriculum or assessment coordinators) and material developers who are involved in directing vocabulary acquisition. These people may range from practitioners who are deeply involved in curriculum development and coordination, who are often in management positions at institutions and who may or may not be involved in the implementation of word lists, to teachers with varying levels of teaching experience who implement word lists in their teaching practice. These individuals may discuss and assist with material development, but they do not have a management role. The tool uses accessible language to summarise the debates in the field in order to help evaluators to consider the most important elements associated with the use of word lists. In early stages of the study learners were included in the target users of the proposed tool, but were limited to practitioners, and learners were eliminated because the researcher considered the process of evaluating word lists to be too complicated; moreover, asking learners to evaluate word lists is challenging, and may not even be an option in some cases.

The tool is meant to be employed before using a list to discern the applicability of that list's usefulness to a target purpose and the measures required to achieve the target objective. It has been designed mainly to evaluate corpus-based word lists developed for pedagogical uses related to English language teaching and learning, but it could also be used for research purposes related to ELT. The tool was designed for the evaluation of word lists containing single-word units and could

also serve as a guide for developing new pedagogical word lists. See the operational definition of word lists in the current empirical work in Section 1.3.

3.4.2 Literature Reviewed to Develop the Proposed Tool

The researcher reviewed the literature related to the development and use of word lists in order to develop the proposed tool. Table 3.5 contains information about the literature that was reviewed to develop each section of the tool. Decisions about the focus of the review were guided by the components of Nation's (2016) framework (see Section 2.6.5) in addition to other issues related to the use of word lists that were derived from the survey (see Section 3.3.5) and which were not addressed in Nation's (2016) framework, namely the 'implementing the list' component. Table 3.5 below summarises how particular publications informed the development of the tool; details regarding how these insights are reflected in the tool are provided in Section 3.4.3.

Table 3.5

Literature Reviewed to Develop the Proposed Tool

Section in the tool	Literature reviewed to inform that section
Section 1: General attributes of the word list Appendix A: Corpus-based Word Lists for English Language Teaching (2000-2020)	The headings in Table 2 (The General Attributes of the Word List) are a replication of the headings in the word list table in Burkett's (2017) work. The word list table (Appendix A) is provided as a supplementary section of the tool that was adapted from Burkett (2017). See Section 3.4.3.1 for more about the content that was adapted from Burkett's work.
Section 2: Context profile	Nation (2016) proposed that a pedagogical word list should be suitable for the target use, the educational context and the learners with whom it will be used. Thus, a section analysing the context of use was added. It consisted of three questions: Uses of word lists (Question A) The purpose for which the list will be used needs to be articulated in order to assess the suitability of a list for such use. Table 3 (Examples of Word Lists in English Language Teaching and Learning Contexts) describes the different uses of word lists for teaching and learning the English language. These are the same purposes that were mentioned in the survey. The examples of each use were adapted from Burkett's (2017) work.

Section in the tool	Literature reviewed to inform that section
	<p>Types of vocabulary (Question B)</p> <p>English vocabulary can be categorised as general English, general academic, and specialised academic vocabulary (Nation, 2016). The category 'professional or occupational vocabulary' is taken from Coxhead (2018).</p> <p>Characteristics of the learners (Question C)</p> <p>The discussion of the factors influencing the creation of word lists in Nation's (2016) work included the learners' ages, proficiency levels and disciplines.</p> <p>The idea that learners with different needs, ages, proficiency levels and disciplines need different word lists was discussed by Brezina and Gablasova (2017) and by Coxhead (2018). These factors were added to the elements used to analyse the educational context.</p>
Section 3: The Evaluation	
<p>Unit of counting</p> <p>Question A: How are words grouped as a unit in the list?</p> <p>Table 4</p>	<p>The idea that different ways of counting words when developing word lists influences the quality and size of word lists means that the usefulness of a word list is affected by the suitability of the unit of counting for the learners and the context (Miller, 2012; Schmitt, 2010).</p> <p>The word family levels (Type, Lemma and Word Family) proposed by Bauer and Nation (1993) were used to classify the main constituents of a word list (the word).</p> <p>Table 4 (The Most Common Word Units Used in Corpus-based Word Lists) presents the definitions and examples of the four word units; it describes the purposes and the audiences that are potentially appropriate for each unit as outlined in the literature, and provides additional notes and an example of a word list for each unit.</p> <p>Type</p> <p>The definition of the type was adapted from Nation's (2016) work. Debates about the suitability of the type for the purpose of vocabulary instruction (for example, productive vocabulary and spelling) and the target vocabulary (such as technical vocabulary), as discussed by Coxhead (2018), Dang and Webb (2016b) and Nation (2016), were summarised. The Applied Linguistics AWL (Khani & Tazik, 2013) was provided as an example of a type-based list.</p> <p>Lemma</p> <p>The description of lemma and example words were adapted from Nation (2016). Reference was made to McLean (2017), who argued that the suitability of the lemma for learners depended on their morphological knowledge. Reference was also made to Nation's (2016) and Schmitt's (2010) discussions of the lemma as a suitable unit of counting for lists of productive vocabulary and for low-proficiency learners. The AVL; Gardner & Davies, 2014) was used as an example of a word list.</p>

Section in the tool	Literature reviewed to inform that section
	<p>Flemma Pinchbeck's (2014) definition of the flemma was used. Reference was made to Dang and Webb (2016b), who discussed the appropriateness of the flemma for learners based on their proficiency levels. The Essential Word List (EWL: Dang & Webb, 2016b) was used as an example of a flemma-based list.</p> <p>Word Family The description of Nation's (2013) word family was adapted. Reference was made to Nation's (2016) argument regarding the suitability of the word family for the development of receptive vocabulary and to Dang, Coxhead and Webb's (2017) argument that the learners' L1 may affect the suitability of the word family. Gardner's (2007) criticism of the wide spread of the word family was mentioned. Schmitt and Zimmerman's (2002) claim against the suitability of the word family for everyone was used to support this argument. The AWL (Coxhead, 2018) was provided as an example.</p>
Question B: What is included in the main list?	<p>The content and guiding notes of Section B in the tool were mainly adapted from the questions in Nation's (2016) work. It was decided to retain this section despite it being mainly related to the construction of word lists, which was not the aim of the proposed tool. As confirmed in Phase 3, this section was irrelevant for the proposed tool; thus, it was deleted.</p>
Question (C): What material (corpus) is the list based on?	<p>Nation (2016) argued that, in order to evaluate a word list, we must assess the corpus from which the list was derived. Factors such as the text type, the age of the learners and the target type of the vocabulary need to be considered when designing a corpus for a word list (Nation & Sorell, 2016), and thus need to be considered when evaluating a word list. Information about the types of texts collected (content) and their representativeness, the size of the corpus, the balance of the corpus, the lengths of the texts and how many corpora there are influence the evaluation of a corpus (Coxhead, 2000).</p>
	<p>Content of the Corpus Notes about the content of the corpus focussed on Biber's (1993) discussion of corpus representativeness, and stated that assessing the suitability of a corpus depended on the learners and on the purpose of analysing the corpus. Nation and Sorell (2016) observed that the nature of the list depended on the source corpus.</p> <p>Informed by factors such as the geographical location, age, and learning situation that were relevant to choosing text types when designing a corpus for creating a word list, as described by Nation and Sorell (2016), Question 17 was added.</p> <p>Representativeness of the corpus Debates about the representativeness of the corpus as discussed in Miller and Biber (2015) were mentioned very briefly.</p>

Section in the tool	Literature reviewed to inform that section
	<p>Size of the corpus</p> <p>Notes related to the size of the corpus that was used to develop the list focused on the idea that assessing the size of the corpus depended on the purpose of the list (Coxhead, 2018; Hunston, 2002). Sinclair's (1991) argument for large corpora to obtain reliable results as opposed to O'Keeffe, McCarthy and Carter's (2007) arguments for a small corpus for specialised purposes were provided as an explanation.</p>
Question D: How are words selected from the corpus and ordered in the list?	<p>Diagram 1 was created to summarise the common criteria for constructing word lists. Examples of qualitative approaches based on corpora were discussed by Coxhead (2018).</p> <p>The notes discussed both the advantages and the disadvantages of these three approaches to making word lists to assist users to assess the appropriateness of an approach in relation to the type of the list being examined.</p> <p>A single criterion or a combination of statistical criteria (frequency, range and dispersion)</p> <p>Gries' (2008) definition of frequency was used to describe frequency. Criticisms regarding the use of frequency to select words were discussed in the notes (Egbert, 2018; Richards, 1970). The definition of range and the advantages of using it to identify words as discussed by Dang, Coxhead and Webb (2017) was summarised. In addition, Nation's (2016) definition of dispersion and how useful it was to identify words was discussed. The limitations of only using statistical criteria to construct pedagogical word lists were then presented (O'Keeffe, McCarthy & Carter, 2007).</p> <p>Statistical results tailored subjectively</p> <p>The idea that there are advantages in adjusting corpus-driven data subjectively to construct word lists was a key idea in evaluating pedagogical word lists in this thesis (Dang & Webb, 2016b; Nation, 2016; Szudarski, 2018).</p> <p>The purely subjective approach to developing word lists</p> <p>Reference was made to Brezina and Gablasova (2017), who described how word lists are developed subjectively.</p> <p>Qualitative/subjective approaches to word lists based on a corpus such as developing word lists for specialised purposes using corpora as discussed in Coxhead (2012; 2018) were also summarised.</p> <p>Disadvantages of the subjective criteria (Schmitt, 2010; Sorell, 2013).</p>
Question E: Implementing word lists	<p>This section was informed by the response of the participants in the Phase 1 survey and the aim of the proposed tool. Some of the complements, such as the size of the list, were mentioned by Burkett (2015).</p> <p>Questions 31 and 33 were adapted from Nation (2016).</p>

In summary, the bulk of the tool's content was based on Nation's (2016) work, but also drew on other research pertaining to the development of word lists and corpus design. These publications

were mainly used to compose the supporting notes, to provide definitions and to restate arguments in the field or research findings that could help the evaluator to respond to the evaluation questions. Details about how and why each source informed the various parts of the tool are described in the following section, Organisation of the Initial Version of the Proposed Tool (Appendix C) under the sub-headings of the main sections of the proposed tool.

3.4.3 Organisation of the Initial Version of the Proposed Tool (Appendix C)

This section describes the proposed tool which for convenience sake is presented in full in Appendix C. The proposed tool comprises three sections. The first section sets out information about the examined list; the second describes the context in which the examined list will be used; the third presents the evaluation questions. The first two sections aim to prepare the evaluator for the evaluation section. This structure was not included in Nation's (2016) original tool and was employed to serve the purpose of the tool and the need to analyse the context before an evaluation (of any type) is conducted (Tomlinson, 2003). Detailed descriptions of how each section was developed from Nation's (2016) questions and what was added based on the findings in the Phase 1 survey and in the review of the relevant literature are presented below. It is important to note that in early stages of this research, the tool was referred to as a framework.

3.4.3.1 Section 1: General Attributes of the Word List. Information about an examined list is necessary to evaluate the suitability of the list for the target audience. After the presentation of the list of questions, Nation (2016) did not pose questions regarding the examined list but, as part of responding to his framework, he provided information about his examined lists, that is BNC/COCA in Table 13.2, within the content of the chapter in which he presented his framework (see p. 133 in Nation, 2016). This was part of critiquing or analysing his lists, but his framework did not pose questions to present such information. In the first section of the proposed tool, information about the list under examination is organised to prepare the evaluator (user) for the evaluation. The evaluator identifies information about the list, including its name, author and year of publication (if

provided), vocabulary type (general English, general academic English and so on), size, unit of counting, corpus details, development criteria, target audience and purpose.

Information about the most popular corpus-based word lists published between 2000 and 2020 (in addition to West's 1953 list) is presented in a table that is provided as part of the tool to assist users to find information about their examined lists (see Appendix A). This corpus-based word list table (Appendix A) was adapted from Burkett (2017). The table consists of 62 corpus-based word lists in total, and contains the following types of vocabulary: general English ($n = 10$), general academic ($n = 8$), general academic for specific disciplines ($n = 19$), and specialised ($n = 23$). If the list under investigation is not included in the summary table, the tool proposes a table (Table 2 in the tool, Appendix C) in which those attributes can be laid out. The headings in Table 2 (The General Attributes of the Word List) are a replication of the headings in the word list table in Burkett's (2017) work.

The purpose of Burkett's (2017) table being presented as an appendix in his PhD thesis was to provide an overview of the lists available in the field as a reference. The table in Appendix A presented in this thesis is designed as part of the tool; thus, practitioners can refer to it when they inspect the specifications of the examined list. The table in Appendix A focusses on recent corpus-based word lists that were published after 2000, while Burkett's (2017) table of 104 word lists included lists from the 1930s, as well as corpus-based word lists and lists developed by combining multiple published lists. Burkett's (2017) comprehensive list of word lists may not be as useful because it presents lists that were published 90 years ago. In this thesis, the word lists that were included in the word list table (Appendix A) were selected from 2000 onwards, with the exception of West's (1953) work. This decision was based on the grounds that information should be up to date and should reflect modern language usage. Moreover, lists that were published between 2017 and 2020 were added as an update.

Burkett (2017) included information about the name of the list, the author and the year of publication, the type and size of the list, the unit of counting, corpus details and notes. Since the purpose of the table in the present study was to assist practitioners to assess the suitability of the list for a given context, I further included a section regarding the target purpose and audience of the lists as claimed by the authors of the lists, together with notes about the criteria that were used to develop the lists. The columns were completed by referring to the published papers that presented these lists. I attempted to include any published assessments of the lists or recommendations in the notes. I divided the lists into four categories, namely general English word lists, general academic word lists, and specialised academic or technical word lists. Within each category, the word lists were presented chronologically according to the publication date. Burkett's (2017, pp. 276-280) list was arranged alphabetically.

3.4.3.2 Section 2: The Context Profile. This section identifies the purpose of using the examined list and the context in which it will be used. Analysing the context is crucial for determining the potential usefulness of the list and should guide the user to the necessary decisions when implementing the list. This section consists of three main questions. The first explores the purpose of using the list; the second pertains to the target vocabulary type. The third question aims to describe the characteristics of the target audience, namely, age range, proficiency levels and academic discipline (if decided).

The first two questions in Nation's (2016, p. 131) framework were 'Was the target population for the word list clearly described?' and 'Was the purpose of the list clearly described?' Reflecting on these two questions, Nation wrote, 'This is poorly done. There is no clear description of the people who will learn from the lists and only a very vague reference to the uses of the lists' (p. 140). The clarity of these descriptions was essential for assessing the lists. However, this was executed poorly not only by Nation, but also in many papers that presented word lists. Therefore, these points must be specified and assessed. These two questions were reformulated in the proposed tool to assist

users to describe the purpose of the word lists and the characteristics of the learners with whom the list will be used. In the proposed tool, I added questions to help to describe the target learners and the factors that should be considered when evaluating word lists. I also added a question regarding the type of vocabulary. The classification of these types was informed by Nation (2016) and Coxhead (2018).

Question A required the user to select one reason for using the word list being examined. The options that were provided were the course/programme design, teaching and learning, testing (assessment), material development, research (related to one of the above purposes) and/or other use. These were the same purposes/uses that were presented in the survey in Phase 1. Burkett (2017, p. 195) adapted Nation's (2016) description of the uses of word lists, and provided a table of examples of word lists used in teaching and learning the English language. This was adapted as Table 3 in the tool to help users to understand the meaning of each use.

Question B inquired about the vocabulary type that was targeted in the educational context, while Question C aimed to identify the learners' characteristics. One of these was the age range, which Nation (2016) suggested as a factor that should be considered when selecting a corpus or designing a word list. Lists for children should be based on the language that the children will be likely to encounter and use. For example, swear words and inappropriate words need to be removed or moved to a marginal list. The proficiency level of the learners was also included because it was one of the factors that influenced the usefulness of a word list, as mentioned in the survey and as highlighted by Brezina and Gablasova (2017), Nation (2016) and Coxhead (2018). Lastly, the factor of discipline was adapted from Coxhead's (2018) work as a factor that needs to be considered when developing a word list even a general purpose one.

3.4.3.3 Section 3: Evaluation Section. This section is the central part of the tool. It includes 33 evaluation questions adapted from Nation's framework and categorised according to five groups, except for the last group (Pedagogical Considerations), which was informed by the survey findings.

1. Word unit
2. Main word list
3. Corpus
4. Criteria for word list development
5. Pedagogical considerations.

For each question, the evaluator chooses the response that best describes the word list. Responses are given on a Likert scale from 1 to 3, where 1 indicates 'not at all', 2 indicates 'somehow', 3 indicates 'most definitely' and 'N/A' is an option if a question is not applicable or the best answer is not known. For some questions, a yes-or-no response is required instead.

Where relevant, indicative answers and explanations of technical terms are provided (the yellow boxes in the tool, see Appendix C). These were largely informed by the relevant literature (see Table 3.5) and could be useful for practitioners who lack experience with or knowledge of corpus tools. Therefore, a few items in Nation's (2016) work were modified to be more accessible, a few were deleted and others were added to suit the purpose of the tool. A detailed description of this procedure is presented below in Tables 3.6–3.10. The left column presents the original questions posed by Nation (2016), and the number of the equivalent question is given in parentheses for questions that were adapted. The right column presents the adapted questions, as well as the new questions that were added to this tool. Each section aims to answer an overarching question (Questions A–E) by addressing it via sub-questions.

Section 3: Question A: How are words grouped as a unit in the list?

The term *unit of counting* is a technical term, and was replaced by the term *word unit* to improve the accessibility of the tool. The user needs to identify the word unit of a list, understand

what it means and then answer the evaluative questions 1-5 accordingly. A table of the four common word units (Table 4 in the tool; see Appendix C) is provided after the questions to help non-expert users. The table defines each unit of counting and gives an example. It also provides purposes and audiences potentially appropriate to each unit, as outlined in the literature, and offers additional notes and an example of a word list based on each unit. In the proposed tool, Nation's original questions were reproduced with simple modifications to their wording to clarify the intended meaning; see Table 3.6.

Table 3.6

Changes to Nation's Questions Related to the Unit of Counting

Nation's questions	Adapted questions
<ul style="list-style-type: none"> • Was the unit of counting suitable to the purpose? (q1) • Was the unit of counting clearly defined, including issues such the UK and US spelling, alternative spellings, part of speech, abbreviations and numbers? (q3) • Was the unit of counting explicitly well-justified? (q4) 	<ol style="list-style-type: none"> 1. Is the word unit (<i>Type, Lemma, Flemma, or Word Family</i>) of the word list suitable to your purpose? 2. Is the word unit of the word list suitable for the proficiency level of your audience? 3. Are issues related to the definition of the "word" unit such the UK and US spelling, alternative spellings, part of speech, abbreviations and numbers dealt with appropriately to your context? 4. If you think the word unit is not appropriate for your purpose and/or audience, is it well-justified? 5. Are there different versions of the list with different definitions of the "word"?

Two questions (2 and 5 in the adapted questions) were added. Question 2 '*Is the word unit of the word list suitable for the proficiency level of your audience?*' was added as it was indicated in the literature (e.g. Stoeckel et al., 2018) that the suitability of the unit of counting to the proficiency level of the learners is an important factor that needs to be considered. In the survey, two respondents explained that they had selected lists based on the unit of counting that was suitable for their learners emphasising the observation that the unit of counting is one of the elements that affect the suitability of the list (see Section 3.3.4.1). Question 4: *If you think the word unit is not appropriate for*

your purpose and/or audience, is it well-justified? This was reformulated based on Nation's (2016) question regarding the justification for using the unit of counting. This question was informed by responses from the survey that were related to the unsuitability of the unit of counting for word lists in many educational contexts. Guidance concerning how to adapt and resolve this issue was provided in subsequent iterations of the tool. Question 5 '*Are there different versions of the list with different definitions of the "word"?*' was added as a solution to the problem of unsuitable units of counting in many of the published word lists. In a later publication, Dang et al. (2020) called for the construction of lists with different units of counting in response to the realisation that one list that suits all purposes does not exist. Notes on Questions 4 and 5 were added as they were Yes/No questions that could not be answered in a straightforward manner.

Question B: What is included in the main list?

Table 3.7 shows the changes to the questions which were concerned with the wording of the items and emphasising evaluating each component in accordance with the intended purpose. Notes on how to approach those questions were added. Question 14 belongs to 'Other lists' section in Nation's framework. As noted in Table 3.5, this section was irrelevant, and evaluating word lists for pedagogical use was challenging; nonetheless, it was retained in the initial version because these questions were posed by Nation (2016), but these questions were deleted in subsequent iterations following the reviewers' suggestions in Phase 3.

Table 3.7*Changes to Nation's Questions Related to the Main List*

Nation's questions	Adapted questions
<ul style="list-style-type: none"> • Was there an explicit description of what would be counted as words and what would not be included? (q6) • Were homographs dealt with? (q7) • Were proper names dealt with, including proper name homoforms? (q8) • Were content bearing proper names distinguished? (q9) • Were hyphenated words dealt with? (q10) • Were transparent compounds dealt with in a way consistent with hyphenated words? (q11) • Were acronyms dealt with, including acronym homoforms? (q12) • Were the proper name lists and other lists revised on the basis of initial output? (q13) • Were marginal words dealt with? (q14) 	<ol style="list-style-type: none"> 6. Is the description of what forms are counted as words suitable to the purpose? 7. Are homonyms and homographs dealt with according to the purpose? 8. Are proper names dealt with, including proper name homoforms, according to the purpose? 9. Are content bearing proper names distinguished? 10. Are hyphenated words dealt with in a consistent well-justified way? 11. Are transparent compounds dealt with in a way consistent with hyphenated words? 12. Are acronyms dealt with, including acronym homoforms? 13. Are the proper name lists and other lists revised on the basis of initial output? 14. How satisfactorily are marginal words dealt with? (from Other lists section in the original framework)

Question C: What material (corpus) is the list based on?

The quality of the list depends on the corpus from which it is derived. Therefore, it is important to evaluate that corpus (Nation & Sorell, 2016). This section explains corpus-related information important to list evaluation, such as the kind of texts collected, their representativeness, the size and balance of the corpus, the length of texts and the number of corpora used (Coxhead, 2000). It is possible for the evaluator to skip this section because access to such information may not be available to them, but doing so may affect the usefulness of their results.

Two questions were added (16 and 17, reproduced in Table 3.8), the former addressing the suitability of the content to the characteristics of the audience and latter addressing situational factors. The question ‘*Were the sub-corpora large enough, of equal size and coherent?*’ in Nation’s

(2016) framework was divided into three questions (20, 21 and 22). Question 18 in the original tool was deleted as it was related to corpus construction, not word list use. A question about other factors was also added (see Table 3.8). Notes about the contents and sizes of corpora for word lists are provided at the end of the section (see Appendix C).

Question 17 was derived from Nation and Sorell's (2016) recommendation that relevant factors (such as geographical location, age and learning situation) should be considered when choosing text types in designing a corpus for the creation of word lists, and should thus be considered when evaluating word lists. Nation's (2016) statement was referenced regarding age, while the geographical and cultural considerations were derived from the comments in the survey.

Table 3.8

Changes to Nation's Questions Related to the Corpus

Nation's questions	Adapted questions
<ul style="list-style-type: none"> Was the content of the corpus suitable to the purpose of the list? (q15) Was the corpus large enough to get a reliable result? (q18) Was the corpus divided into sub-corpora so range and dispersion could be measured? (q19) Were the sub-corpora large enough, of equal size and coherent? (q20, 21, 22) Was the corpus checked for errors? 	<p>15. Does the <i>content</i> of the corpus represent the actual or potential language your audience is likely to be exposed to?</p> <p>16. Is the <i>content</i> of the corpus suitable to the proficiency level and age of your audience?</p> <p>17. Are factors like text type, geographical division, the age of the audience and learning situation considered according to your purpose?</p> <p>18. Is the <i>size</i> of the corpus suitable to get reliable results for the intended purpose?</p> <p>19. Is the corpus divided into sub-corpora? If yes,</p> <p>20. Are the sub-corpora large enough?</p> <p>21. Are the sub-corpora of equal size?</p> <p>22. Are the sub-corpora coherent?</p>

Question D: How are words selected from the corpus and ordered in the list?

In this section, the user must first identify the approach used to develop the lists and then evaluate its suitability. A diagram presenting the most common approaches is provided (a yellow box in Appendix C), followed by a discussion of the advantages and disadvantages of each. The original questions address whether the approach was described; thus, in the proposed version, the

questions address the suitability of the approach, see Table 3.9. Question 27 was adapted from 'Other lists' section of the original tool.

Table 3.9

Changes to Nation's Questions Related to Making the List

Nation's questions	Adapted questions
<ul style="list-style-type: none"> • Were the criteria for inclusion and ordering in the list (frequency, range, dispersion, or some composite measure) clearly described and justified? (q23) • Were the criteria for making the sub-list clearly described and justified? (q25) • Were any subjective criteria used? Were they described and justified? (q24) • Were the lists checked against competing lists not just for coverage but also for overlapping and non-overlapping words? (q26) • Were any supplementary lists used? (q27) 	<p>23. How suitable are the criteria for inclusion and ordering in the list (e.g., frequency, range, dispersion, or some composite measure) for getting reliable results?</p> <p>24. If subjective criteria are used, how well are they formalized and applied?</p> <p>25. Are the criteria for making the sub-list clearly described and justified, if any?</p> <p>26. Are the lists checked against competing lists for coverage and overlapping and non-overlapping words?</p> <p>27. Are other supplementary lists used to exclude or add items from the list? How? (Other lists section)</p>

Question E: How can the list be implemented in context?

Nation's tool does not contain any section focussing on implementing the list, but it does feature two questions related to the availability and criticism of the assessed word list. The questions added to the proposed tool were based mainly on the comments raised in the survey, see Table 3.10.

Question 28 regarding the size of the examined word list was informed by data collected via the survey in Phase 1. The participants reported that the large size of word lists was an issue that influenced their utility. This issue was also mentioned in the literature (e.g. Banister, 2016) and later by Dang (2020) as one of the issues in the use of word lists. Managing the content of the list in the language programme or course is an important factor that needs to be taken into consideration.

Table 3.10*Changes to Nation's Questions Related to Implementing the List*

Nation's questions	Adapted questions
<ul style="list-style-type: none"> • Availability: Are the lists readily available in electronic form for evaluation? (q31) • Self-criticism: Are the weaknesses of the lists are clearly acknowledged? (q33) 	<p>28. Is the size of the list manageable for your intended purpose? Is the context for each word provided (e.g., through concordance lines or example sentences)?</p> <p>30. Are there supplementary materials based on the list to be used in class?</p> <p>31. Do you have access to an electronic version of the list so that you can edit it according to your purpose?</p> <p>32. Is there guidance on how to use the list?</p> <p>33. Are the weaknesses of the lists acknowledged and justified?</p>

Question 29 pertaining to the context of the words was also informed by responses collected from the participants in Phase 1. The lack of context for each word and 'learning contextualised vocabulary' was mentioned in the survey. It was reported that word lists were not related to the teaching materials in some contexts. This is a factor that is important, but the context provided is not always suited to the learner's needs. The type of context provided in authentic language data, such as corpus examples, might not be suitable for the proficiency level of the target learners or their needs, which means that it might be more useful for the teacher to provide the context. A comment was made in the notes related to evaluating the suitability of the context. Practitioners need to ensure that, if context is provided with a word, it needs to be relevant.

Questions 30 and 32 pertained to the availability of supplementary materials to be used with the list and guidance concerning how to use word lists were derived from the survey findings. Support for the end user of the word list is one of the things that a list developer needs to consider when developing a pedagogical list. Recently, such lists have recently been developed.

3.4.4 *How is Nation's (2016) Framework Different from the Proposed Tool?*

The proposed tool differs from Nation's (2016) framework in terms of the purpose, the target users and the process of development, as well as with regard to organisation and content.

The purpose of Nation's (2016) framework was to critique a word list, but the purpose of the actual critique was not stated clearly. His framework was presented within a section about making word lists, in which he stated that the process of evaluating the lists and writing the book assisted in the revision of his list. Thus, it could be said that the purpose of his framework was to evaluate a word list as a step towards creating one. Reading the critique of his lists is clearly of benefit to anyone who is planning to use the BNC/COCA lists. However, if the questions were presented in isolation, or if the user was planning to use another list, it is not clear how usable this framework would be for a practitioner who lacks knowledge and expertise about vocabulary instruction and corpus linguistics. The purpose of the proposed tool overlaps with Nation's intention in that both recognise that word lists need to be evaluated, but the focus of the evaluation in the proposed tool is on use, and it was designed specifically with the non-expert user in mind. The language is accessible for regular language practitioners, and the purpose is to assess the suitability of corpus-based word lists for use in ELT with the target purpose and context in mind.

Accordingly, the target users of the tool are conceived of as practitioners; specifically, not as people who are involved in corpus linguistics or the development of word lists. Guidance and notes regarding how to respond to the questions are provided and proposed answers to some questions are presented. Supplementary guidance is provided in the proposed tool. By contrast, responding to the questions in Nation's (2016) work requires prior knowledge of corpus-based word lists. In fact, in the introduction to his book, Nation stated that his 'book is not written for those who are new to the study of vocabulary. It is aimed at those who know something about the teaching and learning of vocabulary and who want to make use of word lists in an informed way or who wish to create their own word lists for particular purposes' (p. X).

In terms of the process of development, Nation (2016) stated that his framework was based on the previous chapter of his book which, in turn, means that the content was based on Nation's extensive work on word lists, vocabulary instruction and his students' contributions. By contrast, the development of the new tool in this study involved adapting Nation's questions based on observations from the literature and data collected from practitioners; this was followed by reviewing and revising the tool in different iterations based on consultations with experts and practitioners in the development and use of corpus-based word lists (Phases 3 and 4).

As for the organisation and content, a detailed description of how the proposed tool was developed from Nation's (2016) questions, as well as how the survey data and the literature review informed the content of the proposed tool, is provided in Section 3.4.

3.5 Phase 3: Initial Review of the Proposed Framework

3.5.1 Aim

This phase aimed to have the proposed tool reviewed across three criteria— effectiveness, content and efficiency—in semi-structured interviews with a panel of experts and potential users. Those criteria focused on: the usefulness of the tool, completeness and relevance of the content, clarity of wording and accessibility of the content to non-expert users, and the organisation of the tool. In the context of this study, the '**expert users**' are experts in corpus linguistics and word lists development while the '**potential users**' include ELT practitioners who do not necessarily have the expertise and knowledge about corpus-based word lists.

3.5.2 Guiding Questions

This review was guided by the following research questions:

RQ5: To what extent does the proposed tool meet its intended purpose, according to expert users and evaluators of corpus-based word lists?

RQ6: To what extent does the proposed tool meet its intended purpose, according to potential users of the tool (e.g. language teachers)?

3.5.3 *Participants and Their Contexts*

Phase 3 involved a specialised sample of individuals purposefully selected to review and discuss the tool in light of their experience and knowledge. Purposive sampling is a non-probability sampling technique common in qualitative studies when the key concept under investigation requires in-depth exploration and participation is demanding (Cohen et al., 2018). The researcher intended to interview 10 participants with varying levels of experience in the development and use of corpus-based word lists and ESL. Finally, six participants agreed to be interviewed; the researcher selected the participants because she wanted a small subset to apply their particular knowledge to the development of the tool. However, the review process and meeting the researcher was time consuming. The estimated time taken to review the tool was 50 minutes. As the reviews began to become repetitive and helpful, the planned changes were substantial and no more participants agreed to be interviewed; hence, the researcher decided that this small number of participants was sufficient for an initial review.

The same inclusion criteria that were employed in the first phase were applied here: participants would have to be aged 18 or older, be involved in ELT and have used word lists for language teaching and learning. The individuals contacted for this phase were language teachers, curriculum developers, materials and assessment developers, and researchers who had experience with word lists (individual participants may have had more than one role). Table 3.11 briefly describes the contexts and credentials of the six reviewers.

Table 3.11*Reviewers Profiles*

Reviewer	Description	Length of interview
Researcher 1 (R1)	An early career researcher in word list use and development in an English-speaking country; non-native.	25 min
Researcher 2 (R2)	A researcher working on corpus linguistics, ELT and vocabulary instruction in Saudi.	36 min
Expert 1 (E1)	An expert with more than two decades of experience in word list development and use in an English-speaking country.	45 min
Expert 2 (E2)	A researcher with a decade of experience working in word list development and use, as well as an expert, native-English-speaking language teacher in the Middle East.	54 min
Teacher 1 (T1)	A language teacher and early career researcher in Saudi.	42 min
Teacher 2 (T2)	A language teacher and PhD student in the UK.	64 min

This phase focussed on expert practitioners and researchers—that is, individuals with experience in the development and use of corpus-based word lists. They were selected because of their awareness of the characteristics of a useful word list, and they helped the researcher to improve the quality of the proposed tool. The expert practitioners and researchers, R1, R2, E1 and E2, had experience in developing, evaluating and using word lists for language teaching and learning. The criterion for determining that E1 and E2 were experts was that they had theoretical and practical knowledge of corpus-based word lists. Practical knowledge means that they had developed word lists, and their publications proved this. The participants' status and experience were confirmed at the beginning of the interviews (see the professional background interview questions in Appendix D). Researchers R1 and R2, who were recruited in this phase, had theoretical knowledge that involved working on corpus-based word lists but not necessarily constructing them, and not necessarily having published works about word lists, corpus linguistics or any aspect of corpus-based word lists. Their status was confirmed at the beginning of the interviews (see the questions in Appendix D). The study also targeted ELT practitioners who had used corpus-based word lists for language teaching and

learning purposes but who did not necessarily have experience with corpus tools and word list development. Their expertise with word lists and background knowledge was also confirmed at the beginning of the interview. This group, T1 and T2, contributed by assessing the accessibility of the tool and reflecting on the realities of word list use.

Sixty email invitations were sent in total; 30 of the addressees were identified by surveying the names of authors on published word lists, through personal connections and via the snowball-sampling technique. A further 30 emails were sent to participants from Phase 1 who had agreed to participate in a further phase of the research. In total, 15 of the 60 contacted participants responded (a response rate of 25%). Eleven respondents accepted this invitation and were required to sign another consent form; once they had done so, they were sent the document and information about the review process. Five of the 11 participants who consented emailed their comments, but could not participate in the interviews due to their busy schedules: Two were experts in the development and use of corpus-based word lists, while three were ELT practitioners (two in Saudi Arabia and one in the UK). Another four of the 15 who responded to the invitation email declined to participate; they indicated that the task was too demanding or felt they lacked the appropriate experience. Finally, six participants—two of the 30 respondents from the previous survey and four who were recruited solely for this step—agreed to be interviewed. Their expertise and professional backgrounds were confirmed at the beginning of the interviews.

3.5.4 Data Collection

3.5.4.1 Semi-Structured Interviews. In this step, the researcher conducted semi-structured interviews as the primary means of data collection (in addition to making contact by email, where necessary, before and after the study).

A semi-structured interview is a compromise between structured and unstructured interviews (Dörnyei, 2007). Like a structured interview, the interviewer directs a semi-structured interview towards pre-determined topics, but they also maintain adequate flexibility in terms of

allowing the respondent to digress in unexpected directions, enriching the conversation (Dörnyei, 2007). This type of interview proved useful because it allowed the respondents to elaborate on certain issues related to the tool while still collecting data that were relevant and focussed in terms of the aforementioned evaluation criteria. One-to-one interviews were necessary because the reviewers possessed varying levels of expertise and differing opinions were elicited in the process. Such depth would have been difficult to achieve by using questionnaires, quantitative methods or focus-group interviews.

3.5.4.2 Interview Protocol. An interview protocol (see Appendix D) was used to systematise the interviews across all respondents and ensure the comprehensiveness of the topics discussed (Dörnyei, 2007). The interview was divided into six main parts. It covered the four main evaluation criteria, each of which was linked to probe questions that were relevant to the evaluation criteria. Apart from the introductory questions, neither the order of the questions nor their wording were controlled.

- **Introduction:** The interviews began with the researcher asking the respondent about their background in using and evaluating word lists. These factual, easy-to-answer questions were posed to break the ice and create a relaxed atmosphere, as well as to identify the respondent's level of expertise.

The remainder of the interviews aimed to determine whether the tool was a valid tool in terms of the following parameters:

- **Clarity of wording** of the questions, indicative answers and response scale; consistency of terminology throughout the document.
- **Completeness and relevance** of the tool. This was the core element to be judged. The reviewers were asked about the usefulness of the tool and its applicability to different types of word lists, purposes and contexts.

- **Accessibility** of the content to non-expert users. Of particular concern was the accessibility of the guidelines and interpretations.
- **Practicality and organisation** in terms of length, logical order, style and format.
- **Closing questions:** The reviewers were asked whether they would like to change or add anything to improve the tool.

3.5.4.3 Email Review. Three (out of six) participants sent their comments to the researcher before the interview (which still took place). Five participants (out of the 60 potential participants) sent their comments but could not participate in interviews because of their busy schedules. Those emails were included in the analysis.

3.5.5 Procedure

Invitation emails were sent to 60 potential participants. The invitation emails included a link to a QualtricsXM page that explained the purpose and procedure of the study. Consent forms and contact details were collected online, either through the QualtricsXM link or by email. After a participant consented, they were emailed the proposed tool along with instructions for its review; alternatively, they were granted access to the documents and instructions via QualtricsXM. Subsequently, an interview was scheduled at the participant's convenience. Participants had the option of being interviewed online or face-to-face for reasons such as convenience or physical distance. One week after receiving consents, a follow-up email was sent to participants who had provided consent but did not schedule an interview.

Out of the 60 contacted individuals, six agreed to be interviewed, see Section 3.5.3. Each interview was audio recorded with the participant's permission. The interview durations ranged from 25 min to 64 min, with an average duration of 44 min (see Table 3.11). The cumulative duration of all six interviews was 4 h and 43.3 min. The word counts of the interview transcriptions ranged from 1,995 to 6,050 words, with an average of 3,619 words and a total of 21,715 across all interviews.

3.5.6 Thematic Data Analysis

Thematic analysis was employed to analyse the interview data. Braun and Clarke (2006, p. 6) defined thematic analysis as ‘a method for identifying, analysing, and reporting patterns (themes) within data’, and they identified two approaches to such analysis. The first is bottom-up or inductive analysis, in which codes are derived from the data. The second is top-down or deductive thematic analysis, in which codes are derived from the research questions and/or the analyst’s interests. The approach employed to analyse data in the current study can be categorised as primarily deductive (theoretical thematic) as it relied on pre-decided codes based on the evaluation criteria. Other categories were derived during the analysis of the data.

The goal of thematic analysis in the current study was to identify themes, i.e. patterns in the data about the tool that are important or interesting, and then to use these themes to improve the tool. Braun and Clarke (2006) provided a six-phase framework for doing a thematic analysis.

Step 1: Become familiar with the data.

Step 2: Generate initial codes.

Step 3: Search for themes.

Step 4: Review themes.

Step 5: Define themes.

Step 6: Write the report.

The process of data analysis, guided by this framework, is described below:

Step 1: Become familiar with the data. The first step was to prepare the data for analysis by transcribing the audio recorded interviews into texts. During transcription, the researcher checked the transcripts for accuracy while becoming familiar with the data. The process of transcription can be seen as a way of becoming familiar with the data (Braun & Clarke, 2006). The next step involved reading and re-reading the data, including the emails, to obtain an overall understanding of the content. Data related to reviewers’ experience and backgrounds were extracted and summarised in

the profiles of the reviewers because those data were not necessary for the analysis. Each reviewer was assigned a title and number—e.g. the first reviewer, who was a researcher, was assigned 'R1', see Table 3.11. Data that were not relevant to the research questions, such as personal anecdotes and off-topic comments, were not coded.

Step 2: Generating initial codes. The remaining data were broken down into chunks based on the interview questions. Notably, a few of the interview questions were not addressed (see Appendix D), and other points of focus emerged. Those that were related to the research but not the research questions also were coded. The researcher began by reading the transcripts, highlighting with different colours the passages that captured the key elements of the review and writing as many codes as applicable next to the text. Responses (interview answers) that covered two codes were given two codes.

Step 3: Searching for themes. This step aimed to compress the texts into smaller sets of data under fewer themes. A theme 'captures something important about the data in relation to the research question and represents some level of patterned response or meaning within the data set' (Braun & Clarke, 2006, p. 10). In this step, the researcher examined the codes and those which fit together were grouped into a broader theme. For example, codes related to the wording of the items—such as revise technical term and reword—were collated into an initial theme called 'word choice in the questions, notes and instructions'. These themes focused on six areas: the clarity of the wording, the relevance of the content, the usefulness of the tool, the accessibility of the tool to non-expert users, the practicality of the tool and the organisation of the tool.

Step 4: Review themes. This step involved revising the themes derived from the analysed data. When data could not be coded with the existing themes, new codes and themes were introduced.

In **Step 5, the themes were defined.** The aim was to 'identify the "essence" of what each theme is about' (Braun & Clarke, 2006, p. 22). See Appendix E for a full list of the codes and themes

and their definitions. Finally, **Step 6** was **the write-up**, in which a summary of the changes and issues discussed were reported. The findings are presented in Section 3.7 following a discussion of trustworthiness.

3.6 Achieving Trustworthiness

The quality of data in mixed-methods studies is determined by the quality of the quantitative and qualitative phases. Quantitative data are evaluated in terms of validity and reliability of the data and the measurement tools, while qualitative data are evaluated in terms of trustworthiness (Graff, 2016). The aim of trustworthiness is to ensure that findings are 'worth paying attention to' (Lincoln & Guba, 1985, p. 290). Lincoln and Guba's (1994) criteria for achieving trustworthiness are widely accepted by qualitative researchers. Those criteria include credibility, transferability, dependability, confirmability, with authenticity added later (Guba & Lincoln, 1994). Briefly, credibility can be defined as confidence in the 'truth' of findings. Transferability means that the findings are applicable in other contexts, while dependability indicates that the findings are consistent and could be replicated. Confirmability describes the degree of neutrality, or in other words, the extent to which the findings of a study reflect the respondents' opinions and experiences rather than the researchers' biases, motivations or interests. Authenticity concerns the ability of researchers to accurately depict the diverse realities portrayed by data collected from participants.

It is important to note that not all procedures are used in each study. The credibility of this study was enhanced by triangulation. Data were collected using different methods, from different types of participants and different ELT contexts. In Phase 1, open-ended responses were collected along with the quantitative data. In Phases 3 and 4 (see Chapter 4), participants with different levels of expertise were interviewed and in Phase 4 two types of studies were conducted.

To enhance the confirmability or neutrality of the findings and to reduce the researcher's bias, member checking was employed. After analysing the interview transcripts in Phases 3 and 4 (Study 2), the collected feedback was summarised as key points which, along with the planned

actions to address these points, were sent individually to each participant to ensure the accuracy of the identified concepts/for their validation of the researcher's interpretation. The emails included a brief description of each reviewer's role and context. Reviewers were asked to elaborate on certain points; when their suggestions were not accepted, an explanation of the reason was provided. Five of the six recipients replied in Phase 3, while nine of the 11 recipients replied in Phase 4. Two follow-up interviews were conducted in Phase 4 to discuss the participants' feedback in more detail.

To establish trustworthiness and credibility, the researcher attempted to report the procedures undertaken and the steps in the study in sufficient detail. In addition, the researcher attempted to include as many of the respondents' direct quotes as possible instead paraphrasing them. Presenting sufficient descriptive data that can be read and which will help to support the findings and by placing it in the context added to the plausibility and credibility of the research.

3.7 Research Ethics and the Confidentiality of the Participants

Ethical approval was obtained from the university's Research Ethics Committee for each phase of this project. Initial ethical approval was obtained for the first explorative phase, Phase 1. After collecting and analysing the data from Phase 1 and preparing the initial version of the evaluation tool (Phase 2), Phase 3 was designed, and ethical approval was obtained. After collecting and analysing the data, Phase 4 was designed, and another ethical approval was obtained (see Section 4.6.). The reason for obtaining separate ethical approval for each phase was the uncertain nature of the study and the results; I needed to collect feedback, which was then used to plan the subsequent phase.

The participants in all three phases were contacted via email. The student's university email (Outlook) was used for recruitment and communication purposes. The participants were not asked to provide responses via email communication unless they chose to do so. The recruitment emails explained the purpose and procedure of each study, how the data were managed, stored and disposed of, and included a link to each phase's page at QualtricsXM (2018). Qualtrics is compliant

with the General Data Protection Regulations (GDPR). The University of Birmingham has an institutional subscription to the service.

Participation in all three phases was entirely voluntary. The participants had the right to refuse participation, to refuse to answer any question and to withdraw from the project (prior to the deadline) at any time without providing an explanation or experiencing consequences. The researcher's and supervisors' contact details were provided in case the participants had any questions about the research or their participation. The recruitment email also invited the recipients to circulate the invitation to other potential respondents who met the inclusion criteria and who might have been interested in the study.

The QualtricsXM page for Phase 1 contained detailed information about the research followed by a consent form (see Appendix B). Appendix B includes the participant information form, the consent form and the questionnaire questions in the order in which they were presented to the respondents. The participants were required to indicate their consent before beginning the questionnaire by ticking a box. The questions would not appear if the respondents did not consent, and only one respondent did not agree. Upon receiving their agreement, the participants were asked to provide their email addresses; the email addresses were collected to identify the participants if they agreed to be contacted for participation in subsequent steps of the study, requested a summary of the findings or decided to withdraw their data from the survey. The collection of email addresses also allowed the researcher to remove multiple submissions by the same respondent. No personal data other than the participants' email address were collected. The participants' anonymity could not be ensured because some of the participants' identities could be identified based on their email addresses; thus, the participants' confidentiality was ensured by not analysing their email addresses or publishing them as part of the research. A list of the email addresses was stored on a passwordprotected file on BEAR DataShare. Only the supervisors and the researcher have access to

this file, which will be kept until the end of the PhD programme, when the email addresses that were used to contact the participants will be deleted from the BEAR DataShare and from the student's university email programme (Outlook). The participants' contact details were not stored anywhere other than on the University of Birmingham's servers. Both the BEAR DataShare database and the student's university email are stored on the University of Birmingham's server. Compensation for participation was not offered, but a summary of the findings was sent to the participants who requested a summary.

In Phase 3, the invitation emails included a link to the study page at QualtricsXM with a detailed participants' information sheet and a consent form attached (see Appendix B). The QualtricsXM page for Phase 3 explained the purpose and procedure of the reviews and interviews, and included a consent form with a box to tick if they agreed to participate. A few participants consented by replying to the researcher's email. Upon receiving their consent, they were emailed the proposed tool together with instructions for its review; alternatively, they were granted access to the documents and instructions via QualtricsXM if they consented via the link. Given that they contributed to the development of the tool, participants who requested a copy of the evaluation tool received the final version. Contact details were not published as part of the research and were only used for communication regarding this project. The procedures used in Phase 1 to manage, store and dispose of the participants' contact details and responses were also used in this phase.

The interviews were audio recorded using a digital voice recorder for the face-to-face interviews and the recording function on Skype for the Skype interviews, with the participants' consent having been obtained prior to the interviews. A Skype account was created for the research study; this account was deleted at the end of the study and all the participants' online details and data were removed. The researcher took notes during the interviews. The participants' exact words were transcribed, while non-lexical sounds and expressions of emotions and cut/off repeated words

were omitted. The participants were informed that they could request the transcripts to correct any factual errors if they wished to do so, but none of the participants requested the transcripts. The transcriptions of the audio files were saved in the BEAR DataShare provided by the University of Birmingham immediately after the interviews, while the original audio files were deleted immediately. The information provided during the interviews and the notes on the copies were used solely for the purpose of this research project; the data were kept confidential and were anonymised in any publications. Only the researcher and her supervisors had access to the interview transcripts and the participants' notes. Upon completion of the research, all the transcriptions and notes will be stored securely for ten years in compliance with the University of Birmingham's policy.

I did not anticipate any risks associated with participation in the three phases of the study. However, there were many demands on time, and some inconvenience was involved in reviewing the documents and in taking part in the interviews in Phase 3 and in Phase 4- Study 2. The participants were informed they were free to withdraw at any stage during the interview or take a break if they so wished.

No personal data other than the participants' email addresses, which were used to communicate with them about the study, were collected in the three phases and procedures were in place to protect the participants' contact details. The participants were known to the researcher via their email addresses or due to interviewing them. The names of all the participants in Phase 3 were coded to protect the participants' confidentiality and anonymity. The participants' permission to publish their descriptions as they appear in Table 3.11. was obtained after the interviews; their permission was obtained via email after analysing the interview data. Only one respondent did not respond to the post-interview email; thus, as this participant did not provide consent regarding how they were described in the thesis. The researcher only included general information about the

participants and their contexts to prevent them from being identified even by people who know them.

A summary of the key points raised by each participant during the interviews was sent to each participant separately to confirm the accuracy of the researcher's interpretation. Explanations of the changes were given. The participants were asked to elaborate on or to correct any points. Any direct quotations from the interviews, which were made available in the thesis and which may be available in academic publications or other academic outlets, were anonymised to prevent the participants from being identified; care was also taken to ensure that additional information in the interviews that could identify the participants was not revealed.

3.8 Summary of the Reviews – Results

This section summarises the analysed interviews. The reviewers' comments are organised according to the evaluation criteria (themes) to which they correspond, namely clarity of wording, completeness and relevance of the content, accessibility of the content to non-expert users, and organisation. The revisions made in response to the reviewers' comments are presented subsequently. Responses related to using word lists are presented later.

Word choice in the questions, notes and instructions. The reviewers commented on the wording of the questions, notes and instructions. Generally, these comments praised their clarity, identified necessary grammar modifications and suggested revisions or further explanations of technical terms that could further enhance accessibility.

Responses that suggested more appropriate word choices were coded as 'reword'. Areas that received such responses included the terminology used to refer to the various types of vocabulary and the age range under Question B. Four reviewers—E1, R2, T1 and T2—recommended using more common terms, such as 'English for general academic purposes', 'English for specific academic purposes' for general academic English and 'specialised academic English'. They also

suggested rewording 'more than 19 years old' to '19 years old and above'. There was some disagreement over the word 'audience'. E1 suggested it should be amended to 'learners' because learners were the target audience. T2, by contrast, felt that the word 'audience' enabled users of the tool to describe the group targeted for engagement with the word list appropriately.

The tool included several technical terms, and it was recommended that these terms be further explained, illustrated with examples and defined. These comments were coded as 'revise technical terms' or 'explain technical terms'. As an example of comments given the former code, conflicting opinions arose around the term 'word unit', which referred to 'unit of counting' because the latter, more technical term was likely to create misunderstandings for non-experts. Two expert reviewers (R1 and E1) felt the technical term to be more appropriate, but language teacher T2 stated that 'if you say unit of counting, I would say, "What do you mean?"'

Comments regarding units of counting (e.g. lemma and word family) and criteria for making word lists (e.g. frequency and range) were coded as 'explain technical terms' (for accessibility). The recommended explanation could entail adding examples, defining terms in the list or adding hyperlinks for clarification (E1, R2 and T2). T2 inquired, 'Can I ask what is a flemma and lemma?' When Table 4 was referred to, T2 suggested adding 'very simple explanations next to each term ... the example helped me understand the meaning of those terms, [as] the table is too big to be here'. Regarding question B in Section 3, the same reviewer remarked, 'To be honest, these are difficult terms, and I am not sure I can understand [this question]. I can't comment on it'. Similarly, R2 (referring to frequency, range and dispersion) suggested providing 'definitions next to each term' so that users unfamiliar with them could learn what each term meant 'without having to look at the diagram'. E1 said, '[The] lemma, word-family, word types; these are all really technical'.

Usability of the tool. All the reviewers said that the document was long and that the notes interrupted the flow. They warned that this could lead some users to abandon the tool for practical and motivational reasons. Therefore, they recommended that to improve the tool's usability, the

level of detail should be adjusted to the target audience (experts and non-expert practitioners). R1, E1, E2 and T2 suggested creating two versions of the tool to target different levels of expertise and interest. E1 and T2 further suggested dividing the tool's layout into two sections, one for questions and the other for notes.

E1, T1 and T2 thought that the instructions for using the tool were 'vague' and that they would become confused if they were to use the tool. E1 suggested giving an example to prepare users to conduct their own evaluation.

Accessibility of the tool for non-expert practitioners. The tool addresses individuals across multiple roles in vocabulary instruction and research, not to mention varying levels of expertise and knowledge. Thus, the tool's accessibility was evaluated to ensure that practitioners who used word lists but lacked expertise with corpus tools and word list development could still use the evaluation tool effectively. The process of ensuring accessibility affected the information that was included (content), how that information was presented (usability and wording) and whether the finished tool met expectations (usefulness).

All reviewers indicated that the level of technicality in the document would affect accessibility. Those issues were addressed in the reviewers' comments about word choice. The responses generally indicated that the tool was useful but that the level of detail and the way information was presented were impractical. A selection of reviewer comments in this area follows.

For me, as a user of a word list tool, I find it extremely useful, the more detailed, the more helpful. But as a teacher, they do not care much about research and care more about practicality. So, they will [want] something shorter, simpler (R1).

How lists are made, and what their purposes are ... is vital information (E1).

The whole idea of the tool is a great idea ... [but] a lot of the language teachers that I know aren't really interested in vocabulary research or how word lists are constructed. ... They take it in a very pragmatic way ... that probably [they will look at] about 10 questions (E2).

For novice teachers or those who are less experienced with vocabulary, [the tool might be] challenging. ... [A]nother thing is, when you start asking whether the word unit of the word list is suitable to your purpose, you ask a thing that, again, not all teachers will be aware of. ... Keep [the content] for novice teachers [simple and] to the point; maybe have it a couple pages [at most] (E2).

What about designing an infographic about word lists [in which] all the information needed for non-experts is presented in fewer words and in simple language with explanations for the technical terms? (T2).

Content presented in the tool. This section details responses related to the information presented in the tool. The tool is divided into three sections: word list attributes, context profile and evaluation questions. Reviewers found the first section useful and straightforward, but E1 and T2 stressed the need for a separate document that presents details about word lists and explains technical terms to orient non-experts.

Reviewers found that the second section covered the key elements of an ELT context that must be considered when selecting a word list. However, they noted that describing an ELT context is not simple; it would be misleading to restrict the description by choosing specific characteristics. A word list is often used for more than one purpose (Question A), and the types of vocabulary often overlap (Question B). The same applies to describing the audience (Question C). In reality, there is overlap between learners' proficiency levels, ages and, in some contexts, disciplines. Responses that discussed this overlap were coded as 'component fits but revise', the revision prescribed being that more than one option should be allowed to describe the context accurately. R1 noted the need for an illustration of the ways different age ranges affect word list selection.

The third section, which presents the evaluation questions and supportive notes, is the most important of the three. It was noted that the content of the tool, being descriptive and extensive, would affect its usability. For example, R1 'found those notes [to be] really useful, but it's more of a

discussion, and different researchers may have given different points, so I think it's safer if we give them the reference, feed and let them decide which approach [and] which way they want to use it'. Other responses were coded as 'component is important but move'. Areas receiving such responses were the unit-of-counting table, the diagram of approaches to word list construction and the notes. These are important to non-expert practitioners but they complicate the document's logical flow and increase its length. R2 thought that the notes should be located before, not after, the questions to help users follow the logic.

Responses coded as 'add items' included comments like the one from R2, who thought that the 'audience's first language, cultural background, and ... whether they have the same mother tongue' are factors that must be considered. E2 proposed two additional elements that should be considered when evaluating a corpus: its age and location.

R1 and T2 found Question 4 in Section 3A (If you think the word unit is not appropriate for your purpose and/or audience, is it well-justified?) to be irrelevant to using word lists. It concerns constructing word lists, not using them. T1, E1, E2 and R2 found Question B (What is included in the main list?), which included Questions 6–14, to be irrelevant to the tool's purpose, which is to evaluate word lists for implementation. Regarding Questions 15–22 about the corpus Section 3C, T1 and T2 noted that, from a teacher's perspective, evaluating the corpus may be irrelevant. T2, a language teacher, thought that evaluating the corpus should be performed by course coordinators. They noted that they were simply not interested in the corpus.

Usefulness. Responses related to the tool's usefulness for evaluating word lists were generally positive, though some indicated limitations or flaws.

R1 thought that 'most of the questions in the tool are very useful . . . and by answering those questions, we have insight into the list. . . [but it may] not be suitable for a list of multi-word units'. R1 added, '(This is) very useful for the non-expert evaluators this tool is targeting—how about researchers?'

R2 was doubtful about the usefulness of word lists in the first place and, hence, of the tool. R2 said, '[It] might be useful [in predicting] if the word list I have is good or not if I am using it to design teaching materials or a course'. In terms of the tool's ability to evaluate different types of word lists, however, R2 said that 'specialised vocabulary requires experts in the discipline studied'.

T2, despite believing that the idea of the tool was a good one, said, '

From my experience, the teacher does not have the power or voice to say that the material or word list is not suitable, though they are the ones who are using them in reality ...

[Therefore, I believe] you should investigate if coordinators look at the corpus the list is based on and [how they] evaluate word lists.

Regarding the rating scale, R1 thought that the tool should 'describe the list, not give conclusions [about] whether the list is good or not ... [however], it still depends on the user's judgement'. In other words, drawing a conclusion based on a quantitative process may not accurately reflect a word list's utility because evaluating such lists is a complicated process that cannot be conclusive. Several factors must be considered, and it is not always straightforward to organise those factors. In practical and actual terms, the contexts for language teaching are messy. R1 suggested deleting the scale and instead allowing the evaluator to draw a general conclusion as some of the questions might have been difficult to answer. E2 said,

I think it is good to set up a rating scale, and 1–3 is nice because it kind of helps people push one way or the other. I mean, I know a lot of these go 1–5 to give a little more flexibility, but I'm not sure that that would help me.

R2 thought that the rating scale might lead to complicated or inaccurate results because some questions required a yes-or-no response instead of a number. The reviewer thought that 'some of the questions might require open-ended responses because there are many considerations that

need to be taken into account when talking about the language teaching context'. It is important to note that not all reviewers commented on the rating scale.

Implementing word lists. The interviews with the language teachers (T1 and T2) illuminated how word lists are used. Most importantly, the interviews reflected the need to raise awareness of how word lists are created so that they can be better implemented. E1 stressed the importance of educating practitioners about word lists and how the target audience and purpose of use should guide the development of word lists.

T2 reported that the use of word lists was imposed on them. They reported on their experience and how it influenced their attitudes:

[Word lists are] useless. I used word lists because the management forced [me] to do that, but from my point of view, I think if I want to teach vocabulary, yes, I will choose words based on the topic, but if I give them a topic and just the meaning related to that topic, they can't explore and learn more. We are limiting them to one meaning.

It is worth observing that by the end of the interview, T2 held a more nuanced position on word lists—one that indicated that it was the way word lists were used, not the lists themselves that led to undesirable outcomes.

In contrast, T1 reported how shifts in practices improved the usefulness of word lists. 'We changed how we use word lists and teach vocabulary ... [for] the last four years it used to be memorisation only; things are changing now'.

Some practitioners were aware of the issues related to word list use (such as which unit of counting is suitable), and it was interesting to learn how they addressed those issues. T2, discussing the word family in the AWL, remarked,

I don't introduce the difficult word. ... I thought that if the student knows the verb, she should know the adverb, but that was not the case. Sometimes when I check[ed] the adverb

in the Oxford dictionary I would find that it belongs to a higher proficiency level compared to its verb. But we had to do this adjustment because our target was based on the student's level.

3.9 Revisions to the Initial Tool

The reviewers' feedback guided several decisions regarding the design of the tool. These included deleting, retaining, modifying and adding items, as well as changing the tool's overall structure. A summary of the changes made are presented in Appendix F. Explanations for the main changes are discussed here.

In terms of the structure, an introduction to the tool was added to clarify its purpose, design, target users and organisation. This addition is intended to lead the user to the tool and describe its purpose. Following the recommendation to reduce the length of the document and improve the flow of the tool, the researcher decided to present the main questions and add links to the supplementary notes at the end of the document. The shortened solution recommended by Reviewers E2 and T2 was adopted rather than opting for two versions (one expert and one non-expert version) or dividing the layout into two parts. It would be difficult to validate two versions of the tool, and as pointed by E1, the shorter version may not be useful for experts in corpus-based word lists. The recommendation to relocate some information to a newly developed separate introductory document about word lists was adopted (see Appendix F) to ensure that the document reached a practical length and that non-experts are oriented with basic information about word lists and thus would be able to use the tool. E1 stressed that accessibility to non-experts is important to ensure successful implementation of the tool specifically and word lists generally. The introductory document (see Appendix G) covered the following points, as suggested by reviewer E1 and informed by the literature review conducted by the researcher.

- *What is a corpus-based word list?* The document gives a brief definition of word lists and explains the purpose of their construction. The definition of a word list remains an unresolved topic, and in fact, the survey revealed that practitioners have different definitions of word lists. The three primary types of word lists used in ELT are presented alongside popular examples.
- *How are word lists organised?* This section focusses on the main constituent of word lists— the word—and the common units of counting: type, lemma, flemma and word family.
- *How are word lists made?* Here, the common criteria employed in word lists generation are explained.
- *What are word lists used for?* The applications of word lists in the English-language classroom are explored.
- *What is a good word list?* This concluding section presents the most important takeaway of the introductory document.

References were added so that users would be empowered to explore the literature further. An illustrative example was added to the end of the tool to guide the user through the tool's use. The example walks the user through an evaluation of the AWL (Coxhead, 2000), which was chosen because of its popularity. See Appendix H for these revisions.

It is important to note that while not all reviewers commented on the rating scale, the recommendation to delete the scale was adopted. The respondents found that rating some of the items was difficult and might even be misleading—such as rating the suitability of the corpus when some users do not have access to such information and cannot make such decisions.

A balanced level of technicality was necessary to discuss information about word lists in a manner simple enough for non-expert users to understand but not so reductive that expert users become bored. The following adjustments to item nomenclature and technical terms were made.

The word 'audience' was changed to 'learners'. Because the question about age range was deleted, comments on the wording of its options were rendered moot. Brief explanations of technical terms for the units of counting (e.g. lemma and word family) and criteria for making word lists (e.g. frequency and range) were added to both the introductory and main documents. To improve the organisation of the tool, the notes were moved to the hyperlinks or the introductory sections, as appropriate, and where needed, other terms were briefly explained in the main text.

Reviewers noted that there is an overlap of the purposes of using a word list, the proficiency level, age and, in some contexts, disciplines of the learners. The solution was that more than one option was allowed to describe the context in Section 2. Furthermore, the question about the age range was deleted as suggested by R1.

The main changes to the content and wording of Question 4 in Section 3.1, the word unit part (If you think the word unit is not appropriate for your purpose and/or audience, is it well-justified?) was found to be irrelevant to using word lists. However, it is relevant to making word lists, so it was revised. T1 and E2 found section B (What Is Included in The Main List?) to be irrelevant to the purpose of the framework, which is evaluating a word list for implementation. Thus, the recommendation to delete Questions 6–14 and the accompanying notes was adopted. As suggested by E2, two elements should be considered when evaluating the corpus: age and location of the corpus, so Questions 8 and 9 were added to the corpus part, Section 3.2. Appendix F outlines the changes made based on the initial reviews.

3.10 Summary

This chapter began by describing the research design and the phases of the project, Then, it presented the chronology of the first three phases—the preliminary survey, developing the evaluation tool and reviewing the tool. Each section detailed the methodology, and it presented and discussed the findings and their implications for the next iteration of the tool's development. The

chapter included a description of how trustworthiness and ethics were ensured in Phase 3 and Phase 4, as described in the next chapter.

4 CHAPTER 4. REVIEWING AND TESTING THE CORPUS-BASED WORD LIST EVALUATION TOOL

4.1 Introduction

This chapter, the project's fourth phase, comprises two studies (the expert review and the testing study) to address the research questions posed in Phase 3. For the sake of clarity, the procedures and results of each study are described separately. The analysis of the critical review (from surveys administered in Study 1) concerns participants' accounts of the tool. The analysis of participants' experiences (collected in Study 2) centres on evidence from real-life contexts. The findings from these two studies are integrated in the general conclusion (Section 4.7), and they lead to the adjustments to create the final version of the tool (Section 4.8). The synthesis and discussion of the data and insights gathered in all four study phases, as well as the contributions of the evaluation tool, are reserved for Chapter 5.

4.2 Aim

The aim of this phase was to conduct an empirical test of the effectiveness of the evaluation tool to determine the comprehensiveness and relevance of its content and its efficiency. For this project, the question of 'effectiveness' refers to the tool's capacity to generate an intended result, in this case, an evaluation of one or more corpus-based word lists for pedagogical use. The 'comprehensiveness' and 'relevance' of the content of the tool refer to its questions and supporting materials: that is, the extent to which the tool addresses all the important elements of corpus-based word lists (comprehensiveness) and the extent to which the tool relates to different ELT contexts (relevance). 'Efficiency' applies to the physical and practical aspects of the tool—its clarity, accessibility, structure, time spent on it and ease of use. To assess these aspects, two studies were conducted. Study 1, an expert review, sought critical feedback on the tool from a panel of experts. In Study 2, real-world ELT practitioners used the tool to evaluate the suitability of a word list. This served as a practical application of the tool.

4.3 Research Questions

The research questions, posed in Phase 3, were explored by eliciting the perspectives of both experts and practitioners in this phase and thus the same set of questions governed both studies.

RQ5: To what extent does the proposed tool meet its intended purpose, according to expert users and evaluators of corpus-based word lists?

RQ6: To what extent does the proposed tool meet its intended purpose, according to potential users of the tool (e.g. language teachers)?

4.4 Study 1: Expert Review

4.4.1 *Participants*

In Phase 3, two experts in corpus-based word lists and two researchers were recruited to review the initial version of the proposed tool, which was revised based on their feedback. Another expert panel was assembled by convenience sampling of individuals purposefully selected to assess the revised tool (Appendix H). The sampling procedure, similar to those employed in Phase 3 (Section 3.5.3), sought participants knowledgeable about or experienced in corpus-based word lists who were willing to review the tool and give feedback. The expertise of prospective panel members was based either on their professional backgrounds in developing or evaluating corpus-based word lists for English language teaching and learning purposes or on their familiarity with current research on corpus-based word lists. These participants were essential to this study as their feedback assisted in assessing the effectiveness of the tool; they revised the content of the tool and the supplementary materials, and assessed the efficiency of using the tool. They also highlighted the strengths and weaknesses of the tool and suggested further revisions to improve it. Designing evaluation tools is a never-ending process, and it is important to consult experts in the subject matter as it is important to test the tool with potential users. The researcher approached 20 potential respondents, who were previously known to the researcher or identified by their publications on corpus-based word lists. Publication itself was not a prerequisite for inclusion.

The selection criteria were listed in the invitations distributed to potential respondents by email or direct (private) message on Twitter. The recipients were asked to self-assess whether they were eligible and willing to be on the expert panel. As the population of eligible participants was limited, the researcher employed snowball sampling: individuals contacted for this study were also asked to identify others who might be eligible to participate. The study link was opened by 29 recipients; 12 agreed to participate and assess the tool as expert panel members, and eight of them completed their reviews of the tool. This was a response rate of 27.5% of the 29 recipients who opened the link. Four of the 12 respondents did not complete their reviews. Follow-up emails were sent to them, but there was no response.

The eight expert panel members (here identified as E1–8) who did submit reviews were an assembly with diverse professional expertise. Three reviewers (E1, E2 and E3) possessed extensive expertise with corpus-based word lists and publications related to their development and evaluation. Two reviewers (E4 and E5) were highly experienced with multiple publications related to corpus-based word lists; the remaining three (E6, E7 and E8) also were experienced, albeit to a lesser degree. All reviewers said they were confident in their reviews and recommendations, based on their expertise and familiarity with the relevant literature. Table 4.1 presents a summary of each reviewer's experience and professional context. To ensure confidentiality, codes are used to refer to them throughout this document.

Table 4.1*The Expert Panel Profile*

Code	Experience and professional domain
E1	Language teacher, tester, materials developer, course/programme coordinator and researcher. Domain: EAP.
E2	Researcher. Domains: EGP, EAP and ESP.
E3	Language teacher, tester, materials developer, course/programme coordinator and researcher. Domains: EAP, ESP and English for professional and occupational purposes.
E4	Language teacher and researcher. Domains: EGP, EAP, ESP and English for professional and occupational purposes.
E5	Researcher. Domains: EGP, EAP and ESP.
E6	Language teacher, tester and researcher. Domains: EGP and EAP.
E7	Language teacher and researcher. Current Domain: EAP. Past domain: EGP (for young students).
E8	Language teacher. Domains: EGP, EAP and English for professional and occupational purposes.

4.4.2 Data Collection

Study 1 sought to capture critical feedback from experts in the development and use of corpus-based word lists. The invitations described the study, explained what would be required of the participants and encouraged the recipients to ask questions about participating in the study. A QualtricsXM link, included in the invitation, granted respondents access to the online survey, and it allowed them to download the associated documents. Participants were asked to give their consent online via the study link before participating. They were asked to review the tool critically and to share any comments and suggestions that arose. Finally, after having reviewed the tool, they were also asked to complete a brief survey providing critical feedback.

4.4.3 The Critical Feedback Survey

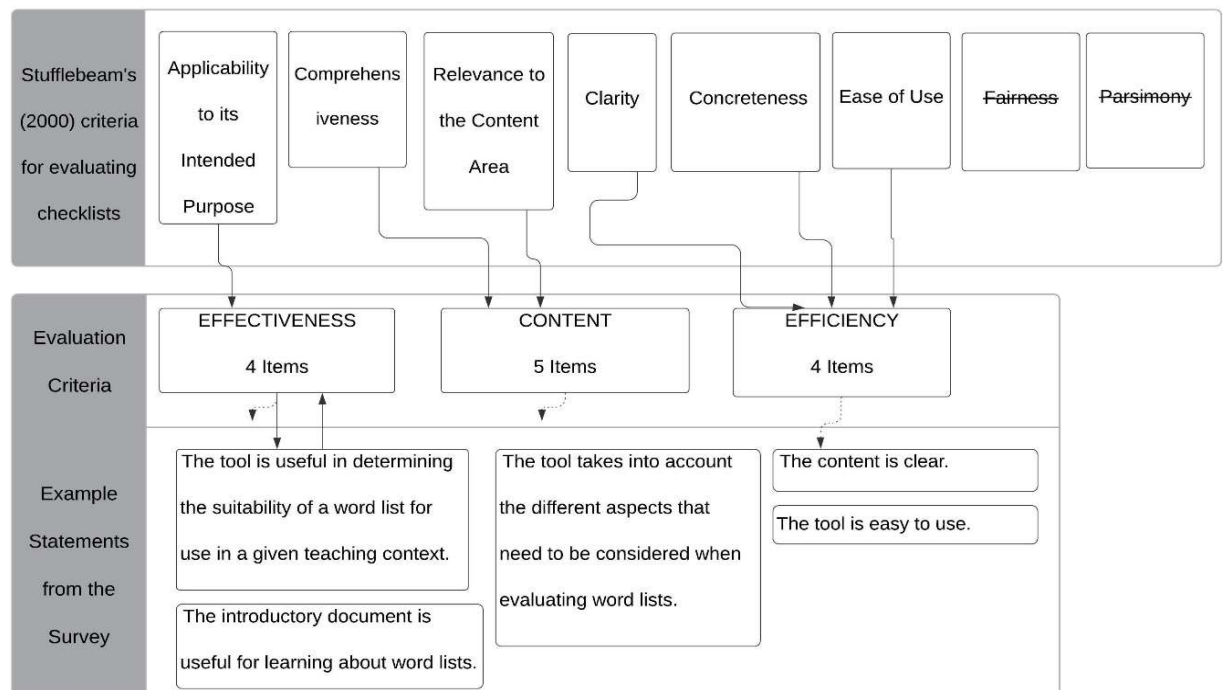
This survey included 14 closed-ended items, four open-ended questions and four demographic questions (see Appendix I). The survey was designed to assess the tool which was developed to evaluate word lists. The survey collected specific data that corresponded to the evaluation criteria, which were based on Stufflebeam's (2000) eight criteria for checklist evaluation: (1) applicability to the intended purpose, (2) clarity, (3) comprehensiveness, (4) concreteness, (5)

ease of use, (6) fairness, (7) parsimony and (8) relevance to the content area (see Figure 4.1).

Although Stufflebeam's criteria were designed for assessing evaluation checklists, the proposed evaluation tool featured the same purpose that a checklist would, i.e. evaluating a resource for selection and implementation. Thus, the criteria were adapted for use in the survey.

Figure 4.1

Criteria for the Critical Feedback Survey



Note. From 'Guidelines for Developing Evaluation Checklists: The Checklists Development Checklist (CDC)' (pp.5-9) by D. L. Stufflebeam, 2000, *The Evaluation Center*.

The survey was divided into three sections corresponding to the three evaluation criteria: effectiveness, content and efficiency. These terms are different from the ones used by Stufflebeam (2000). Effectiveness corresponds to Stufflebeam's 'applicability to the intended purpose', content corresponds to 'comprehensiveness and relevance to the content area' and efficiency combined 'clarity', 'concreteness' and 'ease of use' in Stufflebeam.

Effectiveness was assessed across four closed-ended items. The first and most important statement was item 1 'The tool is useful in determining the suitability of a word list for use in a given teaching context'. The comprehensiveness and relevance of the content were assessed across five items. Item 5 in this section was 'The tool takes into account the different aspects that need to be considered when evaluating word lists'. Other items asked if the questions and the notes were comprehensive. Efficiency was assessed across four items focussing on whether the practitioners understood the tool and found it easy to use. Stufflebeam's criteria of concreteness and fairness were not incorporated as they were irrelevant to the present tool. After each section, there was a space for open-ended comments. There, respondents could elaborate or mention specific examples. They were asked about the strengths and weaknesses of the tool and any changes they would recommend to improve the tool. Each closed-ended item used an interval-response format ranging from 1 to 5, where 1 = *strongly disagree* and 5 = *strongly agree*. Reviewers were asked to report the extent to which they agreed with the items and to comment on each criterion. At the end of the survey, four additional closed-ended questions prompted the reviewer to assess and report their level of confidence in their review. These questions were at the end of the survey because the survey's purpose was to collect critical feedback on the evaluation tool, regardless of the critic's level of expertise.

4.4.4 Data Analysis

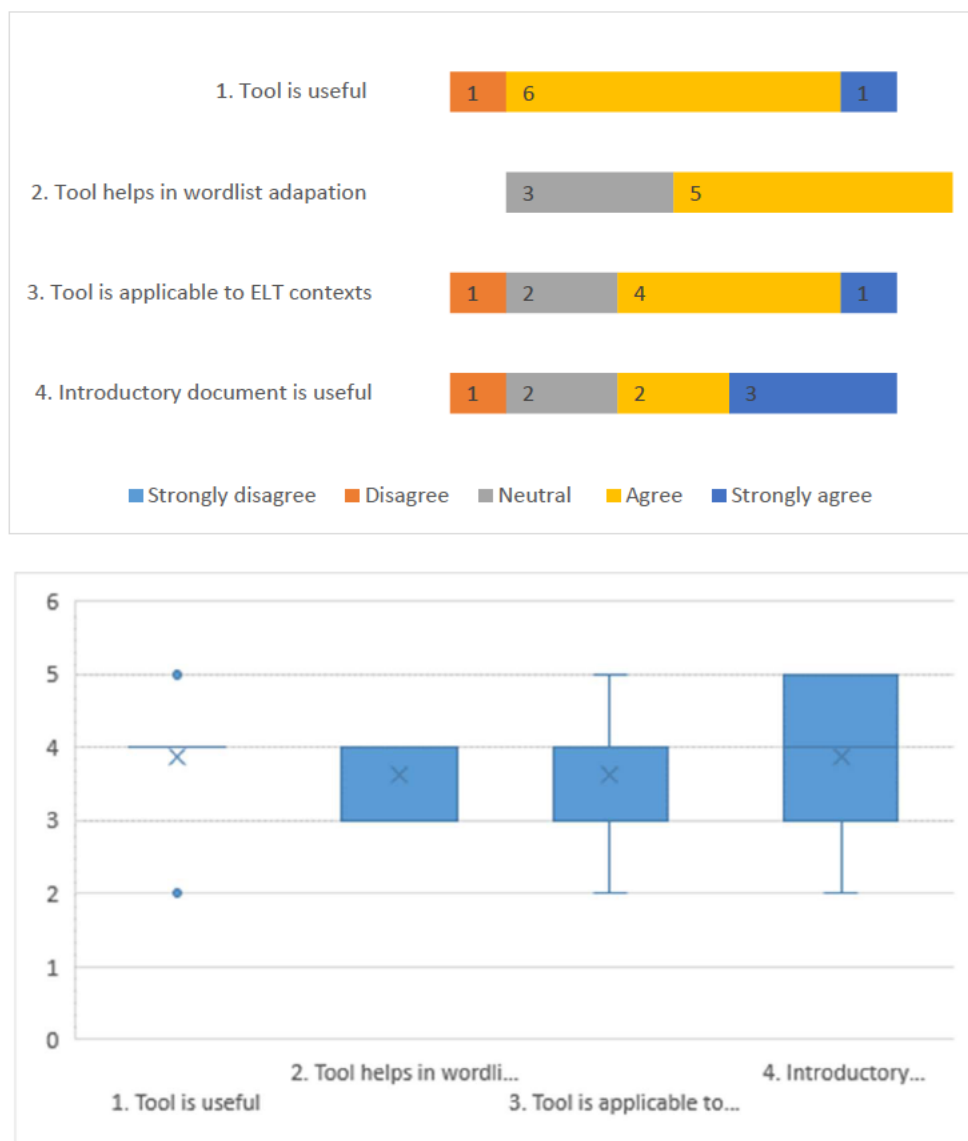
The feedback included both quantitative and qualitative data, but because of the largely qualitative orientation and the small sample of the study, it was not appropriate to apply statistical measures to the quantitative data. Instead, it was decided that it was more appropriate to present descriptive statistics in boxplots and stacked bar charts (see Section 4.4.5) to show the extent to which respondents rated a particular aspect of the tool. Stacked bar charts are useful in showing the counts of the reported results. They allow for comparing the opinions of the reviewers regarding an aspect (Robbins & Heiberger, 2011). Boxplots show the distribution of the ratings for each aspect

along with the median and range. They show where the results are skewed (Cohen et al., 2018). Each rectangular box contains the middle 50% of the scale for each item (the second and third quartiles). For example, the small boxes in Figure 4.2 for items 2 and 3 mean that ratings are in a narrow range, while the large box for item 4 indicates that the middle 50% of the ratings were in a wider range. The disappearance of the box for item 1 indicates that the data were not dispersed; the responses were in alignment. The lines above and below the boxes (the whiskers) show the range of the highest and lowest scores, excluding outliers. For example, in Figure 4.3, item 8 has an upper short whisker (i.e. a narrow range) and a long lower whisker. The x in each box is the median of the responses. The small circles are outliers, i.e. scores that are more than 1.5 times outside the interquartile range. Outliers are true scores, but they may skew the mean (Cohen et al., 2018).

As for the qualitative data, a deductive thematic analysis of the open-ended questions and optional comments was undertaken. The responses to questions and the optional comments were coded using categories developed by the researcher regarding the evaluation criteria and the questions. All the responses that corresponded to a single criterion were categorised using the same code. Verbatim examples of responses to the open-ended comments have been integrated into the results section below. Each response is identified by reference to the expert reviewer who made it.

4.4.5 Results

4.4.5.1 The Effectiveness of the Evaluation Tool. Respondents seemed to agree that the tool was useful in assessing the suitability of a word list for use in a teaching context (see Figure 4.2). The first figure shows the stacked bar of the responses. The second one shows the boxplots.

Figure 4.2*Experts' Responses Regarding the Effectiveness of the Tool*

Some respondents expressed similar concerns. In particular, they seemed uncertain about the tool's usefulness in helping practitioners adapt word lists to their teaching contexts (item 2). They believed that the lack of explicit instructions on how to adapt a list (as observed by E4 and E6) could undermine the tool's utility. Reviewers who believed in the tool's capacity for adapting word lists (E1 and E7) also felt that the adaptation process would depend on the user's knowledge and expertise in vocabulary instruction. They thought that potential or target users of the tool would need more

practical advice. They did acknowledge that some advice was included in the example that comes with the tool.

Figure 4.2 shows that the reviewers were not confident in the tool's applicability to different ELT contexts. E3 explained: 'The tool's applicability is closely linked to its usefulness, thus, if the user (the appropriate user) can use the tool efficiently and make use of it, they can apply it to any ELT context'. E4 again stressed that the usefulness of the tool depended largely on the user's expertise with word lists and vocabulary instruction in general.

Most reviewers praised the introductory document's usefulness in presenting the basic information about word lists. E6 stated that the document served as a helpful summary that covers all the key terms 'in a very concise manner'. However, E6 did raise concerns about the level of detail that had to be absorbed by teachers unfamiliar with word list studies or who had never read about corpus-based research. E6 suggested that 'conducting a training workshop will be beneficial [in helping] those teachers to apply evaluation practice. One document probably won't be enough for them to understand and see the value of this work'. E5 noted that the definitions of the word list units were clear, but the term 'flemma' might pose challenges to less-experienced users. That reviewer concluded by proposing an addition to the example to clarify that term and how it is used.

4.4.5.2 The Comprehensiveness and Relevance of the Content and Evaluation Questions.

As shown in Figure 4.3, all respondents strongly agreed that the tool accounted for each of the aspects that must be considered when evaluating word lists. E1's response was illustrative of the other reviewers' sentiment: 'It brings together a lot of the issues regarding word lists that some teachers may not be familiar with but need to consider when choosing a word list'.

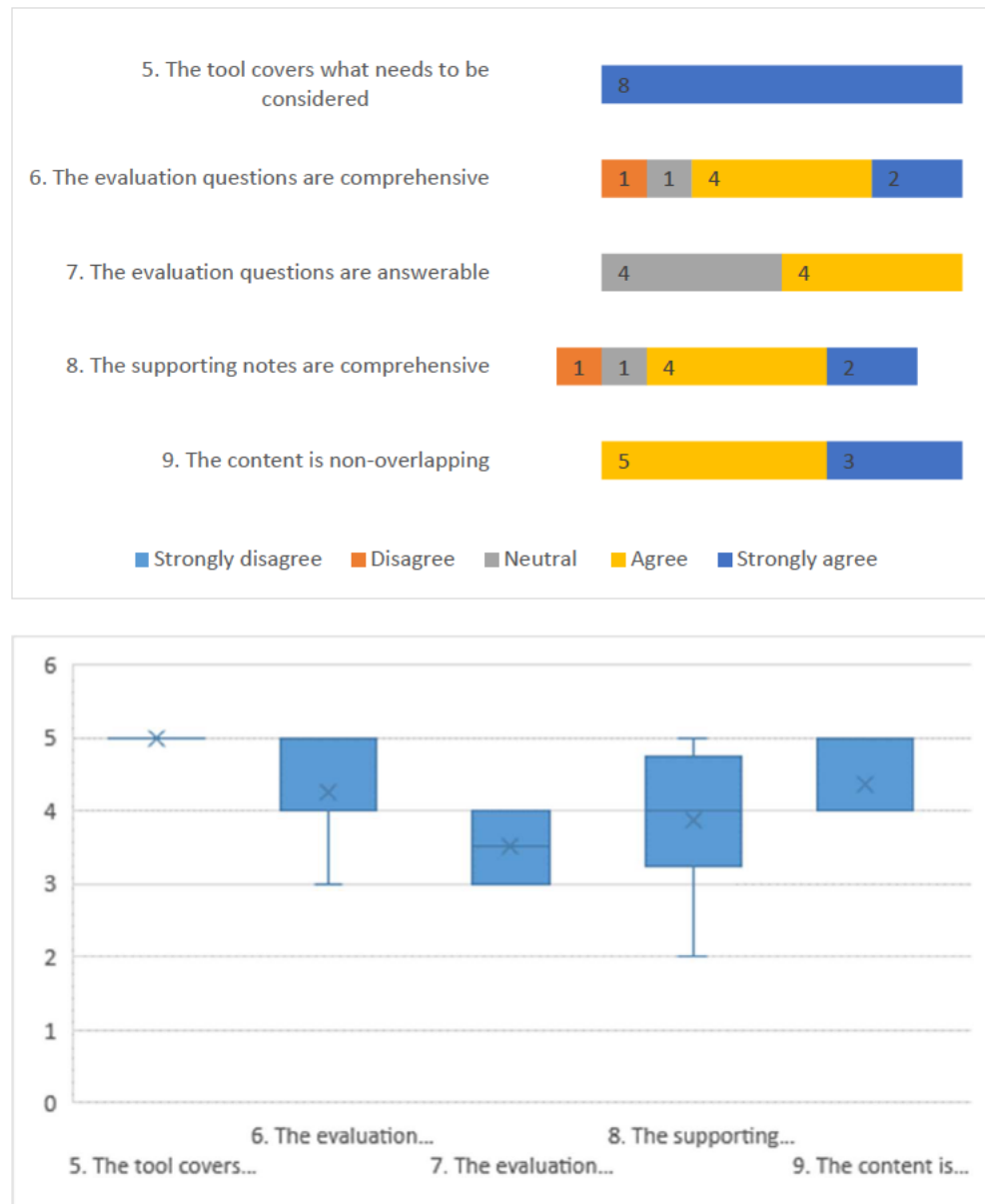
Figure 4.3*Experts' Responses Related to the Content and Evaluation Questions*

Figure 4.3 shows that reviewers tended to assess both the notes and the evaluation questions as comprehensive (items 6 and 8). E4, who found the evaluation questions to be both highly comprehensive and non-overlapping, reported, 'They are clear for someone who has some experience with corpus tools and methods. For such evaluators, the comprehensive notes can surely help to put things into perspective'.

While Figure 4.3 shows a positive tendency towards the comprehensiveness of the notes, E1, E3 and E6 were concerned about the comprehensiveness of the notes. In the responses to the open-ended questions, the reviewers identified a few minor content issues whose resolution could make the supporting notes more comprehensive. For instance, E6 advised that practical recommendations about the size of a corpus should be added to the notes about size. E3 proposed, 'Perhaps concepts such as size could be addressed by having a scale or tick box' that asks how large the corpus or corpora used for the word list in question are. E3 concluded, 'Then a note could pop up to say whether the number they put in seems big enough based on [the] guidelines'. Similarly, E1 pointed out that for Question 15, about supplementary materials, teachers may need to know what tools are available for use with word lists.

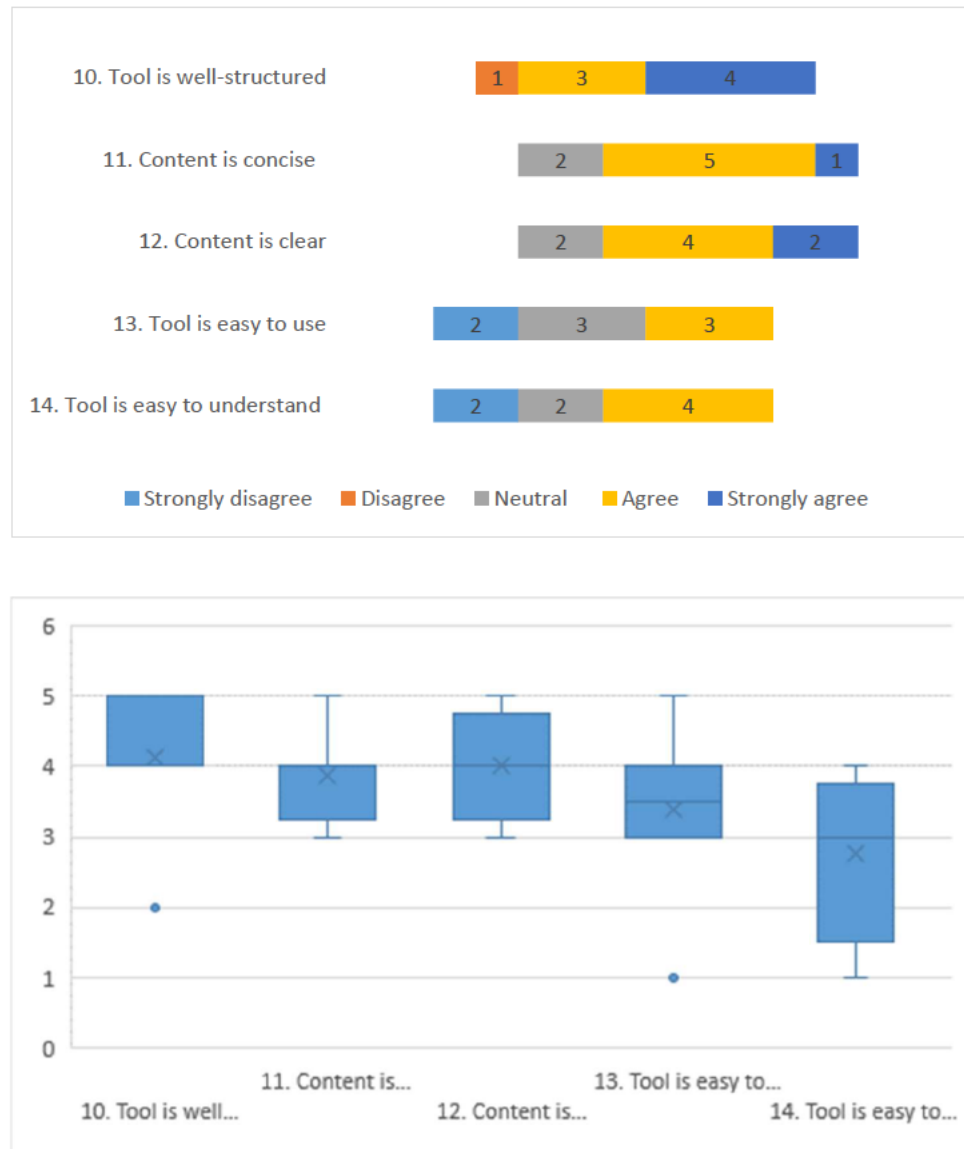
Respondents were less sure that the evaluation questions could be answered (item 7 in Figure 4.3); although the overall trend was affirmative, a full 50% of respondents answered 'Neutral'. E6 explained that whether the evaluation questions are answerable depends on the user's access to published information about a given word list and the user's familiarity with the terms and concepts underlying word lists. E5 was neutral about the answerability of the questions because a few specific questions were difficult to answer. E4 noted that in Sections 3.1, 3.2 and 3.3 (in the tool) about the word unit, the corpus and development of word lists might be difficult for regular teachers. E4 doubted 'that teachers will be interested and/or can evaluate how the list was created'. E4 thought that Section 3.4 about implementing a word list in context was relevant to teachers and easy to understand and that they would answer the questions appropriately. E4 believed it might be difficult for general practitioners who are not very familiar with word lists to fill in Table 1 in Section 1. E4 noted this was especially true about corpus details, size of word unit, and audience, because 'not all practitioners may be able to locate this info and understand it'. Other comments about these items were:

- E5 thought that ‘dispersion’ in section 3.3 was defined clearly, ‘style’ did not seem to be defined. E5 did not know what ‘intensive words’ are, and thought other users also might not.
- E7 observed, ‘Section 2.3.1 didn’t just use the CEFR levels (instead of proficient/ independent/ basic, etc.)’.

4.4.5.3 The Efficiency of the Tool. Figure 4.4 shows that the reviewers agreed that the structure of the tool was good, with one exception. E1 explained that using the tool meant dealing with three separate files, often jumping back and forth from the notes to the relevant sections of the tool. To amend this, E1 suggested hosting all content online. That would make it possible, for example, to summon the relevant notes next to a question by hovering the cursor over an icon. E1 further noted that if the tool remains in MS Word, it could benefit from adding links at the ends of the notes so users could jump back to the relevant question. E1 stated:

It wasn’t always clear which note referred to which question. For example, when I look at Question 13, and I click the link called ‘See note 6’, I am taken down to section D, which has notes for questions 14–17. The inconsistency between labels in the document (note 6 / section D) is a little confusing also.

E3, though, gave a score of 5 for the structure of the tool, and for the open-ended responses noted, ‘Due to its length, language teachers may be reluctant to use it’.

Figure 4.4*Experts' Responses Related to the Efficiency of the Tool*

There were some concerns about whether the tool could be used efficiently. E6 and E8 disagreed with the statement 'The tool is easy to understand'. They explained that the efficiency of the tool depended on the experience of the user. For E6, an expert researcher, the tool was quite easy to use and understand. E5, who was neutral, explained that teachers and practitioners without a background in corpus-based research may need to take time to absorb all the terms used in this tool

before they could truly understand it and use it effectively. An example response was ‘Again, it depends on the background of the practitioners, some might feel the need to do some background reading, but I think everything is explained clearly and concisely’.

Concerning the clarity and consistency of the content, Figure 4.4 shows a tendency towards agreeing that the content of the tool was clear and concise. Comments by E4 and E2 reflect those who were neutral. E4 stated that while the content is concise and clear, the tool would not be easy to follow or comprehend for teachers who did not have prior knowledge of corpus-related terms and concepts. E4 suggested changing the intended audience of the tool to researchers and programme designers (who might have time to read the comprehensive notes and improve their understandings of how word lists are developed). E4 explained, ‘Busy teachers who are most likely not familiar with corpus concepts might find it extremely difficult to answer several items’.

E2 suggested dedicating a section of the notes to explaining the terminology used to refer to the target vocabulary types, as that information may not be known to all ELT practitioners. E2 added, ‘Because I think that some parts of the tool are unclear, I cannot estimate how easy different people will think it is and especially people without any background in linguistics’.

Other, lesser issues included the following:

- ‘In section 3.1, I didn’t understand Question 3. I also think that teachers may not be able to answer Question 4. I suggest giving them some options or examples of how a list can be adapted in terms of its unit’ (E8).
- ‘In section 3.2, Question 7 asks ‘Is the *size* of the corpus suitable to obtain reliable results for the intended purpose?’ Given the admitted (as in your “Notes on questions 5–9”) subjectivity of assessing a corpus in terms of its size, I wonder whether teachers should be expected to be able to answer this question’ (E2).
- The ‘supporting notes’ seemed clear to me, except in two places. In Table 4, I didn’t understand the ‘drawback’ mentioned in relation to the lemma word unit. Why are word

lists with lemmas as their units 'restrictive'? The comment that 'derivations in English are not difficult' is also unclear to me. 'difficult' in terms of what, in which respect?' (E5). E5 further noted that the quote from Stoeckel et al. (2018) in Table 4 may be difficult to understand.

4.4.5.4 Responses Related to the Strengths and Weaknesses of the Tool and Further

Recommendations. Reviewers were asked to report their perceptions of the strengths and weaknesses of the tool. According to E3, a one important strength of the tool was that it had a foundation of solid work in evaluating word lists; the tool collected in one location the most important elements of word lists. Thus, helps users to consider thoroughly how a word list meets (or does not meet) their needs. The reviewers agreed that the tool was quite comprehensive, yet it was concise, and it included examples and notes that could make it easy to use for those with some background in corpus linguistics.

Other reviewers maintained that the tool might not be suitable for all users. Although the researcher had provided a detailed explanation of technical terms, most reviewers felt that it would not be easy for a busy teacher to grasp these concepts simply by reading the definitions and notes. Another critical issue was its practicality. As assessed by the reviewers, the tool was long and time-consuming, and for some users, it might even require prior training.

Reviewers were also asked to propose any additions, revisions or deletions to any component of the tool and to suggest any changes that they thought would improve it E3 and E6 suggested that piloting the tool with teachers with varying levels of experience in linguistics may help to improve it further. It should be noted that these participants were not aware that a testing study (Study 2) was being conducted while their feedback was being collected. The reason for conducting these two studies at the same time was related to the time constraints for the project. It would have been more useful to review the tool based on the expert's reviews and to then test the revised version with the practitioners. Nevertheless, the revisions that would have been made were not

essential to the content and usefulness of the tool. Other responses to this inquiry have been reported in previous sections. The remaining responses are summarised here.

- E6 suggested amending Question 12 to state explicitly: ‘If evidence is provided that the list has been evaluated/validated’, list users can likely use the list with greater confidence. E6 felt that this question should be expanded and perhaps moved out of the criteria for list development in Section 3.3 to the implementation of word lists in Section 3.4. E6 also suggested that the researcher explain why users should choose this tool instead of Nation’s (2016) framework and briefly state how the two tools differ.
- E1 inquired about ‘links to tools that would allow teachers to make use of word lists or to understand them better’. E1 offered helpful examples of the sorts of tools they imagined would be useful: ‘a brief summary of the VocabProfile section of the LEX Tutor site or a link to Laurence Anthony’s pages where AntConc etc. can be downloaded’.
- E3, echoing the sentiments of other reviewers, stated, ‘I think that questions about the way a corpus was composed and its sample will probably be difficult for practitioners to consider if they don’t have any background in corpus linguistics’. E4 suggested developing ‘two separate evaluation sheets: one for specialised researchers and another for teachers. The one for teachers need not include such specialised questions’.
- E4 recommended excising Section 1 entirely as ‘it might be off-putting even to some researchers’. E4 explained, ‘If the [examined] list is not included in the Table [1: The examined word list’s attributes] you attached, then the researcher will need to read about the list and find specialised information about the word list’.

4.5 Study 2: Practitioners Testing the Tool

The objective of this study was to test the effectiveness, content and efficiency of the tool from the point of view of potential end-users in real-world contexts and to explore the users’ experience. Thus, the interview protocol (see Appendix J) focused on exploring how participants used

the tool and the conclusion they arrived at, as well as any benefits derived from, or issues encountered with the tool. It has been a common practice in the development and validation of evaluation tools and checklists to conduct quantitative tests on the reliability and the face and content validity of newly developed tools (e.g. Walker-Egea, 2014). Another approach to validating such a tool involved the agreement of a panel of experts after several rounds (e.g. Llorens-Vernet & Miró, 2020).

In this phase, it was of critical importance to test the tool in the real world. However, it was not possible to compare the participants' responses quantitatively. Nevertheless, each participant's response and experience provided a deeper understanding of and confidence in the tool's usefulness and efficiency. Evaluating word lists is a complex process in which many factors must be analysed. They were asked to evaluate a word list with which they were familiar so they could reflect on their experience with that list and their reflections on it based on using the evaluation tool.

4.5.1 Participants

Mirroring the recruitment procedure detailed in Phase 3 and Study 1, purposive, convenience, and snowball sampling techniques were used to secure participants in this qualitative study. The inclusion criteria were that the practitioner (a) be involved in ELT and (b) have used, developed, or evaluated a corpus-based word list for language teaching and learning, materials development, course design and/or vocabulary testing. The researcher approached 46 potential participants, who were previously known to the researcher or who were identified by their social media profiles or publications on corpus-based word lists. These were different from the expert users involved in the first study in this phase of the research.

Of the 46 participants contacted, 27 initially consented to assess the tool and take part in an interview, but only 11 completed the process. The details of the practitioners, as well as the word list they evaluated, are shown in Table 4.2. Each of the 11 practitioners is pursuing or has completed a postgraduate degree related to language learning or applied linguistics and is currently working as an

ELT practitioner (involvement in research was not an inclusion criterion). Ideally, all participants would have used the evaluation tool to examine the same word list; that would allow for data comparison. However, that might have led to unrealistic results. Thus, each practitioner was asked to choose a word list that they had previously used. The purpose of this process was to allow them to compare their prior experience using the lists with the results from the tool.

Table 4.2

Practitioners Profiles

Code	Profile	Evaluated word list
R1	Language teacher in Europe with many management roles. PhD researcher in TESOL, applied linguistics, vocabulary learning and instruction in English language teaching.	The Oxford 3000 word list included in Navigate Course book by Oxford University Press
R2	Private EFL tutor; involved in teaching EGP to young learners. Masters student in English language teaching and TESOL. This participant has designed their own teaching material, which focuses on basic speaking and listening skills.	Cambridge Young Learners Pre A1 Starters Vocabulary List
R3	University lecturer specialised in English language translation to and from Arabic. PhD researcher in applied linguistics. This participant has some expertise in corpus tools, mostly focusing on ESP.	Combined published specialised word lists
R4	English language instructor at a university language institution. The participant's target learners are beginners who struggled to pass the English language requirement. They have experience in teaching EGP and have worked with corpus tools for ELT purposes as part of their postgraduate studies (Masters).	The English Unlimited word lists; tagged with English Vocabulary Profile level descriptors
R5	EFL/EAP teacher at an English-speaking country university. PhD researcher interested in assessment and vocabulary acquisition.	New General Service List 2 (NGSL2: Browne, 2013)
R6	A manager in a language examination board organisation, focused on developing and validating assessment products with expertise in corpus-based methodologies. Holds a PhD in Education.	The Essential Word List (EWL: Dang & Webb, 2016b)
R7	ESL/EAP instructor at a university where English is an EFL; PhD researcher in Applied Linguistics interested in vocabulary instruction.	New General Service List 2 (NGSL2: Browne, 2013) and New Academic Word list (NAWL: Browne et al., 2013)

Code	Profile	Evaluated word list
R8	Content teacher teaching ESP in an English medium university where English is an EFL. A researcher in Geography.	Academic Word List for Geography, unpublished word list
R9	EAP/ESP language teacher at a university where English is an EFL; pre-intermediate learners in Humanities Colleges.	An adapted version of the Academic Word List (AWL: Coxhead, 2000)
R10	Assistant professor of applied linguistics. Published research in L2 acquisition, vocabulary, and corpus linguistics. Works with language teachers to incorporate word lists in the language classroom.	Phrasal Expressions List (PHRASE List: Martinez & Schmitt, 2012)
R11	EAP language professor and researcher at university where English is an EFL. Interested in SLA vocabulary learning and acquisition of formulaic sequences.	New General Service List 1 (NGSL1: Brezina & Gablasova, 2015)

Of the 27 participants who consented, four withdrew. They explained that, although they met the inclusion criteria, they did not feel qualified to participate because of their perceived lack of experience with the development and use of corpus-based word lists. For example, one practitioner explained, 'I think I may not have the experience required to participate. I use word lists but I do not know how they are made'. A lack of autonomy concerning lesson planning represented a challenge for some participants. One practitioner advised, 'I do not have a choice of what I teach, so even if I evaluate my vocabulary lists, I cannot change them. I don't know how much help I can be'. Another responded to the study invitation, 'we have the course coordinators decide on what is taught, and we have to ensure everyone in the level is taught the same content. I'm afraid most of the questions are not applicable to my context'. Another factor that might have affected the number of participants is that, according to the interviewees, reviewing the tool was a task that demanded time and effort, and the average time to complete the tool was one hour. Busy or overwhelmed professionals who cannot foresee the benefits of employing a tool are not likely to be interested in assessing it. Thus, many potential participants were unwilling to spend an hour or more to assess the tool, in addition to the time required for the interview. Two other participants agreed to take part but did not complete the evaluation because of their busy schedules. Those who fully participated ($n = 11$) were highly motivated to use and benefit from the tool, and they were also regular users of

word lists. This may affect the generalisability of the results, but the intent of this testing study was not to generalise the findings to the broader ELT community because that is difficult to achieve with qualitative data (Creswell & Clark, 2017).

4.5.2 Procedure

The invitations explained the purpose and requirement of the study and encouraged the participants to ask any other questions they had about participating. They were asked to consent either by replying to the invitation email or following a QualtricsXM link. They were also asked to disseminate the invitation and identify other potential participants (snowball sampling).

After they consented but before reviewing the tool, participants were asked to read an introductory document on the basic characteristics of word lists, review a sample application of the evaluation tool to understand its proper use and select a word list to evaluate. They could choose any corpus-based word list with which they had existing experience; the word lists they evaluated are shown in Table 4.2.

After selecting a word list, the participants were asked to use the tool to evaluate the suitability of their list for their real-world teaching context and send their responses to the tool's questions to the researcher. Next, the researcher conducted an online interview with each participant to discuss their experiences with the tool and their opinions concerning its usefulness. Participant R4 asked to complete Section 3 of the tool (the evaluation) during the interview because of the complexity of the task for them and lack of time to orient themselves with word lists. The 11 participants provided information about the processes they followed when using the tool and supplementary materials. They also noted the benefits of the tool, reported any issues they experienced and provided recommendations for improvement.

4.5.3 Data Analysis

The research questions were the same as those used in Phase 3 and Study 1 (see Section 1.7). As in Phase 3 (see Section 3.5.6), the thematic analysis of the interviews used a deductive approach

following the guidelines from Braun and Clarke (2006). The first step involved transcribing the interview recordings and getting to know the data, which in this case included the interview transcripts and the documents from the completed corpus-based word list evaluation tool. The data were approached with pre-determined codes that captured the key elements of the review based on the criteria suggested by Stufflebeam (2000): effectiveness, content and efficiency. Other codes emerged during the coding process. Then, all the data were coded with the initial codes list and themes were identified (see Appendix K for a list of all the themes and codes). Example codes for effectiveness are presented in Table 4.3. They included ‘potential benefits of the tool’, ‘uncertainty about the usefulness of the tool’, ‘helpfulness of supplementary materials’, and ‘change practice after using the tool’. Additional themes included ‘using word lists’, and ‘issues’.

Table 4.3

Example Theme and Codes

Theme	Definition	Codes
EFFECTIVENESS of the tool and supplementary materials	Positive and negative comments on the effectiveness of the tool, the notes and the introductory documents in serving their intended purposes.	Potential benefits of the tool. Uncertainly in the usefulness of the tool Helpfulness of supplementary materials. Change practice after using the tool. Awareness of corpus-based word lists

4.5.4 Results

The interviews yielded rich data about the participants’ experiences and reflections concerning the use of the tool. Valuable information on the participants’ general practices regarding word lists also was collected. Though that information did not address the research questions of this phase directly, it is reported in this chapter and discussed broadly in Chapter 5 in relation to the research questions addressed in Phase 1.

4.5.4.1 Responses Related to the Effectiveness of the Tool and Supplementary Materials.

The potential benefits of using the tool to evaluate word lists were coded as ‘potential benefits of the tool’. This included positive responses in which participants expressed how the tool did and would help them evaluate a word list for their purposes and context. Participants said that the tool indicated the potential usefulness of the list, and it provided guidance on what a practitioner should consider when using or evaluating word lists. Examples of responses related to the benefits of the tool are detailed below.

R6 suggested that the tool would encourage practitioners to have a conscious awareness and appreciation of how word lists function and to critically assess whether a particular word list is genuinely useful for them. R6 contrasted this conscious assessment with the common approach to selecting word lists:

[Using the tool is preferable to] randomly picking the first word list they come across, or one that’s been recommended by someone they know, or they’ve just heard of. So [the tool is] designed to prompt them to think a bit more consciously about whether this word list is correct for them. And to kind of guide them towards the information they need to think about.

R2 also said the tool was a very useful mechanism for self-reflection, by which a practitioner could analyse their context and identify the most appropriate decision. R2 asserted that such reflection is necessary for effective practice and that although they have been using their word lists for a long time, the tool helped them to reflect on the syllabus they had developed based on that list.

R1 said the tool ‘would help me choose a good word list ... it would help me to make an informed decision related to facilitating my students’ needs’. Similarly, R5 reported that the tool helped them to think more thoroughly and critically about their word list. However, they noted,

[ELT practitioners] would still need to use their professional judgments and weighing things up overall to inform that decision. But it wouldn't actually give them a clear-cut answer either way as to whether or not it's that particular word list to how they should use.

A secondary aim of this study was to raise awareness of corpus-based word lists and promote their usefulness by familiarising practitioners with how to use them effectively. This seemed to be achieved when R8 noted that the tool helped them look at word lists from another perspective. They had been using word lists to teach specialised vocabulary, but they were not aware where this vocabulary came from and how to critically think about and assess the lists they use. Likewise, R7 noted the need to raise awareness of corpus research and the effective implementation of word lists. This is emphasised in the tool, particularly the introductory document and the supplementary notes. Such notes were coded as 'awareness of corpus-based word lists'.

Interestingly, while the tool's focus was on lists of individual words, R10 used the tool to evaluate a list of phrases and found the tool very useful in evaluating their list. In fact, they stated,

I have used the Phrase list by Martinez and Schmitt, and the process I was going through in terms of choosing specific items from the list were exactly the kinds of questions that you are asking in your tool. So, this would be a real example of how I have used a given corpus-based list.

Concerns about the effectiveness of the tool were coded as 'uncertainty in the usefulness of the tool'. R3 and R4 wondered about the benefits of the tool for certain practitioners or certain purposes or contexts. However, they both confirmed that if those practitioners were interested in learning about word lists and willing to invest a non-negligible amount of time learning to use the tool, then they, too, would benefit. R4 (an English language instructor at a university language institution) noted that, although they were personally interested in assessing their list and their approach to teaching vocabulary, they were not sure whether they would use the tool in their

current teaching practice. This was because of their heavy workload and a lack of autonomy concerning teaching materials. Similarly, R3 stated,

I would like to say that this tool is useful, but not everywhere. You may need to test it or develop it more. For example, I teach translation, and frequency lists are [a] good source of vocabulary focus. So, this tool would really help me save time in assessing the lists and thinking about them before deciding which one to use. So, because I have been conducting frequency-list evaluation and sometimes adaptation, I can see how useful the tool might be for me, though I did not really consider the corpus and approach of making word lists before. Maybe because I was focusing on frequency and only frequency lists, and with regard to the corpus, I saw it as the source.

Similarly, R2 (a private English language tutor) said,

I think maybe some course designers, syllabus designers at universities or like myself, would perhaps find it more useful because of course. ... This would save someone a lot of time; it might be 30 minutes or an hour of thinking through and writing everything down ... But it is a well-paid time.

Concerning the applicability of the tool in different contexts, R3 concluded,

I think this depends on the role and expertise of the user. If I am in a role with freedom to change in my curriculum or teaching materials, then yes, I would be able to transfer what I learned from evaluating my word lists to my teaching classroom, but this also requires some level of expertise and background on corpora and how word lists are made.

R6 suggested that the extent to which the tool might inform decisions would depend on the specific teaching context, as some 'teaching environment[s] can be really restrictive, there's only so much you could probably do with that, right. So, you'll find that the goal is to make sure you have a functional tool'. R5 (an instructor and assessment developer) thought that the tool was useful

because it encouraged them to 'reflect on the sort of key things. I didn't really know what the key questions were'. However, they did not think they would immediately change their day-to-day teaching practice after using the tool because the selection of their word list was based on practicality and availability. They stated that the questions posed in Sections 3.2 and 3.4 were most relevant to their managerial role (assessment development). R2 and R7 said they would not implement any practice changes after using the tool. However, the tool did cause R2 to reflect on their students' graded readers, particularly how the level of difficulty of the proficiency-one-level reader varied from one publisher to another. R2 also wondered how the vocabulary in those readers suited their learners in terms of theme and level. R7 reported that they had become more aware of different aspects of vocabulary lists that they should consider while planning how to use those lists in language teaching.

In fact, the participants also praised the introductory document. They said it would help practitioners to understand word lists more. Such comments were coded as 'helpful supplementary material'. R1, R5 and R3 noted that the introductory document helped them understand the key concepts before answering the questions in the tool. All the participants referred to the supporting notes during the evaluation. R10 and R11 found the example at the end of the tool very helpful in explaining how to use the evaluation tool.

In summary, the participants considered the tool a useful way to reflect on the potential effectiveness of their word lists in an ELT context. They said that the tool structured the analytical process and fulfilled its purpose. However, because of the time needed to use the tool and its technical complexity, most participants suggested that ELT teachers might not be willing to devote time and energy to the tool unless they could foresee its benefits or become supported. Participants also stressed the importance of the tool being adaptable to practitioners' particular contexts and uses of word lists.

4.5.4.2 Responses Related to the Content of the Tool and Supplementary Materials.

Participants' comments about the questions in the tool, the supplementary notes or the introductory document were coded as 'content'. This included queries regarding how the content of the tool related to the practitioners' teaching contexts and word lists and the difficulties users might have answering some questions. The participants commented on the wording and structure of some questions, and they provided suggestions for improvement. For example, R2 (a teacher of younger learners) reported that some questions were irrelevant to their context. For instance, Question 2.2. was 'What is the target vocabulary type in the context?' To answer this, R2 was not sure which option best referred to the general English vocabulary they taught to their students.

R1 noted that Question 16, 'How useful is each entry in the list? Are there items that need to be deleted or added?' is a compound question addressing two issues. Furthermore, R1 was not sure how to answer Question 17, concerning the availability of guidance on using lists. They queried whether such a question was helpful to users. R1 also felt that the second part of Question 18, which asks how problematic the acknowledged weaknesses of the list are in the user's teaching context, might be challenging for some practitioners. Those comments were coded as 'answerability of the questions'. Similarly, R3 noted,

Yes, the questions [in the tool] were clear, but answerable? Not all of them. As I said, answering some questions depends on access [to] information and understanding the concepts behind making word lists, but I think you are aware of that because you wrote that those sections can be skipped.

R10 explained that teachers don't think about methodological approaches in the same way as expert practitioners or researchers do so. Sections 3 and 4 could be paraphrased or shortened to be more accessible to practitioners. R10 thought that the questions in Section 4 (implementing word lists) were important for teachers. Teachers tend to think about how they would use word lists in the classroom.

R6 described the content of the tool and supporting notes as ‘extremely thorough’, but noted that meaningfully answering some questions required some level of expertise in corpus-based word list development. R5 was also challenged by the phrasing of some questions. They thought the tool addressed important concepts but needed to communicate those concepts more accessibly. For example, they suggested that practitioners with less experience in word lists might not be familiar with the concept of ‘coverage’ and, therefore, might struggle to answer the questions. Regarding the phrasing of the questions, R10 recommended paraphrasing Questions 9 and 15. The word ‘location’ in Question 9 needs to be revised to reflect the intended meaning, i.e. language variety or geographical location. Question 15 needs to be revised to include examples of supplementary materials. By supplementary materials, R10 thought of guidance documents provided with the list (this is addressed in Question 17).

R5 was also confused about how to select a word unit and assess its suitability. Looking at Table 4 in the tool (uses and issues of the common word units) and referring to their word list, they could not find the term ‘modified lexeme’. They considered the unit useful but it was not an option in Table 4. Nevertheless, they noted that some questions in the tool were crucial to ELT practitioners. These were coded as ‘relevance’:

For example, Question 17, ‘Is there guidance on how to use the list?’ is really important to me, thinking about my colleagues, the kind of people that I work with because on CELTA [and] DELTA, and even on some Masters courses, you’re not really taught how to use word lists. And they’re not really a part of materials design, either. That’s another part of materials design, course books and stuff that we’re not really told about, and what we’re supposed to do with them, so I think that one would be, from a practitioners’ point of view, really, really useful.

R10 highlighted the importance of the diagram of approaches in Section 3 to explain how word lists are developed. If multi-word unit lists are to be examined, they noted that some questions

related to identifying and ordering multi-word units in a list need to be added. R10 wondered whether skipping some questions in Sections 2 and 3 (as stated in the tool) would skew the results or diminish the value of the tool.

4.5.4.3 Responses Related to the Efficiency of the Tool in Practice. Responses related to the physical format, structure and ease of use of the tool, as well as the time taken to complete it, were composed within the theme of ‘efficiency’.

All participants commented that the tool and the Word List Table were lengthy documents, and three specifically mentioned the number of pages in each document. Participants also commented on the format of the tool. They said it was long, complicated and sometimes difficult to follow. They noted the need to go back and forth between the questions and the explanatory notes. R5 said that they could not understand the task at first and overestimated the time they would need to complete the tool. However, once they learned how to use the tool, they finished it more quickly than they predicted. Such comments were coded as ‘length’, ‘complicated’ and ‘time’. On the other hand, R10 and R11 found the instructions in the email very helpful, and they suggested incorporating them into the tool. R10 further noted, ‘The explanations in the box are good. It would be useful to draw the user’s attention to them’.

Practitioners took between 50 and 90 minutes to complete the tool. They observed that the time required to use the tool, relative to its benefits to the typical classroom teacher, was something that bears heavy consideration. R1 suggested that practically speaking, most teachers would answer approximately 10 questions or spend between 10 to 15 minutes on the tool. Thus, they would not completely benefit from the tool. Working with the researcher, R4 (who had little experience in word lists) spent 90 minutes responding to the tool’s questions. They also noted that the example should have been placed at the beginning of the tool instead of the end.

R2, in contrast, asserted that the tool was worth the time invested and it would generate valuable returns, particularly for course designers, materials writers and publishers. R2 argued that

despite requiring a high level of engagement, the explanatory information in the tool was necessary, contending that a valuable reflection cannot be completed in just 10 to 15 minutes. They concluded that although the tool took approximately one hour to complete and practitioners may not be particularly willing to allocate such time, it would be time well spent, considering the benefits of the tool. R10 suggested that specifying the time needed to complete each section might help the efficiency.

Presenting the tool in a Microsoft Word document was the issue most cited. Such comments were coded as 'interface'. Four participants asked if the Word document was the final format. R6 suggested that the interface could be improved for longer-term usability because its current presentation might deter practitioners. R1 suggested, 'If you really want it; if it's really something that could become useful for teachers, I would really factor in the usability, the "user-friendly" part of it'. One way to improve usability, as suggested by R1 and R6, would be to transfer the tool into a more user-friendly digital format. R7 noted that they expected an interactive interface of the tool where they would be guided step by step on how to evaluate the word list and then receive some kind of feedback based on their responses.

R2 stated that they printed the document and handwrote their responses, although they were not sure that this approach would be suitable for everyone, especially since the current presentation gives the impression that the tool is text-heavy. R6 noted,

I think a lot of practitioners wouldn't require much support to use the tool. Maybe some might, but a wider question is whether they're put off by the Word document, and whether it looks like there's a lot of information to put in, and not necessarily which information is really critical.

One participant, R1, found the open-ended questions to be inappropriate for practitioners, suggesting that if they were retained, practitioners would benefit substantially from receiving feedback on their answers from the researcher after evaluating their lists. R1 (a researcher and

language practitioner with a management role) noted that some practitioners may not be willing to spend as much time on the tool as R1 did. They expanded:

If a researcher is doing it, maybe they're going to be more likely to take the time. I don't think teachers will. I don't think teachers, depending on how long [it takes] – if [the tool uses] yes-or-no question[s], then I think it [will] probably be okay.

R11 noted that when they first skimmed the tool, they expected some feedback, but after using the tool they realised that this may not be possible and that the questions were useful in pushing them to think about their word list. R10, as an expert and having a specific word list in mind, found the tool easy and straightforward to use.

4.5.4.4 Issues Reported with the Tool. Most participants reported that they did not experience or foresee any issues or problems because of using the tool. However, some participants highlighted the tool's (in)efficiency and level of difficulty for some users, as well as the effort required by users. Those issues are reported under the theme of 'efficiency'. One such response came from R3, who said,

I think [the usability of the tool] depends on the lists you are evaluating. For example, if the target list is not published, or I cannot find information about it, then I might not be able to answer some questions. I mean, even if it is published but the list is not in your table, I am not sure how much time I would spend looking for such information. So, you need to update your table constantly.

All the participants completed Table 1 (Section 1) which asked for information about the word list they evaluated, although R4, did not answer the sub-sections about the corpus and their approach to creating their word list. This was because information about how the list was developed and the corpus from which it was derived was not available. Two participants had to search for additional information about their word lists, as they were not included in the Word List Table.

Another contacted the publisher of their word list to learn about the corpus from which the list was derived and how the list was compiled. It is important to note that this participant contacted the publisher before the study, as they were interested in how the list was made.

4.5.4.5 Recommendations for Improvement. Most of the recommendations were noted in Sections 4.5.4.1–4.5.4.4. Generally, participants' thoughts about improving the tool focused on efficiency and usability. R1 suggested developing a digital, point-and-click version of the tool: 'If it could be in a kind of ... digital format, where users can literally just click ... on things, and it just easily takes them through the tool, that would work really well'. Similarly, R6 suggested improving the format of the tool to allow practitioners to navigate more easily, recommending digitising the tool or adding more hyperlinks.

R1 praised the PDF introductory document and suggested, 'I think maybe you could design the questions to work a little bit ... more closely with that PDF sheet [so] that they could [be] a little bit more clearly aligned'. R6 recommended including a short video tutorial that would illustrate how to complete the tool 'so [users] can kind of see the sequence and see the fact that it looks like it's a lot of information to put in, but actually, this is not a difficult process'.

R1's final recommendation was to include a follow-up component, in which users could provide feedback on whether an evaluated word list was indeed useful in practice and note any strengths and weaknesses. R2 made a similar proposition, observing that the tool would be more useful if users could provide feedback after they used the tool, whereas R6 stated that it would be interesting to see how practitioners used the tool and changed their practices based on it.

To improve the structure of the tool, R10 suggested: (a) moving the notes section to a separate document and (b) moving Tables 3 and 4 to the notes section to shorten the length of the document. Making two versions of the tool, one for the expert researchers and another for the teacher could increase the usability of the tool. Some content is not relevant to the non-expert teacher or it might be difficult for them.

4.5.4.6 Responses Related to the Use of Word Lists in ELT. Throughout the interviews, ‘the use of word lists’ emerged as a consistent theme though it was not addressed in the interview questions or by the research questions raised in Phases 3 and 4.

The participants raised issues related to using word lists in their professional contexts that may affect how the tool is used. For example, R1 noted that some teachers have negative attitudes towards word lists, stemming from the traditional, rote-learning approach that was popular approximately 50 years ago when some participants were language learners. Importantly, the participants themselves all had positive attitudes towards word lists and reported using word lists to guide their teaching and support learners. R2, for example, extensively discussed using and adapting word lists to their context.

R5 noted that their main criteria for selecting word lists were practicality and the availability of ready-made supporting tools. They said they were aware of the criticisms of their list, the New General Service List 2 NGSL2 (Browne, 2013), but chose it because it was practitioner friendly.

Practitioners who chose word lists from course books had confidence in those lists, although they could not always cover the whole list in their teaching. As R4 said,

Those lists are directly related to the teaching materials and the tests. We use those items in grammar exams, for example, in multiple-choice [questions]. This motivates students to learn those words, not only for the vocabulary course but language usage and comprehension.

Regarding the selection of word lists, participants reported that they relied on published evaluations of their lists. For example, R5 wrote, ‘I’ve seen favourable comparisons of coverage with GSL & the other NGSL’. Another important factor that contributed to the selection of the NGSL, in particular by R5, was that the developers of the list provide support, guidance and ready-made tools for practitioners. R2 noted that they would use the tool in the future but, if it were not available, they would continue to select word lists based on their target learners and the reputation of the word lists’ publishers.

R7 discussed how they employed the NGSL2 (Browne, 2013) and the NAWL (Browne et al., 2013), saying they were more appropriate because they were based on more up-to-date, larger and well-balanced corpora. They also discussed the applications they used with their students and the lack of valid and corpus-informed vocabulary tests. R7 mentioned the gap between language teachers and corpus tools and said they lacked awareness of readily available lists that they could use for setting vocabulary planning goals. They found the Word List Table a useful resource that organised published word lists and information about them.

R8 reported on the need to update word lists and to learn how to add new items because this required expertise and knowledge. Before learning about how word lists are made, they relied on the items students confuse and their intuition for updating their word lists noting that 'As the field of geography is developing the needed to add new items, if I knew how, it would have been very useful for me and the students'.

4.6 Research Ethics and the Confidentiality of the Participants

The procedures for ensuring the confidentiality of the participants' identities and responses and for collecting the informed consents were the same as those used in Phases 1 and 3 (see Section 3.7.). Study 1 in Phase 4 sought to capture critical feedback from experts in developing and using corpus-based word lists via an online survey. A QualtricsXM link, which was included in the invitation email, included a detailed participant information sheet and a consent form. The participants were asked to give their consent online via the study link by ticking a box before participating. The respondents were then granted access to the online survey, and were allowed to download the associated documents.

They were also asked to complete a brief survey to provide important critical feedback. At the end of the study, the participants were asked to report on their confidence in their review and level of expertise (see Question 3 in Appendix I). These descriptions were reported with the participants' consent in Table 4.1, Section 4.4.1. In Study 2, which involved online interviews, the

invitations explained the purpose and requirements of the study, and encouraged the participants to ask any questions they had about participating. They were asked to consent either by replying to the invitation email or by following a QualtricsXM link. Only audio recordings took place, even though some participants switched on their cameras.

No personal data were collected in this phase other than participants' email addresses, which were used to communicate about the study, and steps were taken to protect their contact details. The names of all the participants were coded to ensure confidentiality. However, it is understood that, because the geographical locations of the participants and the names of the word lists were given in Table 4.2, Section 4.5.1. (to show the contexts of the participants) and because the names of famous researchers who are experts in word lists have been cited, hypothetically speaking, these participants may be recognisable by people who know them in person. Moreover, the participants' permission to publish these descriptions as they appear in the thesis was obtained after the interviews. A description of each participant, as shown in Table 4.2., was sent to them together with a summary of their feedback to confirm the accuracy of the researcher's interpretation, and their permission was obtained via email after analysing the interview data.

4.7 General Conclusion Based on Study 1 and 2

This section details the primary points reported across both the expert review and testing studies, Study 1 and Study 2.

1. The quantitative and qualitative data collected from the participants suggested that the tool is useful in determining the suitability of a word list for a teaching context. This is because the tool accounts for the elements that must be considered when evaluating word lists. Some teachers might not be familiar with those elements, but they must be aware of them when selecting a word list for practical use. Further, the tool provides clear guidelines to help the user of a word list gain deep insight into its applicability for

their specific purposes and context. For those who are not familiar with corpus-based research and word list studies, the tool acts as a shortcut for developing a basic understanding of the field.

2. A challenge arises in balancing the interest in detail and complexity against the interest in efficiency. In its current form, the way the material in the tool is presented should be improved. Housing the tool and all associated content online was a popular proposal that may improve the efficiency of the tool.
3. Because of the lack of explicit instructions on how to adapt lists, respondents seemed less certain of the tool's usefulness in helping practitioners adapt word lists to their teaching contexts, especially if they lack experience with corpus-based word lists for ELT.
4. Generally, reviewers praised the introductory document's usefulness in presenting the basic information about word lists concisely. However, the document is still cumbersome for teachers who are not familiar with word list studies and corpus-based research. Admittedly, a single document is not likely to be enough to convey the value of this work to non-experts. Pre-use training in evaluation practice might be beneficial for such teachers. They also suggested better links between the document and the questions.
5. The tool applies to different ELT contexts, although further investigation is merited.
6. There are minor content issues related to the comprehensiveness of the questions and notes.
7. Respondents were not sure that all evaluation questions could be answered. The ability to answer questions seems to depend on the user's familiarity with the terminologies and concepts underlying word lists and their access to published information about the word lists—particularly, information on the list's development.
8. The audience of the tool should be revised to include researchers and advanced programme designers, but not teachers. Teachers and practitioners without backgrounds

in corpus-based research may require too much time to absorb all the terms used in this tool for it to serve them as an intelligible and efficient tool.

4.8 Final Adjustments to the Evaluation Tool

In light of the comments from experts and practitioners, the researcher made the following adjustments to improve the *effectiveness* of the tool. The final iteration of the tool is presented in Appendix L:

- Modified the target users of the tool and stated in the introduction that users need basic knowledge of or an orientation to corpus-based word lists, and they could refer to *The Guide* in Appendix M if needed.
- Stated clearly that users can skip some questions because the purpose of the tool is to evaluate by reflection on the context and the examined tool, not to generate a numeric value.
- Encouraged teachers to believe in their perspectives. This has been noted in *The Guide* (Appendix M).

To improve the *comprehensiveness* and *meaning* of the items of the tool, the researcher made the following adjustments (See Appendix L):

- Matched the options for proficiency levels with descriptions in Table 3.
- Paraphrased Question 9 '*Is the location of the corpus suitable for the intended purpose?*' The word 'location' was revised to reflect the intended meaning, namely, language variety or geographical location.
- Revised Question 12 to explicitly state '*If evidence is provided that the list has been evaluated/validated, list users can likely use the list with greater confidence.*'
- Revised Question 15, '*Are there supplementary materials based on the list for use in the class?*', to include examples of supplementary materials, for which the reviewers suggested guidance documents provided with the list (as Question 17 also addresses).

- Revised Question 18: *'Are the weaknesses of the list acknowledged and justified? How problematic are they to your context?'* It is a methodological question that might be challenging to the language teacher.

To improve the *efficiency* of the tool, the researcher made the following changes:

- Added hyperlinks in the notes section to take the user back to the questions.
- Added clearer instructions in the introduction.
- Added an estimate of the expected time for using the tool in the instructions. However, estimated times for individual sections are difficult to calculate because some sections require more information, and users may skip some questions.
- Moved the example to the beginning of the document.

4.9 A Training Intervention for Corpus-based Word Lists: Plan A

The researcher had intended to conduct Study 2, Phase 4, in a different way. The planned training intervention was motivated by the observation that practitioners needed guidance to implement corpus-based word lists effectively in their actual teaching practice. There has been a substantial increase in the number of pedagogical word lists published for use in ELT in the past two decades. The data generated from the survey results (Phase 1) and the interviews (Phase 3), as well as the small amount of published research (Banister, 2016; Burkett, 2015, 2017), showed that there was a lack of awareness about corpus-based word lists in terms of how they are defined and how they can inform vocabulary instruction, as well as that there were misconceptions and negative attitudes that needed to be clarified. For example, a key concept related to corpus-based word lists is frequency, as many lists are based on the frequency criterion. As has been highlighted in the literature (Aizawa, 2007; Egbert, 2018; Richards, 1970; Schmitt et al, 2021) frequency has limitations as a criterion for determining a word's usefulness if used in isolation. This research supports Nation's (2016) argument that most corpus-based word lists need to be adjusted subjectively for the specific purpose and context of planned use. The literature (e.g. Chambers, 2019; Römer, 2006) also

identified a gap between corpus linguistics research and practice and this gap remains in regard to corpus-based word lists (see Section 2.5.3.1).

Based on this need, the planned intervention study aimed to offer in-service teachers a training in the general features of word lists and how corpus-based word lists could be used in vocabulary instruction, as well as to explore teachers' perceptions of using word lists in their classrooms. After the introductory workshop, it was intended that they would have been asked, firstly, to use the tool to evaluate the word list, and then, secondly, to reflect on the tool after using the list. The main aim of this intervention would have been to bridge the gap between word list research and practice by helping practitioners, particularly teachers, to understand where word lists come from and what they have to offer to the language classroom. Furthermore, this training would have helped the non-expert user in using the tool and the evaluation process as confirmed in Study 2-Phase 4 (see Section 4.7.). Teachers were selected because they have immediate use of word lists. Also, this group of practitioners is the least expert in corpus linguistics and has the least resources and access to research.

The results of this intervention would have explored in-depth perceptions of word lists before and after the training intervention and the effectiveness of the training in changing practices and assisting practitioners use word lists effectively. I would have focussed on practitioners (teachers) who were not aware of word lists or who had not used word lists. This would have been an attempt to bridge the gap between corpus-based word list research and ELT practice, as well as to dispel some of the myths about word lists.

In the circumstances, I had to recruit participants who had used word lists for vocabulary instruction, and to ask them to use the tool to evaluate a list with which they had had experience. Instead of the training intervention, I asked them to review *The Guide* (Appendix M) regarding its

content, accessibility and potential usefulness. *The Guide* summaries the content of the intervention (see Section 5.3. for information about this document).

Putting this content online at the beginning of the COVID-19 outbreak was not possible due to the timescale of the project. The researcher's lack of skills and experience in delivering content online at that time was another reason for cancelling this intervention. It was not clear whether individuals would have welcomed this intervention. In terms of testing the tool, it would be more beneficial to have it tested by regular language practitioners after the intervention and to determine how useful it may be. It is unclear how useful *The Guide* (Appendix M) is in introducing basic word list concepts in the current format.

While this was not what had been planned, the results yielded by this study were sufficient to claim the potential usefulness of the proposed tool, at least for a particular group of practitioners. It was also found that *The Guide* was potentially useful, but I would need data from the target audience to validate this finding. The interviews shed light on issues relating to the use of word lists and attitudes towards them, but these were reported by practitioners who had already used word lists.

5 CHAPTER 5. GENERAL DISCUSSION

This chapter is organised around the themes that guided this research and were derived from the data collected. The first section features discussions, based on the results of the study, on how word lists are exploited in ELT and how practitioners decide on a suitable word list for their pedagogical use. It also describes the reported attitudes towards word lists to discuss how those attitudes influence the exploitation of word lists. The subsequent section concerns the proposed evaluation tool. Chapters 3 and 4 focussed on reporting the reviewers' analysis of the tool, and they discussed how these reviews informed the development of the tool. This section presents those analyses and examines in greater depth the potential usefulness and issues of the tool in its different iterations. The last section presents an offshoot of the tool targeted to practitioners who lack knowledge about corpus-based word lists. It is important to note that despite the differences in the aims of the phases of this study, there is an overlap of the themes explored. This chapter discusses these themes and the research questions that addressed the proposed tool in general.

5.1 Practices and Views related to Word Lists in ELT

5.1.1 Word List Exploitation in ELT

Word lists are a well-established feature of vocabulary instruction, but research into the exploitation of word lists in ELT is limited. At the beginning of this study, it was not clear how the proliferation of word lists for ELT has influenced the use of word lists or the extent to which published word lists are used in ELT and for what purposes. The scant research that was available focussed on exploring how word lists are utilised for teaching academic English and the challenges of using them (Banister, 2016; Burkett, 2015, 2017).

The first research question posed in the initial phase (RQ1) was *What distinct uses are made of corpus-based word lists by practitioners, researchers and learners involved in L2 teaching and learning?* Directed towards the wider ELT community, Phase 1 of the current study explored how much word lists were used and what practices related to word lists were used in ELT. A detailed

discussion of those practices is presented in Chapter 3, Section 3.3.4.1, the key findings are given here. During the interviews in Phases 3 and 4, though the conversations focussed on analysis and review of the tool, the participants talked about how they used word lists for their teaching and research.

The data collected throughout the study revealed an acknowledgement of the usefulness of word lists. Practitioners used word lists as a resource for prioritising the vocabulary that merits instruction, and using word lists in this way aligned with the cost-benefit principle. This suggests that, by learning very frequent words, learners receive the best return for their learning effort (Nation, 2013, 2016).

The online survey revealed that almost 80% of the 74 practitioners have used word lists, while approximately 20% have not. Burkett (2015) reported that 60% of the participants in the 95 university intensive English programmes surveyed used word lists. Rossiter et al. (2016) surveyed vocabulary teaching practices of 30 ELT teachers, and they learned that 60% used word lists. Similarly, Banister's (2016) investigation of how the AWL is used revealed that the 193 EAP participant teachers used it. Those studies showed that corpus-based word lists are indeed used.

The data revealed that the most common purpose of using word lists was to select the words that students are expected to learn, which is an obvious purpose of making word lists, and it has been noted in the literature (e.g. Nation, 2016). In addition, practitioners indicated that they used word lists to develop day-to-day practical activities; for designing deliberate, language-focussed learning activities and material; and for assessing the reading level of assigned texts and other resources. These reports also show that practitioners have heeded Nation's (2007) call to use the four strands principle for vocabulary instruction, which suggests that deliberate vocabulary learning (the language-focused learning strand) should account for 25% of the time in a well-balanced vocabulary learning programme. The other strands are meaning-focused input, meaning-focused output, and fluency development. Nation (2016) argued that word lists can inform vocabulary

development through four strands (see Figure 2.1). The practitioners in this study reported using word lists for informing meaning-focused input and output and planning deliberate learning, yet no reference was cited related to using word lists for fluency development.

Burkett (2017) categorised word list uses into four categories: course design, teaching and learning, assessment, and materials development (see Table 2.3). He noted that these categories are interrelated and lists might be used for some aspect before another. For example, course design would likely precede assessment or materials development. Following Burkett (2017), the reviewers in Phase 3 noted that it was difficult to disentangle the purposes of word lists, as the initial version of the evaluation tool required users to choose one purpose of using the word list being reviewed. Burkett (2017) explained that because language teaching and learning are complex processes, if you use a list for teaching then you are likely to test the students on how well they learned the items on the list. The aim of the proposed tool was to evaluate word lists in terms of their purpose. To satisfy that objective, the revised version of the tool allowed users to select more than one option.

However, how word lists are used to inform direct language teaching and learning and to what extent these uses are successful are yet to be determined. Nation (2016) had observed that the process in which corpus-based word lists are used to support materials development, course design and tests is still unclear. He attributed this to a lack of knowledge about vocabulary acquisition and the value of word lists. On a broader level, even though there is evidence that corpus materials are used effectively for materials development, the question of why they are not used to their full potential deserves more study (Curry et al., 2020). In terms of how word lists are implemented, two reported experiences are worth mentioning. First, T2 (Phase 3) reported that they were forced to use the AWL (Coxhead, 2000) in their EAP programme. T2 thought that the list was not always relevant to the topics in the teaching materials and that the word list presentation led to a restricted understanding of the word's meaning. T2 was aware of the unsuitability of the word family concept for their students despite their lack of knowledge of corpus linguistics and adapted the word family

accordingly. This participant's experience reveals the challenges of using word lists that might have engendered negative attitudes towards word lists leading some to think that word lists are useless. Despite publications such as those by Dang (2020), Nation (2016) and Schmitt and Schmitt (2020) to highlight the pivotal role word lists can play in vocabulary instruction, the value of word lists is still underestimated because of the challenges that come with using them. By contrast, the experience of T1 (Phase 3) showed how practitioners' practices influence the usefulness of word lists. In the institution where they teach, they used to use a word list as a self-study tool. However, this encouraged rote memorisation and diminished the usefulness of their list. However, under a new administration, when they started using word lists as part of the curriculum and in context, they found word lists to be more useful and facilitative in the learning process. This confirmed the importance of understanding how word lists are constructed and how to use them effectively (Schmitt & Schmitt, 2020). Their approach to word list implementation employed Nation's (2007) four strands of balanced vocabulary instruction, in which word lists are used to provide meaning-focused input and to guide learning (see Section 2.5.2).

The second question that guided the initial survey (RQ2) was *What methods are used to generate pedagogical corpus-based word lists?* A recent trend in word list implementation is to replace existing word lists with new lists. Dang (2020), however, called for making existing lists more suitable for particular learning and teaching contexts. She encourages word list developers to facilitate the use of word lists by creating smaller lists that match the proficiency levels of particular groups of learners. As an example, she cited the EWL (2016), which targets beginner learners: function and content words are presented in separate lists because it is easier to introduce content words before function words, though the latter are more frequent. Memorisation and mastery of the 800 items in the list enable beginners to recognise 60% and 75% of the words in a wide range of spoken and written English texts, respectively. By reducing the number of items in a word list for

pedagogical purposes, it is possible to focus attention on the words that provide the highest return in the short term. Another example is Ward's (2009) engineering word list for Thai students.

Forcing a long list of words into a language course is among the most common problems for practitioners that use them (Nation, 2016). Solutions proposed by Dang (2020) included adapting word lists to learners' proficiency levels and dividing the lists into smaller, more manageable subsets to fit the individual courses in a language programme. She cited the Academic Spoken Word List (Dang et al., 2017) as an example. The list's developers divided the list into four levels, and they claimed that learners could focus on the levels that apply to their current level of general vocabulary. Each level is further divided into sub-lists to allow easy incorporation of the list in language learning programmes.

In Phase 4, the data revealed that practitioners mostly used word lists for classroom teaching and learning and that they usually adapted word lists for their context. One issue they noted for the successful implementation of word lists was the gap between word list research and ELT practitioners' knowledge and training. The practitioners observed that many of their colleagues are not aware of word lists and cannot use them independently, and they know little to nothing about corpus linguistics. These comments broadly support previous research that identified the gap between teacher cognition and education and training in relation to corpus-based tools for language teaching (e.g. Frankenberg-Garcia, 2012; Latif, 2020; Mukherjee, 2004). Using word lists effectively requires some basic knowledge about how they are developed and organised (Schmitt & Schmitt, 2020) and how vocabulary frequency and control can inform the teaching and learning of vocabulary (Nation, 2016). This coincides with the practitioners' note of the neglect of teaching and learning of vocabulary training in their teacher training or Master's courses. To emphasise the importance of training, R11 mentioned they lacked knowledge about corpus analysis and vocabulary control even after a Master's degree in applied linguistics. However, taking a course in corpus linguistics during the PhD programme, they noted that, even though they worked in a restrictive teaching environment

where the teaching content and assessment are fixed, they adapted the list to their teaching.

Chambers (2019) and Nesi et al. (2021) also noted the gap between corpus linguistics research and practitioners' training, and they identified it as an aspect of teacher training that warranted additional attention.

5.1.2 Practitioners' Attitudes towards Word Lists

One aim of this research was to explore what ELT practitioners know, believe and think about corpus-based word lists for vocabulary instruction and how their cognition shapes their teaching practices. The fourth research question posed in the first phase (RQ4) aimed to explore *How are word lists perceived by ELT practitioners, researchers and learners?* In each phase of the study, practitioners who used word lists reported positive attitudes and effective implementation of word lists. They also noted negative associations and misconceptions. Thus, it is possible to conclude—in line with Borg's remark (2003)—that the participants in the study made judgements about word lists based on their own learning experiences, their classroom teaching, and the ideas adopted during training.

The Phase 1 survey revealed that approximately 20% of the 74 participants had not used word lists. Eight had not thought of using a list, three thought word lists were useless, and two were not aware of word lists. These reasons for neglecting word lists can be attributed to a lack of awareness and negative associations with corpus-based word lists. Of the 80% ($n = 59$) of the 74 participants who used word lists, 76.2% thought that word lists were useful, 18.6% doubted the usefulness of word lists. Burkett (2015) reported that 60% of the participants in the 95 University intensive English programs surveyed used word lists; 90% believed that word lists are useful, and 80% were somewhat aware of word lists. Rossiter et al. (2016) surveyed vocabulary teaching practices of 30 ELT teachers to find out that 80% were aware of word lists, but only 60% used them. Similar positive attitudes were expressed in Banister (2016) where the participating teachers thought the AWL was useful because it was relevant to their students and because it was based on scientific

research—not intuitive judgments. These findings suggest that many ELT practitioners are aware of word lists and have positive attitudes towards them, but they are not necessarily using them.

Nevertheless, some practitioners in ELT are not aware of the existence of corpus-based word lists. Again, caution must be taken when interpreting these numbers, as the samples and cited literature are biased toward academic contexts.

Based on this study's results and reported explanations, the positive trend could be attributed to the observation that practitioners and researchers who are aware of how word lists are developed have used word lists effectively and, thus, have positive attitudes. These practitioners understand the underlying principles of developing word lists and how frequency can inform vocabulary learning. They see word lists as a resource and a guide, not only as a self-study tool for students. Teachers outside academia may not have this background (Latif, 2020). Thus, it could be said that the gap between word lists and teaching practice should be addressed.

One possible explanation for the different numbers – and an interesting finding – may be based on how word lists are defined. One respondent in the preliminary survey, a language learner, defined a word list as a list of words learned in class. Another language teacher reported that they do not use word lists and said in the comments sections that they use a list of words for the novel they teach, but it is not a frequency list. This example reflects a lack of awareness about what corpus-based word lists are. They could be of any size, and they would not have to be frequency based. Earlier word lists were developed intuitively, and they were found to be a good reflection of usefulness.

These beliefs may stem from experience and knowledge: two practitioners (Phase 4) from the Middle East self-taught themselves, six others (not from the Middle East) were introduced to corpus linguistics during their higher education, while the other three (Middle East) knew nothing about corpus linguistics. Their beliefs align with those of experts such as Folse (2004), Nation (2007, 2013, 2016) and Schmitt (2010), who suggest that effective intentional vocabulary instruction should

occur. In Phase 3, however, while T1 reported effective adaptation of the AWL to their teaching context, they thought that list-learning was useless, particularly the decontextualising factor. Talking with the researcher, the teacher discovered that their problem with word lists was related to teaching and that they should not feel guilty. Many misconceptions and negative attitudes emerged from misunderstanding and lack of awareness of word lists, so raising awareness about corpus-based word lists and corpus-linguistics, in general, could be beneficial.

On the other hand, the Phase 1 survey revealed doubt about the usefulness of word lists. Similarly, participants in the interviews in Phases 3 and 4 conveyed negative beliefs and challenges related to using word lists in their teaching. Borg (2001) offered an explanation for negative attitudes towards using word lists in ELT. They might be based on prior experience of English language learning, lack of knowledge about word lists, teacher education and contextual factors such as the challenges of using word lists, teaching theories and erroneous practices. Each of these factors is discussed in the following paragraphs.

First, these negative attitudes could be attributed to the traditional approach of rote learning that was popular approximately 50 years ago and from which those teachers learned vocabulary. For instance, R1 in Study 2 recounted that they learned French vocabulary by memorising word lists, but they forgot most of the words. The effect of teachers' language learning experiences and their beliefs may provide an explanation for teachers who prefer alternative methods of learning as opposed to those they had (negative) experience with explicit vocabulary learning (Gao & Ma, 2011).

The challenges related to using word lists for classroom teaching and learning, such as dealing with the large size of such lists, decontextualisation and meaning, and the perceived irrelevance of the list to the teaching materials, might have contributed to those negative attitudes that sometimes led to neglecting word lists. R6 noted that contextual factors have a substantial influence on attitudes. Schmitt and Schmitt (2020) explained that negative attitudes towards word lists can be attributed to the misconception that word lists are intended for decontextualised

vocabulary learning. Similarly, Banister (2016) reported that even though they praised word lists, participants thought that list learning did not match their communicative approach and that there are better approaches to teaching vocabulary.

Finally, management of word lists without knowledge of their origins or how to use them effectively might have led to negative attitudes and misuse. Participants in Phase 4 reported that implementing word lists requires some base knowledge, and teacher training such as CELTA does not provide support in terms of implementing word lists and corpus linguistics. This gap between teaching practices and corpus research has been highlighted in the literature, and in response, several researchers (e.g. Frankenberg-Garcia, 2012; Latif, 2020; Mukherjee, 2004; Nesi et al., 2021) have conducted research to explore teacher cognition and perception after training in corpus linguistics. From these studies, it is apparent that there is a lack of awareness of and training in corpus-based applications for language teaching. Teachers must be guided to those resources and trained to incorporate them effectively.

This discussion highlighted the importance of raising awareness about corpus-based word lists among ELT practitioners. Understanding how word lists are compiled and how they can be used is important in two ways. It facilitates using them successfully (leading to positive attitudes), and most importantly, it bridges the gap between word list research and practice.

5.1.3 Evaluating Word Lists for Pedagogical Use

This study set out to address the issues of assessing corpus-based word lists for pedagogical use. The concept of assessing word lists for claims of usefulness received little attention in prior research (Schmitt, 2016) and in studies that were done, results were not conclusive or practical for ELT uses (Dang & Webb, 2016a; Schmitt, 2016), see Section 2.6 for more examples and a discussion of word list evaluation. In Phase 1, the third research question (RQ3) was, *Which measures can be used to evaluate a word list's utility for ELT practitioners, researchers and language learners?*

Studies conducted to evaluate word lists (e.g. Dang & Webb, 2016a; Hernandez, 2017; Newman, 2016; Qi, 2016) and the responses from participants showed that practitioners and researchers take different approaches to evaluating word lists. Practitioners rely on their experience and intuition when choosing a word list or selecting words for teaching. Researchers (and some research-oriented practitioners) focus on the lexical coverage of the list or the published paper that presented the list. Furthermore, practitioners consider whether supplementary materials are available, the reputation of the textbook that includes the list, the suitability of the list for their purpose and their students, and access to the word list itself.

To illustrate, even though Dang and Webb (2016b) excluded the NGSL2 (Browne, 2013) from their evaluation of high-frequency word lists when compiling their EWL, two practitioners in Study 2 reported that they were actually using the NGSL2, and one of them was aware of the criticism of the NGSL2. These practitioners explained that the list developers provided supplementary materials and tools, as well as guidance and support as they target the teacher, not the research community. Researchers seemed to prefer the NGSL1 (Brezina & Gablasova, 2015). Dang and Webb (2016a) incorporated it in the EWL, and one practitioner from Study 2 reported using it for both teaching and research. Similarly, researchers exploring the use of word lists (e.g. Banister, 2016; Burkett, 2015, 2017) found that most EAP teachers and institutions used the AWL, not the AVL, because their programmes and materials had been designed based on it. A detailed discussion of the key measures, as practitioners reported in the Phase 1 survey, is presented in Section 3.3.4.3. They are recited here besides responses related to selecting and evaluating word lists reported during the interviews in Study 2 (see Section 4.5.4.6.).

Lexical coverage has been widely used to assess word lists. While it helps to indicate usefulness, it is not feasible for all practitioners to calculate the lexical coverage of a word list. Calculating and interpreting lexical coverage requires knowledge about corpus linguistics, and it involves making decisions related to the corpus, the type of coverage and the unit of counting. To

assess the validity or suitability of a word list, Schmitt and Schmitt (2020) suggested examining how it is compiled, its contents and how the claims of the list's compilers are validated. While this is a good proposal, many practitioners do not have easy access to such documentation about word lists, which is available to researchers who work in universities. It also seems like it takes a lot of time, especially if someone is comparing several lists for possible use. Such practitioners need an orientation so they can judge the usefulness of a list and decide how they would implement it. This study showed that the proposed tool in its different iterations and supporting documents (Appendices G and M) are approachable and accessible to practitioners who lack information about word lists.

Recent research (e.g. Dang et al., 2020; He & Godfroid, 2019; Simpson-Vlach & Ellis, 2010) has called for teachers' perspectives to be considered when evaluating the usefulness of word lists. Such research has shown that teacher perceptions are good indicators of usefulness. It has been shown that practitioners' experience and perspective reflect usefulness because they are more aware of the teaching context. The participant in this study agreed that evaluating word lists is a complex and challenging task which regular language teachers require more support to accomplish. One practitioner recommended that, rather than evaluating the word list itself, teachers could assess an evaluation or a presentation of the word list, such as the NGSL2's website. However, for maximal effectiveness of implementing word lists, practitioners must be aware of the origins of word lists, and they must analyse their decisions and practices. Most importantly, practitioners need some basic knowledge about word lists to evaluate them accurately.

Finally, it could be claimed that using the proposed tool to evaluate word lists systematically could yield useful insights into those word lists. Practitioners in Phase 4 praised the fact that the tool would push practitioners to analyse their teaching context and it could encourage them to assess the lists they are using. The following section discusses the development and usefulness of the tool.

5.2 The Evaluation Tool of Corpus-based Word Lists

Because of the sheer number of word lists available, practitioners must select a list that suits their purpose and their learners' needs. This study aimed to propose a solution for ELT practitioners who must assess the suitability of corpus-based word lists. The literature review highlighted the paucity of evaluative research that indicate what should be considered before implementing a word list in the language classroom. To the researcher's best knowledge, this is the first study to develop a tool for evaluating word lists for pedagogical use. The tool targets ELT practitioners with varied expertise in corpus-based word lists, and it is based on consultations with experts in the field and potential users of the tool. The data collected in Phases 3 and 4 aimed to assess the proposed tool in its different iterations from the point of view of experts in word list development and use (RQ5) and ELT practitioners (RQ6). The reviews focused on the effectiveness of the tool, its content and efficiency guided by the following questions:

RQ5: To what extent is the proposed tool suitable for its intended purpose, according to expert users and evaluators of corpus-based word lists?

RQ6: To what extent is the proposed tool suitable for its intended purpose, according to potential users of the framework (e.g. language teachers)?

Four key areas emerged from the data that relate directly to the tool. They merit more attention and possible modification:

- evidence of the usefulness of the tool to evaluate word lists
- the suitability of the tool for the target users and the comprehensiveness of its content
- the efficiency of the tool in practice
- the need for a consciousness-raising resource for practitioners who use word lists for teaching ESL

Each of these is discussed in the following sections.

5.2.1 *The Usefulness of the Tool in Achieving its Intended Purpose*

Evaluation checklists and tools outline the criteria to consider, and they function as a resource that simplifies the evaluation process and the recall of information (Hales et al., 2008). During data collection, practitioners reported that they sometimes selected a word list with little or no evaluation. Those selections were based on the popularity, the availability of the list, the reputation of the publisher and whether supplementary materials were available. Zahan and Begum (2013) reported similar reasons for selecting EFL textbooks. Checklists to evaluate textbooks have been developed to address this issue. The proposed tool was developed to help users to select word lists that are suitable for the target context. The participants in this study supplied positive responses regarding the usefulness of the tool in determining the suitability of a corpus-based word list for a teaching context. In Phase 3, the reviewers regarded the tool as useful, but issues related to its efficiency and accessibility required further consideration. Similarly, the quantitative and qualitative data collected from the reviewers in Phase 4 suggested that, in general, the revised version of the tool was useful for evaluating word lists because the tool guides practitioners to reflect on the word lists they use and on their practices in a structured and thorough manner. The tool accommodates the various elements that practitioners must consider when evaluating word lists for pedagogical use. Furthermore, the tool establishes well-informed guidelines that help word list users gain deep insight into a word list's usefulness for users' specific purposes and context.

Nevertheless, some reviewers expressed concerns regarding the effectiveness of the tool for certain practitioners or certain purposes. The most frequently cited purpose when using the tool was to assess a corpus-based word list for use in classroom teaching and learning. In this context, it is the teachers who use the list; thus, assessing the list and reflecting on how it could be implemented would be useful for them. However, the reviews suggested that course and syllabus designers, materials and assessment developers, coordinators and publishers might find the tool more relevant than would regular language teachers because teachers would still need to have expertise in corpus

linguistics to inform their judgements. Researchers (e.g. Çalışkan & Gönen, 2018; Frankenberg-Garcia; Latif, 2020; Mukherjee, 2004) have shown that many language teachers lack knowledge about corpus linguistics, and they rarely use corpus-based tools in the language classroom. To compensate for this, the tool is supplemented with a guide (see Appendix M: A Guide for Corpus-based Word Lists in the Language Classroom) to support practitioners with limited expertise and knowledge about corpus-based word lists. Thus, even if the teachers and practitioners might not fully benefit from the tool, *The Guide* (Appendix M) should help bridge the gap between research and practice. Training in and learning about corpus-based word lists might have an immediate effect in the classroom and the teacher's practice. Course/programme coordinators often have more time to invest in the tool and, most importantly, using the evaluation tool to assess word lists for a particular context would be of direct benefit for them.

The expert reviewers in Study 1 (Phase 4) suggested that the extent to which the tool might inform decisions would depend on the specific teaching context and the role of the practitioner. Some teaching contexts are restrictive; in other words, teaching in them is pre-determined and fixed. Decisions related to selecting a word list for classroom use are often made by course/programme coordinators. Considering the teachers' limited time and the resources needed to complete the evaluation using the tool, as well as the instructions needed for some, using the tool may be a burdensome task. Regular teachers may need to make a tremendous effort to use the tool, yet receive little benefit. Another limitation is that most of the reviewers did not report immediate changes in their teaching practice after using the tool, which is something that a regular teacher would expect. Thus, decisions made after using the tool might not be implemented. It should be noted that it is still unclear how practitioners use word lists to inform materials development, testing and courses design. In testing the tool, the purposes for which the participants used the word list evaluation tool were mostly for classroom learning and teaching. In only one case was a list evaluated for materials development use.

The question was raised whether the tool would lead to conclusions similar to those drawn after using the examined word list. Practitioners (Phase 4) said they lacked sufficient confidence to address this issue because most of the questions in the tool, apart from the last section on word list implementation, concerned the research element of word lists, particularly, developing corpus-based word lists. Practitioners need to consider how they can use these lists in class, how to communicate the lists' content to students and how to use the lists both for teaching and as a self-study tool. The experts in Phase 4 noted that the successful implementation of word lists was based on practitioners' awareness of corpus linguistics and word list development, as well as their expertise in ELT. Thus, even though the tool does not provide conclusive judgements similar to those drawn from using word lists, it provides support and access to information needed for assessing word lists and planning how to use them. It should be noted that pre-use and post-use evaluation requires different procedures, are conducted for different purposes and lead to different conclusions (Tomlinson, 2003).

The tool was designed and tested to be used in all English language teaching and learning contexts. Based on the reviews in Phase 4, the tool was useful for assessing word lists for use in these contexts. However, making claims about the usefulness of the tool in the ESP context would require further testing because assessing specialised, technical word lists requires some level of expertise in the discipline in which it will be used and only one specialised list was evaluated in the present study.

A secondary aim of this study was to highlight the need to raise awareness about corpus-based word lists and promote their usefulness by showing practitioners how to use them effectively. In Study 2, R8 appears to have achieved this aim when they noted that the tool helped them examine word lists from another perspective (see Section 4.5.4.1). They had used word lists to teach specialised vocabulary, but they were not aware of the origins of the vocabulary and how to think critically about the lists they used. Likewise, R7 noted the need to raise awareness of corpus research

and effective implementation of word lists, which the tool, especially its introductory document and the supplementary notes, addresses.

Concerning the supplementary materials, the participants praised the introductory document (Appendix G). They said it would help practitioners understand word lists more. R1, R5 and R3 (with no background in corpus linguistics) reported that the introductory document helped them to understand the key concepts before answering the questions in the tool. All the participants said that they referred to the supporting notes during the evaluation. R10 and R11 both found the example at the end of the tool helpful in explaining the requirements and how to approach the questions. Thus, it could be said that the tool could help to bridge the gap between corpus linguistics research and language teaching by raising awareness and providing accessible resources for ELT practitioners.

While the tool's focus was on lists of individual words, R10 used the tool to evaluate a phraseology list. They found the tool useful in evaluating their list, but they thought the questions about word list development should be revised to reflect the process of developing word lists with multiple units. This is not critical, as R10 and R11 noted that issues related to word list development are not necessary.

I have used the Phrase list by Martinez and Schmitt, and the process I was going through in terms of choosing specific items from the list were exactly the kinds of questions that you are asking in your tool. So, this would be a real example of how I have used a corpus-based list (R10).

In summary, the participants considered the tool a useful means for reflecting on the potential effectiveness of word lists in an ELT context, and they reported that the tool structured this analytical process and fulfilled its purpose. However, because of the complexity of the tool, most participants suggested that regular ELT teachers might not be willing to devote time and energy to the tool unless they can foresee its benefits or receive support. Furthermore, they praised the

supplementary materials. Participants also stressed the importance of the tool's adaptability to practitioners' particular contexts and uses of word lists.

5.2.2 *The Content of the Tool: Accessibility, Comprehensiveness and Relevance*

It was important to ensure that the tool covered all the necessary elements for evaluating word lists and that the content was accessible to the target users (Scriven, 2007). The tool was designed with the non-expert practitioner in mind, so its accessibility to the target audience was of paramount importance. In the development phase (Phase 2), in terms of the wording of the tool, the researcher provided simplified definitions for technical terms, explained challenging concepts related to word lists, and provided supportive notes. This affected the quantity of the information, and how the tool presented that information in terms of accessibility and wording. The content of the tool was based on Nation's (2016) book and other research as well as data gathered from practitioners, but it also reflected the points that Schmitt and Schmitt (2020) recommended for assessing word lists.

It was also a priority in Phase 3 to ensure the accessibility of the tool to the non-expert practitioner. The issues reviewers raised in that phase were addressed by the researcher (see Section 3.9). Key changes included developing an orientation document to introduce the basics of corpus-based word lists and improving the guidance on using the tool.

In Phase 4, the researcher provided the necessary guidance and explanations, but the level of technicality remained challenging. Participants observed that those who are interested in learning about corpus-based word list development may need to spend more time with the tool to absorb those concepts or complete training. However, practitioners with substantial expertise found the tool accessible. Scriven (2007) noted that while evaluation checklists should balance the comprehensiveness of the content, its clarity (accessibility) and the practicality of the tool (in terms of length and time spent on it), priority should be given to clarity. Thus, it could be said that the tool achieved comprehensiveness and clarity at the expense of practicality. Two reviewers suggested that the audience for the tool should be revised to remove teachers and practitioners without a

background in corpus-based research. They might require too much time to engage with the tool, and they might find the content of the tool challenging. Those teachers might not absorb the concepts related to word lists only by reading the notes, and reviewers in Study 1 wondered whether questions 2 and 3 could even be answered. This observation is in line with the identified gap between teacher training and research in corpus linguistics.

In terms of the comprehensiveness of the content, the researcher undertook many changes to the initial version of the tool, based on the issues raised by reviewers in Phase 3. Those changes are presented and justified in Section 3.9. The results from Phase 4 confirmed that the tool covered all the elements that should be considered when evaluating word lists for pedagogical use. There were minor issues in terms of the wording of questions 9 and 15 that were addressed by the researcher. Other issues reported by the reviewers, related to the proficiency levels of the learners and the size of the corpus, were addressed.

5.2.3 *The Efficiency of the Tool*

The key issues that influenced the efficiency of the tool were its structure or format, length and presentation. They arose from the effort to balance the need to include necessary details and the complexity of evaluating word lists against the interest in efficiency. Reviewers noted that the format of the tool was long, heavy with text, complicated and sometimes difficult to follow. However, such details are necessary to promote reflection on word lists because evaluating them is a complex, time-consuming process. Length is a common problem with evaluation tools (e.g. Kashoob, 2018), and Scriven (2007) emphasised that while brevity is desirable, clarity is essential.

In support of the tool, the participants asserted that it was worth the time investment and it would generate valuable returns, particularly for course designers, materials writers, publishers and independent teachers. R2 (Phase 4) argued that the explanatory information in the tool, despite the high level of engagement that it requires, is necessary and in regard to the time a teacher cannot complete a reflection in a mere 10 to 15 minutes. Participants in Phase 4 concluded that, although

the tool took approximately 1 hour to complete and practitioners, regular teachers who often do not have that time, might not be willing to allocate so much time, the benefits of the tool would ensure that the investment was time well spent (see Section 4.5.4.3 for responses related to the efficiency of the tool).

In Phase 3, the researcher significantly revised the physical format of the tool, based on reviewers' feedback. The modifications are noted in Chapter 4, but the key ones are discussed here:

- Following the recommendation to reduce the length of the document and improve the flow of the tool, the researcher presented the main questions and added links to the supplementary notes at the end of the document. This allowed users to analyse the questions without distraction from the explanations and notes. Participants in Phase 4 said they referred to the notes as needed. While the notes were useful to some users, others observed that the need to alternate between the questions and the notes might complicate the flow of the evaluation.
- References were added so that users would feel empowered to explore the literature further.
- Although not all reviewers commented on the rating scale in Phase 3, the recommendation to delete the scale was adopted. Some reviewers had trouble rating certain items, which may be misleading. Two reviewers in Phase 4 agreed on use of the open-ended questions, though one reviewer suggested that feedback may not be beneficial.
- A balanced level of technicality was necessary to present information about word lists accurately and in a manner simple enough for non-expert users to understand—but not so reductive that expert users would be bored. An explanation and definitions were added, which participants in Phase 4 found useful and clear. However, the concepts of

corpus linguistics are abstract and cannot be fully understood by users who have little or no background in corpus linguistics.

- Presenting the tool in a Microsoft Word document was the most-cited issue in Phase 4. In its current form, the presentation of the material requires improvement, and housing the tool and all associated content online is a popular proposal that is expected to improve the efficiency of the tool. If there had been additional time for this project, this recommendation would have been followed. This would improve the efficiency of the tool in terms of moving from the notes to the questions and the presentation of the text and questions. Nevertheless, using the tool requires some learning on the part of the non-expert user (mostly regular language teacher). Presenting The Guide (Appendix M) in video tutorials may enhance the learnability of the content of these documents, but the user would still need to engage with the content of the video.
- Others recommended making a short video tutorial to illustrate how to complete the tool so users could see its sequence.

5.2.3.1 The Efficiency of the Tool in terms of Practitioners' Time Resources. The

participants in Phase 4 reported that they took between 50 to 90 minutes to complete an evaluation using the proposed tool.

The typical classroom teacher often does not have time to allocate to such an assessment. Furthermore, they may not benefit directly from the results yielded by the tool because they often do not make the decisions regarding vocabulary selection or resources used. Thus, in its current format, the tool may not be useful or practical for them. Another issue that requires consideration is that using the tool requires some investment in learning about corpus-based word lists if the user does not have this background. This adds to the time that would need to be allocated to using the tool.

Nevertheless, it is important to note that the participants considered that the tool was worth the time invested and would generate valuable returns if time resources were available. It was observed that the tool would be more useful for certain groups of practitioners, particularly for course designers, materials writers, publishers and independent teachers. In fact, it was noted that evaluating word lists is a complex task that cannot be conducted in 10-15 minutes. It requires analysing the examined list and the teaching setting, and then considering the main elements (criteria) with regard to the setting. Following this, decisions concerning how to implement the list and addressing the list's issues need to be considered.

A proposed solution to this issue is to construct two versions of the tool: One for the expert practitioner in which all the relevant questions are presented and the supporting notes of which these experts are aware are deleted. The other would be for non-expert practitioners, in which complicated questions about the unit of counting, the corpus and the approaches to constructing word lists would be deleted while focussing on analysing the context and the target learners; considering the last section related to implementing word lists, issues such as the size of the list and supplementary materials would be assessed. In the latter case, the evaluation tool may not be as comprehensive but it would be more practical considering regular teachers' time resources. This version would help teachers to think critically about their lists and how to implement them, but would not necessarily lead to an evaluation of the suitability of the examined list. Such practitioners would benefit from learning about word lists, where they come from and how they can be used. This content would be included in *The Guide* (Appendix M) document, which could be in its current format or could be moved online as digitalised content.

One of the solutions suggested by the reviewers in Phase 4 was to specify the time needed for each section. This solution is not possible because responding to these sections might take different amounts of time for different users.

In summary, the proposed tool requires a time investment. Nonetheless, while tools such as Text Inspector (2021) and LEX Tutor (Cobb, 2015), which are widely used by teachers, are handy, easy to use and have been proven to be useful, these tools are pre-programmed to provide certain information and need to be used with specific word lists. Using these tools does not allow teachers to be critical about the word lists and does not guide teachers to question the lists they are using. In this case, ease of use is not the best feature and considerations about time investment can be made by teachers based on what they can accomplish as result of the assessment outcome. In this study, although the reviewers acknowledged that using the tool was time consuming, they observed that word list evaluation is not a process that can be done in 10-15 minutes.

5.3 A Guide for Corpus-Based Word Lists in the Language Classroom (Appendix M)

Two main conclusions were derived from the previous discussion. First, the cumulative reviews suggested that the tool met its goal, but because of the complexity of evaluating word lists, practitioners who lack expertise and knowledge in corpus linguistics may find the tool challenging. Second, data on practitioners' beliefs and practices (cognition) in ELT related to word lists showed there was a lack of awareness of word lists as well as the need for some background on corpus linguistics if corpus-based word lists are to be implemented successfully. Some interviewees noted this gap and referred to their experience in learning about corpus linguistics. Researchers have found that ELT teachers are not sufficiently aware of corpus-based applications and how to use them (e.g. Çalışkan & Gönen, 2018; Frankenberg-Garcia, 2012; Latif, 2020; Mukherjee, 2004). This may deprive teachers of the benefits of such resources. Those researchers noted that raising awareness about corpus-based word lists led to positive attitudes, and they encouraged teachers to use corpus tools. To bridge the gap between research in corpus linguistics and teaching practice and to ensure the optimal usefulness of the tool, this study aimed to raise awareness about corpus-based word lists in ELT through a stand-alone self-access consciousness awareness document (See Appendix M).

The Guide targets teachers who use (or plan to use) corpus-based word lists for vocabulary instruction and lack sufficient knowledge and expertise in corpus linguistics. It aims to support language teachers in learning about word lists and to guide them to think critically about issues that influence their use of word lists. These include the size of the word list and the availability of supplementary materials (adapted from Section 4: Implementing Word Lists in the tool). That document examines implementation-related issues, functions as a shortened and simplified version of the proposed tool and aims to reconceptualise beliefs towards word lists through raising awareness and critical reflection on the use of word lists for language learning. This guide should be subjected to validation and improvement by a similar process, but that is beyond the scope of this study.

5.4 Summary

This study aimed to provide an evaluation tool that ELT practitioners can use to assess the suitability of word lists for their context and purpose. This chapter confirms the usefulness of the proposed tool for evaluating pedagogical word lists for ELT. A secondary aim that emerged during the study was to highlight the need to raise awareness of corpus-based word lists among ELT practitioners to help bridge the gap between corpus linguistics research and teaching practices. To achieve this, the study provides a resource, *The Guide* (Appendix M). This chapter shows that using a corpus-based word list for vocabulary instruction in ELT is multifaceted. The value of word lists has been influenced by teaching approaches throughout the centuries and the advances in corpus analysis tools. The status of word lists and other contextual factors shaped practices related to word lists. Knowledge of word lists and using them in ELT is still limited, despite the benefits of such lists for vocabulary instruction. The researcher developed an offshoot conscious raising tool for language teachers who are not aware of word lists. It is hoped that this tool and its supplementary material will raise awareness and guide practitioners to use word lists successfully.

6 CHAPTER 6. CONCLUSION

This study was motivated by the paucity of accessible guidance on evaluating corpus-based word lists that accounts for the practitioners' perspective and the considerations that should be made before implementing a word list in an ELT context. The study comprised developing and testing an evaluation tool as well as exploring practices and views related to corpus-based word lists. The findings in Chapter 5 were structured around two main themes: practices and views related to corpus-based word lists in ELT and the usefulness of the evaluation tool from the point of view of practitioners and experts. This chapter reflects on the research process on a broader level, and it summarises the answers to the research questions that guided this study. It highlights the contributions and implications for pedagogy and research, and it notes the limitations of the study. Finally, it offers recommendations for future research and practice.

6.1 Reflection on the Research Process and Results

One beneficial contribution of corpus linguistics to language teaching is corpus-based word lists, which can direct practitioners and learners towards important words that are worth their limited instructional time (Folse & Youngblood, 2017). Corpus-based word lists as we know them today went through multiple and overlapping stages. The review of the chronology of word list development in Chapter 2 summarises this journey. It also shows how issues from the early 20th century among members of the Vocabulary Control Movement related to the unit of counting and approaches to making word lists for example are still debated (Section 2.4 Issues in Corpus-based Word List Construction). It also reflects the intertwined and complicated value of word lists in ELT, and how they are influenced by advances in corpus linguistics and approaches to language teaching over the centuries. Interestingly, the findings derived from this study show how these negative and positive perceptions towards word lists still exist. In light of teacher cognition, the study explored these perceptions and practices. This exploration shed light on the gap between corpus linguistics research and ELT practice, and in response, this research attempted to raise awareness of corpus-based word lists among ELT practitioners by proposing *The Guide* (Appendix M).

Looking at the wider menu of corpus-based lists available, ELT practitioners must ensure they use the most suitable list. Because of the lack of accessible guidance on how to evaluate a corpus-based word list for ELT contexts, this thesis aimed to develop an evaluation tool for use by ELT practitioners based on Nation's (2016) work adapted and reviewed through consultations with ELT practitioners and experts. Before developing a well-informed tool, it was necessary to understand the status of word lists in ELT contexts and to decide how the tool would be developed and validated.

There was limited published research on using word lists for ELT at the start of this study. Thus, it was not clear if word lists were used or if they are worth the investigation in the context of ELT practice. Published articles introducing new lists loosely described factors related to corpus-based word lists for pedagogical uses. These factors included their purpose(s) of the list, the target audience and how to use the list. In Phase 1 of this project, the initial explorative phase, 74 responses were collected from ELT practitioners, researchers and learners through an online survey designed and administered to explore the following questions:

RQ1: What distinct uses are made of corpus-based word lists by practitioners, researchers and learners involved in L2 teaching and learning?

RQ2: What methods are used to generate pedagogical corpus-based word lists?

RQ3: Which measures can be used to evaluate a word list's utility for ELT practitioners, researchers and language learners?

RQ4: How are word lists perceived by ELT practitioners, researchers, and learners?

The results showed that word lists are indeed used; 80% reported using word lists while 20% reported not using them. The Phase 1 survey showed that practitioners, learners and researchers were using word lists for different purposes, mostly for classroom language teaching. This emphasised the significance of the evaluation tool for the ELT community, and it helped to identify the target audience and contexts. In terms of how word lists should be made (RQ2), the survey respondents thought that statistical measures (i.e. frequency, range and dispersion) based on corpus

evidence were important, but they stressed the value of the expert's subjective modification to tailor the lists to the needs of the users. Most respondents thought that the purpose of a list should guide its development. This aligns with Nation's (2016) observation that pedagogical word lists should be tailored, and practitioners should have confidence in their perspective. Recently, Dang et al. (2020) incorporated teachers' perceptions in evaluating word lists, and they found that their predictions were accurate. With more studies like this and increased awareness, teachers can become more confident in their perspectives. The third and fourth research questions aimed at exploring how practitioners chose and evaluated word lists for ELT, and what they thought about word lists. These two questions led to a focus on the gap between word list research and practice, and it was revisited throughout the study. A general conclusion was that a list should be suitable for the purpose and context and of a manageable size. Practitioners reported problems devoting the time and resources needed to make one. There were inconsistent criteria for including words, different units of counting, words out of context and a lack of supplementary material for the word lists. The most remarkable finding was the negative attitudes toward word lists which often came from misusing them. These revealed potential drawbacks related to word list use. Even though RQ4 was posed only in the first phase, the participants in the interviews in Phases 3 and 4 proffered their attitudes and perspectives of word lists.

The next step was to develop the tool. In Phase 2, the major themes of the survey and the literature review formed the basis for adapting Nation's (2016) framework. The tool was designed based on Stufflebeam's (2000) guidelines and on the principle that a single word list is unlikely to be useful in all contexts and for all learners, as indicated by Nation (2016). At that point, the tool was called a framework and it had a rubric. The second part of Chapter 3 relates to how the initial version of the tool was designed, reviewed and revised. Developing and testing the tool in its different iterations addressed the following research questions:

RQ5: To what extent is the proposed tool suitable for its intended purpose, according to expert users and evaluators of corpus-based word lists?

RQ6: To what extent is the proposed tool suitable for its intended purpose, according to potential users of the framework (e.g. language teachers)?

The next two phases used different methods to review and revise the tool from the perspectives of experts and practitioners. The literature review on developing and validating evaluation checklists in different disciplines dealt with numeric values and lists of items to be checked.

In Phase 3, the tool was reviewed and revised by a panel of experts and practitioners focussing on the tool's clarity of wording, relevance, accessibility and practicality. The reviewers found the tool to be useful but its accessibility to non-expert practitioners and its structure needed to be revised. The next iteration incorporated the reviewers' feedback. Then, two studies were conducted to test the tool. Study 1 was a critical review by another expert panel. In Study 2, the tool was used to evaluate the suitability of a word list from the perspective of ELT practitioners in real-world scenarios.

The participants thought the tool would allow practitioners to make judgements about word lists and adapt them for maximal usefulness, even though the tool might not provide conclusive answers. They also observed that, because of the complexity of evaluating word lists, practitioners who lacked expertise and knowledge of corpus linguistics might find the tool challenging. The qualitative and quantitative results showed that knowledge about how corpus-based word lists are constructed was necessary for implementing them successfully. To bridge the gap between research into corpus linguistics and ELT teaching experience, the researcher developed a supplementary resource to help practitioners learn about word lists and to engage them in thinking critically about issues that directly influence their use of word lists (Appendix M). Minor adjustments were made to the tool, based on the feedback from Phase 4, resulting in its current version (Appendix L). This iteration is expected to be useful for its intended purpose, but it needs further testing on a scale which is beyond the scope of this study.

Considering the pragmatic philosophy to research, this study employed different techniques for data collection and analysis and conducted them in different settings and with different participants. The study not only filled in a gap in the field of using word lists for language teaching but also a gap in real educational settings by focusing on a real-world problem and proposing solutions. The potential impact of this tool on the use of word lists might not be immediate, and the tool needs further refinement, which is beyond the scope of this study. Certainly, improving the evaluation and implementation of word lists will require effort, and even though the current study involved feedback from practitioners, the tool needs more development.

6.2 Contributions of This Study

The principal contribution of this study is the evaluation tool to help ELT practitioners assess the suitability of a corpus-based word list for their purpose and context. The tool is aimed at practitioners with varying degrees of familiarity with corpus linguistics and word lists. It should be of interest to a wide range of ELT practitioners, given the popularity of word lists, the different ways word lists can inform L2 vocabulary instruction and research, and the need for tools to assess their usefulness in specific contexts. Assessing the suitability of a given word list and planning effective implementation would help to reduce the increasing number of published word lists in the field of applied linguistics. Despite the proliferation of word lists in recent years, this study, as well as previous studies (e.g. Burkett, 2015; 2017), found that many practitioners and programme coordinators considered popular lists unsuitable for their purposes and contexts. One reason for this is that no vocabulary list can be suitable for all learners in all contexts; instead of making a context-specific word list that is appropriate for particular learners and contexts, the proposed tool could assist by adapting an existing word list to the target purpose and context.

Another problem this study addressed is that many of the published lists have been constructed using inadequate validation evidence or a clear intended purpose and context (Schmitt, 2016). When validation has been conducted, it often involves calculating lexical coverage. Lexical coverage is a good indicator of a word list's usefulness, but there are problems (see Section 2.6.3 for

a discussion of these issues). For researchers who are involved in making word lists, this tool contributes by providing more rigorous and systematic procedures for the evaluation of word lists during the process of constructing new word lists for pedagogical use. These include the specification of a list's purpose, the intended learners and the educational context, the particular type of vocabulary being targeted and ensuring that the lists are suitable with regard to these elements. These elements were highlighted in Nation's (2016) book, but the proposed tool provides this information as an organised set of procedures, and focusses on the elements related to language teaching. It adds factors that need to be considered when evaluating or making word lists for pedagogical use based on data collected from practitioners, and addresses issues in real-world use reported by practitioners who use word lists.

This study contributes to bringing applied linguistics research and ELT practice closer. The purpose of the tool overlaps with the aim of Nation's (2016) framework, but he acknowledged that his work was aimed at people who understood vocabulary. Word lists are one of the contributions of corpus linguistics research, but practitioners need to have knowledge about word lists and how to assess them for maximal usefulness. Applied linguistics research should be concerned not only with understanding practice, but should also involve practice.

Another contribution of this study is that it explored practices and views related to word lists in ELT. Research on how word lists are used and viewed has been limited despite the recent renaissance of interest in word lists. Findings from the different phases of this study and data from the various participants have contributed to our understanding of the current status and views of word lists. The findings emphasised the need for greater awareness of how word lists are developed, selected and used in different contexts. Furthermore, reviewing the literature and collecting data from practitioners indicated that there is still a lack of accessible and freely available resources pertaining to word lists for practitioners. This study calls for support for the end users of word lists; for example, the provision of detailed users' manuals. The tool includes two supplementary stand-alone documents and supporting information for learning about word lists (see Appendices G and

M). It is hoped that these materials will lead to using word lists more efficiently and mitigating the effects of negative experiences with them. Furthermore, word lists have suffered from misconceptions and issues for a long time, as highlighted by Folse, (2004). This study attempted to explain these misconceptions and misuses from the perspective of teacher cognition and to propose solutions. It is hoped that this study will contribute to reviving interest in word lists and to improving their usefulness in practice.

The findings of the study confirmed that there is a gap between corpus/applied linguistics research and the practitioners' perspective. An underlying theme that was repeated in the different phases of this project was that practitioners' voices were often unsolicited and/or lacked authority in word list research, even though they are the end users. This study contributes to improving the usefulness and applicability of research related to word lists by giving practitioners authority and taking their perspective into consideration with regard to judging what is useful for their learners. In the process of developing word lists, great emphasis has been placed on frequency to determine the value of words. Researchers working on word lists for pedagogical purposes (most of the published word lists are directly or indirectly targeted at language learning) need to be aware that, even though frequency is a valuable criterion, the teachers' perspective is just as valuable. The teachers' perspective has been taken into consideration in a small number of recent studies evaluating word lists, such as the work of Dang et al. (2020) and Omidian et al. (2017), see Section 2.6.4. for more details about these studies. Hence, using the proposed evaluation tool by researchers would systemise the process of involving teachers in the evaluation of word lists, which would ensure that all the elements that need to be considered are taken into account.

Evaluation tools help practitioners assess practices and resources. The literature on developing evaluation checklists and tools revealed that these systems often ignore the regular practitioner (Kashoob, 2018) and often give a numeric value to the examined object. The study has filled a gap in the methods of developing and testing evaluation tools, using an approach that involves practitioners in their design and considering the target users and context.

6.3 Implications for ELT Research and Practice

The findings of this study have several pedagogical implications for ELT practice: First, the corpus-based word list evaluation tool shows great potential as a self-assessment instrument for practitioners who want to determine the suitability of a word list for pedagogical use in their context and to achieve their purpose. For practitioners who are using or planning to use word lists, this tool provides insight into the key elements that should be considered when choosing a word list. The tool helps practitioners to be critical about their choices, and it encourages them to adapt selected word lists to their teaching needs. Even though some practitioners might find the tool challenging and others cannot change their teaching practices after using the tool, the tool and its supplementary materials raise awareness about word lists which should help in using and adapting word lists effectively (even in restrictive teaching environments).

The supplementary materials developed with the proposed tool are offered as resources to raise awareness about word lists. They include the *Word Lists Table* (Appendix A), which was adapted from Burkett (2017). It is a good resource for finding information about word lists, but it should be updated regularly. The *Introduction to Word Lists* (Appendix G) presents the basics of corpus-based word lists in one page, and *The Guide* (Appendix M), an offshoot of the original tool, is meant for language teachers who are not familiar with word lists.

The findings related to the practices and perceptions of word lists, and the challenges of using them could also shape decisions for teacher training programs. This study reflects the lack of knowledge about corpus-based tools and resources among language teachers, which may deprive them of the benefits of these resources. Those programs must address teachers' needs. In fact, despite the reintroduction of vocabulary instruction, it is still neglected in some parts of the world, including the Middle East. Researchers and those involved in teacher development must address the needs of in-service teachers and provide resources and training. As researchers (e.g. Le Foll, 2021) heeded the call from Römer (2006) and others to bridge the gap between corpus linguistics research and teachers, greater awareness of word lists and support for practitioners who use them is needed.

This study also has implications for researchers involved in word list development and use. The tool can inform corpus-based word list development for pedagogical purposes. It highlights issues and considerations that must be noted by list compilers. For example, list compilers should provide a guide to their lists to help practitioners identify the purpose of the list and how it can be used. One example is the guide for the Phrasal Expressions List (the PHRASE List) by Martinez and Schmitt (2012). Compilers can divide a list into subsections for more practical implementation, provide different versions of units of counting to address different contexts, consult or allow for teachers' perspective on the usefulness of the words, offer resources and supplementary materials based on the list, and consider factors and issues related to different teaching contexts.

It is hoped that this study is a worthwhile contribution to the literature on the development of evaluation tools and that it helps to address the process of devising and validating such tools, including checklists.

6.4 Limitations of This Study

The results suggest that the tool for evaluating corpus-based word lists has met its intended purpose, yet it is a new instrument that still requires repeated revisions and more testing in different ELT contexts to validate it more thoroughly. As with all evaluation checklists (Stufflebeam, 2000), several years of field-testing and use are required to adequately validate and refine the tool. However, that is beyond the scope of a PhD project. Because of time constraints, it was not possible to conduct an additional review of the tool by practitioners after the modifications and improvements made in Phase 4. However, these were minor modifications, mostly structural changes that should not affect the usefulness of the tool.

A second limitation was the small number of participants who reviewed the tool in Phases 3 and 4. The eligibility criteria required the reviewers to be familiar with word lists. Some practitioners who use word lists did not consider themselves qualified to be reviewers. Moreover, the task of reviewing the tool required considerable time and effort, and that is another reason that participants were so few. Several prospective participants declined the invitations in Phases 3 and 4

because they could not allocate an hour or more to the study (see Section 4.4.1 for potential participants' reasons to decline participation). Nonetheless, those who did take part in Phases 3 and 4 volunteered their valuable time, and their cooperation and interest in this research are gratifying.

The nature and size of the sample make it difficult to draw generalised observations about the findings of the study. However, it is important to note that the primary purpose of the study was to explore the use of word lists in ELT and to develop the evaluation tool. The reviews and feedback reflected both the potential benefits of the tool and its drawbacks, and this helped to improve the quality and features of the tool. Furthermore, responses on the use of word lists confirmed the gap between research and practice, but were not sufficient to make claims about how different groups of practitioners use or view word lists.

Finally, as for many researchers around the world, the COVID-19 pandemic had an impact on this study. It affected the last phase of testing the tool. The plan to conduct face-to-face workshops to orient practitioners and discuss the use of word lists was cancelled. The pandemic also slowed the process of interviews, restricted interaction with the participants and limited the depth of analysing the documents. Data collection in Study 2, therefore, focused on practitioners who had some background in corpus linguistics or word lists. It would have been more instructive to train practitioners with limited prior expertise with word lists and then test the tool to explore their attitudes before and after the training (Plan A, see Section 4.9.). Shifting to online interviews is likely to have affected the amount of information and specificity of detail provided by participants.

Nevertheless, it is important to note that virtual qualitative data collection could be useful during a pandemic if used effectively and if its quality were ensured (Lobe et al., 2020). Online data collection enabled me to interview practitioners from different parts of the world. I was flexible in scheduling interviews, using different platforms that were accessible to the participants. I also used the share screen mode to discuss the documents during the interviews. I opened the camera to enhance interaction, but I left it to the participants to do so. I sent participants an outline of the interview and key points to be discussed beforehand to enhance communication and help them

prepare for the interview. I exchanged emails for clarification and explanations before and after the interviews, and for two participants I conducted the interviews in two sessions so they could elaborate on some points (they preferred the interviews rather than sending emails).

6.5 Recommendations for Future Research

Despite these limitations, I believe the evaluation tool proposed in this study will allow practitioners to reflect on using word lists in their contexts and raise awareness of corpus-based word list development and use in ELT. I invite other researchers and ELT practitioners to contribute their effort and expertise to evaluate the tool further. The tool is a work in progress, so additional research is needed to examine the tool: (a) in different ELT contexts, (b) by practitioners with different roles, (c) for different pedagogical purposes and (d) using various types of word lists. This would confirm the appropriateness and utility of the evaluation tool. As discussed in Section 5.2.1, new versions of the tool can be developed and tested to address the needs and circumstances of different groups of practitioners.

The interviewees did not express any intentions to change their practices based on using the tool, or to revise their decision-making about developing course materials. However, course design or vocabulary instruction, in general, might have reflected the tool's usefulness in changing teaching practices.

Similarly, I could not follow up on the impact of *The Guide* (Appendix M) and the introductory materials (Appendix G) on raising awareness about corpus-based word lists and their usefulness for pedagogy. It would have been interesting to explore how the tool is used in long-term practice and decision-making. This is an area where further research may influence teacher training, the design of courses and additional materials and the use of word lists itself.

Previous research reported positive attitudes towards corpus analysis tools after training, but, apart from Latif (2020), they did not follow up on the practitioners' efforts. More longitudinal studies are needed to explore the effect on teachers' attitudes and practices from raising awareness and training the teachers on corpus linguistics. Recently many resources like Corpus for Schools

(2021) and *Creating Corpus-Informed Materials for the English as a Foreign Language Classroom* (Le Foll, 2021) have been published to target ELT practitioners.

The scope of this study was restricted to exploring teachers' uses, attitudes and beliefs concerning word lists. The experiences and attitudes of learners and of practitioners who do not use word lists were not a focus. Although the first phase included learners and non-users, it did not follow up on their responses. It would be interesting to explore how useful the tool is with those who have never used word lists or have negative attitudes towards them. As observed in Chapter 3, there were limitations in the design of the survey. It did not identify the specific roles of the respondents to enable differentiation amongst them in terms of how they used word lists or their attitudes towards them. Future researchers are invited to identify the ways in which different practitioners use and view word lists.

When discussing the use of research tools, it is worth remembering that surveys capture respondents' recollections of their experiences. As they are not observations of practice, the accuracy of the data respondents provide can be affected by memory. Classroom observations and interviews with practitioners who are using word lists regarding how they are used would reflect the status of word lists.

Future research can use this model for designing and testing evaluation tools. Many evaluation checklists are based on numerical rubrics, which forces the evaluator to rate an aspect. Language teaching and learning contexts are complex so word lists, like many pedagogical resources which are used in conjunction with other tools, are difficult to rate numerically. The current study's approach to developing and testing the tool based on consultations with the practitioners can be adopted in future studies.

6.6 Conclusion

The present research developed and tested an evaluation tool for use by ELT practitioners, particularly those who lack expertise in corpus linguistics and word list development. Despite its limitations, the tool has fulfilled its purpose of guiding practitioners through the process of assessing

the suitability of lists and reflecting on their effective implementation. Besides the potential usefulness of the tool, this study highlights the importance of evaluating pedagogical resources and practices, and it contributes to research on the development of evaluation tools. Finally, this thesis makes a valuable contribution to research on teacher cognition and including practitioners' voices in research. It attempts to bridge the gap between corpus linguistics and ELT practice by raising awareness about corpus-based word lists, and highlighting the importance of taking the practitioners' perspectives into account, particularly in the development, adaptation and implementation of corpus-based word lists. It sheds valuable light on practices and views related to corpus-based word lists in the context of ELT.

References

- Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon*. John Wiley & Sons.
- Aizawa, K., & Iso, T. (2007). Estimating word difficulty: The divergence from frequency levels. *ARELE: Annual Review of English Language Education in Japan*, 18, 111–120.
- Aluthman, E. S. (2017). Compiling an OPEC word list: A corpus-informed lexical analysis. *International Journal of Applied Linguistics and English Literature*, 6(2), 78–91.
<https://doi.org/10.7575/aiac.ijalel.v.6n.2p.78>
- Anthony, L. (2014). *AntWordProfiler* [Computer software].
<https://www.laurenceanthony.net/software/antwordprofiler/>
- Archer, D. (2009). *What's in a word-list?: Investigating Word Frequency and Keyword Extraction*. Ashgate. <https://doi.org/10.1093/llc/fqp041>
- Babbie, E. (2010). *The practice of social research methods*. Cengage Learning.
- BALEAP. (2008). *The Competency framework for teachers of EAP*.
<https://www.baleap.org/resources/baleap-publications>
- BALEAP. (2013). *The BALEAP can do framework for EAP syllabus design and assessment*.
<https://www.baleap.org/resources/baleap-publications>
- BALEAP. (2014). *The TEAP accreditation scheme handbook*.
<https://www.baleap.org/resources/baleap-publications>
- Ball, H. L. (2019). Conducting Online Surveys. *Journal of Human Lactation*, 35(3), 413–417.
<https://doi.org/10.1177/0890334419848734>
- Banister, C. (2016). The Academic Word List: Exploring teacher practices, attitudes and beliefs through a web-based survey and interviews. *Journal of Teaching English for Specific and Academic Purposes*, 4(2), 309–325.
- Bardel, C., Gudmundson, A., & Lindqvist, C. (2012). Aspects of lexical sophistication in advanced learners' oral production: Vocabulary acquisition and use in L2 French and Italian. *Studies in Second Language Acquisition*, 34(2), 269–290. <https://doi.org/10.1017/S0272263112000058>

- Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography*, 6(4), 253–279. <https://doi.org/10.1093/ijl/6.4.253>
- Beck, I. L., McKeown, M. G., & Kucan, L. (2013). Choosing Words to Teach. In *Bringing words to life: Robust vocabulary instruction* (2nd ed., pp. 19–39). Guilford Press.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371–405. <https://doi.org/10.1093/applin/25.3.371>
- Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, 36(2), 81–109. <https://doi.org/10.1017/S0261444803001903>
- Borg, S. (2006). The distinctive characteristics of foreign language teachers. *Language Teaching Research*, 10(1), 3–31. <https://doi.org/10.1191/1362168806lr182oa>
- Boulton, A. (2009, July 20–23). *Corpora for all? Learning styles and data-driven learning* [Conference paper]. 5th Corpus Linguistics Conference, Liverpool, United Kingdom. <https://hal.archives-ouvertes.fr/hal-00384914>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Breeze, R. (2013). Lexical bundles across four legal genres. *International Journal of Corpus Linguistics*, 18(2), 229–253. <https://doi.org/10.1075/ijcl.18.2.03bre>
- Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics*, 36(1), 1–22. <https://doi.org/10.1093/applin/amt018>
- Brezina, V., & Gablasova, D. (2017). How to produce vocabulary lists? Issues of definition, selection and pedagogical aims. A response to Gabriele Stein. *Applied Linguistics*, 38(5), 764–767. <https://doi.org/10.1093/applin/amx022>
- Browne, C. (2013). The new general service list: Celebrating 60 years of vocabulary learning. *The Language Teacher*, 37(4), 13–16.

- Browne, C. (2016, September 12–14). *The NGSL, NAWL, TSL, and BSL Word Lists: Corpus-based Word Lists and Tools for Success* [Conference paper]. Vocab@Tokyo Conference, Tokyo: Meiji Gakuin University.
http://vlijournal.org/vocabattokyo/vocabattokyo_handbook_2016.pdf
- Browne, C., Culligan, B., & Phillips, J. (2013). *The New Academic Word List*.
<http://www.newgeneralservicelist.org/>
- Browne, C., Culligan, B., & Phillips, J. (2017). *The New General Service List-Spoken*.
- Brysbaert, M., Stevens, M., Mander, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, 7, 1116.
<https://doi.org/10.3389/fpsyg.2016.01116>
- Burch, B., Egbert, J., & Biber, D. (2017). Measuring and interpreting lexical dispersion in corpus linguistics. *Journal of Research Design and Statistics in Linguistics and Communication Science*, 3(2), 189–216. <https://doi.org/10.1558/jrds.33066>
- Burkett, T. H. (2015). An Investigation into the use of frequency vocabulary lists in university intensive English programs. *International Journal of Bilingual & Multilingual Teachers of English*, 3(02), 71–83. <https://doi.org/10.12785/ijbmte/030202>
- Burkett, T. H. (2017). *An Investigation into the use of word lists in university foundation programs in the United Arab Emirates* [Doctoral dissertation, University of Exeter].
<https://ore.exeter.ac.uk/repository/handle/10871/33151?show=full>
- Çalışkan, G., & Gönen, S. İ. K. (2018). Training teachers on corpus-based language pedagogy: Perceptions on using concordance lines in vocabulary instruction. *Dil ve Dilbilimi Çalışmaları Dergisi*, 14(4), 190–210.
- Campion, M. E., & Elley, W. B. (1971). *An academic vocabulary list*. New Zealand Council for Educational Research.

- Capel, A. (2012). Completing the English vocabulary profile: C1 and C2 vocabulary. *English Profile Journal*, 3. <https://doi.org/10.1017/S2041536212000013>
- Carroll, J. B., Davies, P., & Richman, B. (1971). *The American Heritage word frequency book*. Houghton Mifflin Boston.
- Carter, R., & McCarthy, B. (1988). *Vocabulary and Language Teaching*. Longman Press.
- Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching*, 52(4), 460–475. <https://doi.org/10.1017/S0261444819000089>
- Chen, Y. H. (2011). To wordlist or not to wordlist? The dilemma and challenges for language learning and testing.
- Chung, T. M., & Nation, I. S. P. (2004). Identifying technical vocabulary. *System*, 32(2), 251–263. <https://doi.org/10.1016/j.system.2003.11.008>
- Cobb, T. (2015). VocabProfile Home [Online search engine]. <http://www.lex tutor.ca/vp/>
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (5th ed.). Routledge. <https://doi.org/10.4324/9781315456539>
- Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, 32, 45. <https://doi.org/10.1017/S0267190512000074>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238. <https://doi.org/10.2307/3587951>
- Coxhead, A. (2011). The academic word list 10 years on: Research and teaching implications. *TESOL Quarterly*, 45(2), 355–362. <https://doi.org/10.5054/tq.2011.254528>
- Coxhead, A. (2012). Specialised vocabulary in secondary school classrooms: Teachers' perspectives. *Teaching language to learners of different age groups*, 194–205.
- Coxhead, A. (2018). *Vocabulary and English for Specific Purposes research: Quantitative and qualitative perspectives*. Routledge. <https://doi.org/https://doi.org/10.4324/9781315146478>

- Creamer, E. G. (2017). *An introduction to fully integrated mixed methods research*. Sage Publications.
<https://doi.org/10.4135/9781071802823>
- Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research* (3rd ed.). Sage Publications.
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.
- Curry, N., Love, R., & Goodman, O. (2020). Adverbs on the move: Investigating publisher application of corpus research on recent language change to ELT coursebook development. *Corpora*.
- Dang, T. N. Y. (2018). A hard science spoken word list. *ITL-International Journal of Applied Linguistics*, 169(1), 44–71. <https://doi.org/10.1075/itl.00006.dan>
- Dang, T. N. Y. (2020). Corpus-based word lists in second language vocabulary research, learning, and teaching. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 288–303). Routledge. <https://doi.org/10.4324/9780429291586-19>
- Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, 33, 66–76. <https://doi.org/10.1016/j.esp.2013.08.001>
- Dang, T. N. Y., & Webb, S. (2016a). Evaluating lists of high-frequency words. *ITL-International Journal of Applied Linguistics*, 167(2), 132–158. <https://doi.org/10.1075/itl.167.2.02dan>
- Dang, T. N. Y., & Webb, S. (2016b). Making an essential word list. In I. S. P. Nation (Ed.), *Making and using word lists for language learning and testing* (pp. 153–167). John Benjamins.
- Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The Academic Spoken word list. *Language Learning*, 67(4), 959–997. <https://doi.org/10.1111/lang.12253>
- Dang, T. N. Y., Webb, S., & Coxhead, A. (2020). Evaluating lists of high-frequency words: Teachers' and learners' perspectives. *Language Teaching Research*.
<https://doi.org/10.1177/1362168820911189>
- Davies, M. (2008). *The Corpus of Contemporary American English (COCA): One billion words, 1990–2019* <https://www.english-corpora.org/coca/>

- Davies, M. (2021). The TV and movies corpora: Design, construction, and use. *International Journal of Corpus Linguistics*, 26(1), 10–37. <https://doi.org/10.1075/ijcl.00035.dav>
- de Chazal, E. (2013). The general–specific debate in EAP: Which case is the most convincing for most contexts? *Journal of Second Language Teaching & Research*, 2(1), 135–148.
- Ding, C., & Reynolds, B. L. (2018). In I. S. Paul Nation: Making and Using Word Lists for Language Learning and Testing. *Applied Linguistics*, 39(4), 602–606.
<https://doi.org/10.1093/applin/amx052>
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.
- Durrant, P. (2014). Discipline and level specificity in university students' written vocabulary. *Applied Linguistics*, 35(3), 328–356. <https://doi.org/10.1093/applin/amt016>
- Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes*, 43, 49–61. <https://doi.org/10.1016/j.esp.2016.01.004>
- Egbert, J. (2018, February 20). Frequency is overrated: Using text dispersion to measure word importance [Lecture]. University of Birmingham, UK.
- Farr, F. (2008). Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users. *Language Awareness*, 17(1), 25–43.
<https://doi.org/10.2167/la414.0>
- Farrell, P. (1990). Vocabulary in ESP: A Lexical Analysis of the English of Electronics and a Study of Semi-Technical Vocabulary. CLCS Occasional Paper No. 25.
- Faucett, L., & Maki, I. (1932). *A study of English word-values statistically determined from the latest word-counts*. Matsumura Sanshodo.
- Faucett, L., Palmer, H., Thorndike, E. L., & West, M. (1936). Interim report on vocabulary selection. *PS King*.
- Foley, A. (2009). Using academic word lists in the communicative classroom. *English Australia Journal*, 25(1), 17–26.

- Folse, K. S. (2004). Myths about teaching and learning second language vocabulary: What recent research says. *TESL Reporter*, 37(2), 1–13. <https://doi.org/10.3998/mpub.23925>
- Folse, K. S., & Youngblood, A. (2017). Survey of corpus-based vocabulary lists for TESOL classes. *MexTESOL Journal*, 41(1), 1–15. <https://doi.org/10.3998/mpub.23925>
- Frankenberg-Garcia, A. (2012). Raising teachers' awareness of corpora. *Language Teaching*, 45(4), 475–489. <https://doi.org/10.1017/S0261444810000480>
- Fries, C. C., & Traver, A. A. (1950). *English word lists*. George Wahr.
- Friginal, E. (2018). *Corpus linguistics for English teachers: Tools, online resources, and classroom activities*. Routledge. <https://doi.org/10.4324/9781315649054>
- Gao, X., & Ma, Q. (2011). Vocabulary learning and teaching beliefs of pre-service and in-service teachers in Hong Kong and mainland China. *Language Awareness*, 20(4), 327–342. <https://doi.org/10.1080/09658416.2011.579977>
- Gardner, D. (2007). Validating the construct of *word* in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, 28(2), 241–265. <https://doi.org/10.1093/applin/amm010>
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305–327. <https://doi.org/10.1093/applin/amt015>
- Gerami, M. R., & Noordin, N. B. (2013). Teacher cognition in foreign language vocabulary teaching: A study of Iranian high school EFL teachers. *Theory & Practice in Language Studies*, 3(9). <https://doi.org/10.4304/tpls.3.9.1531-1545>
- Ghadessy, M. (1979). Frequency counts, word lists, and materials preparation: A new approach. *English Teaching Forum*, 17(1), 24–27.
- Gilner, L. (2011). A Primer on the General Service List. *Reading in a Foreign Language*, 23(1), 65–83.
- Gladman, T., Tylee, G., Gallagher, S., Mair, J., Rennie, S. C., & Grainger, R. (2020). A Tool for Rating the Value of Health Education Mobile Apps to Enhance Student Learning (MARuL): Development and Usability Study. *JMIR mHealth and uHealth*, 8(7). <https://doi.org/10.2196/18015>

- González-Fernández, B., & Schmitt, N. (2017). Vocabulary acquisition. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 280–298). Taylor & Francis. <https://doi.org/10.4324/9781315676968-16>
- Goulden, R., Nation, I. S. P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11(4), 341–363. <https://doi.org/10.1093/applin/11.4.341>
- Grabowski, Ł. (2015). Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *English for Specific Purposes*, 38, 23–33. <https://doi.org/10.1016/j.esp.2014.10.004>
- Graff, J. C. (2016). Mixed methods research. In H. R. Hall & L. A. Roussel (Eds.), *Evidence-Based Practice* (pp. 47–66). Jones & Bartlett Learning.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*, 2(163–194), 105.
- Guidy-Oulai, A.-M. (2009). The development and application of a checklist for evaluating e-learning in organizations. [Doctoral dissertation, Western Michigan University]. <https://scholarworks.wmich.edu/dissertations/665>
- Ha, A. Y. H., & Hyland, K. (2017). What is technicality? A technicality analysis model for EAP vocabulary. *Journal of English for Academic Purposes*, 28, 35–49. <https://doi.org/10.1016/j.jeap.2017.06.003>
- Hafner, C. A., & Candlin, C. N. (2007). Corpus tools as an affordance to learning in professional legal education. *Journal of English for Academic Purposes*, 6(4), 303–318. <https://doi.org/10.1016/j.jeap.2007.09.005>
- Hales, B., Terblanche, M., Fowler, R., & Sibbald, W. (2008). Development of medical checklists for improved quality of patient care. *International Journal for Quality in Health Care*, 20(1), 22–30. <https://doi.org/10.1093/intqhc/mzm062>

- Hartshorn, J., & Hart, J. (2016). Comparing the Academic Word List with the Academic Vocabulary List: Analyses of frequency and performance of English language learners. *Journal of Language Teaching and Learning*, 6(2), 70–87.
- Hashimoto, B. J., & Egbert, J. (2019). More than frequency? Exploring predictors of word difficulty for second language learners. *Language Learning*, 69(4), 839–872.
<https://doi.org/10.1111/lang.12353>
- Haywood, S. (2007). *The AWL highlighter* [Online search engine].
<https://www.nottingham.ac.uk/alzsh3/acvocab/awlhighlighter.htm>
- He, X., & Godfroid, A. (2019). Choosing words to teach: A novel method for vocabulary selection and its practical application. *TESOL Quarterly*, 53(2), 348–371. <https://doi.org/10.1002/tesq.483>
- Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). *RANGE*. In [Computer software].
<https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs>
- Hernandez, M. M. (2017). *Comparing the AWL and AVL in textbooks from an intensive English program* [Master's thesis, Brigham Young University].
<https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=7458&context=etd>
- Hirsh, D., & Coxhead, A. (2009). Ten ways of focussing on science-specific vocabulary in EAP classrooms. *English Australia Journal*, 25(1), 5–16.
- Horn, E. (1926). *A Basic Writing Vocabulary, 10,000 Words Most Commonly Used in Writing*. College of Education, University of Iowa.
- Howatt, A. P. R., & Widdowson, H. G. (2004). *A history of English language teaching* (2nd ed.). Oxford University Press.
- Hubbard, P. (2006). Evaluating CALL software. *Calling on CALL: From theory and research to new directions in foreign language teaching*, 313–338.

- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.
<https://doi.org/10.1017/CBO9781139524773>
- Hyland, K., & Tse, P. (2007). Is there an 'Academic Vocabulary'? *TESOL Quarterly*, 41(2), 235–253.
<https://doi.org/10.1002/j.1545-7249.2007.tb00058.x>
- Jablonkai, R. (2010). English in the context of European integration: A corpus-driven analysis of lexical bundles in English EU documents. *English for Specific Purposes*, 29(4), 253–267.
<https://doi.org/10.1016/j.esp.2010.04.006>
- Jones, M., & Durrant, P. (2010). What can a corpus tell us about vocabulary teaching materials? In A. O'Keefee & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 387–400). Routledge. <https://doi.org/10.4324/9780203856949-28>
- Karpenko-Seccombe, T. (2021). *Academic writing with corpora: A resource book for data-driven learning*. Routledge.
- Kashoob, M. (2018). *Evaluating ELT materials with specific reference to colleges of applied sciences general foundation programme in Oman: Towards a viable checklist* [Doctoral dissertation, University of York]. <https://etheses.whiterose.ac.uk/20915/>
- Kennedy, G. (1998). *An introduction to corpus linguistics*. Longman.
- Khani, R., & Tazik, K. (2013). Towards the development of an academic word list for applied linguistics research articles. *RELC journal*, 44(2), 209–232.
<https://doi.org/10.1177/0033688213488432>
- Kilgarriff, A., Charalabopoulou, F., Gavrilidou, M., Johannessen, J. B., Khalil, S., Kokkinakis, S. J., Lew, R., Sharoff, S., Vadlapudi, R., & Volodina, E. (2014). Corpus-based vocabulary lists for language learners for nine languages. *Language Resources and Evaluation*, 48(1), 121–163.
<https://doi.org/10.1007/s10579-013-9251-2>
- Kremmel, B. (2016). Word families and frequency bands in vocabulary tests: Challenging conventions. *TESOL Quarterly*, 50(4), 976–987. <https://doi.org/10.1002/tesq.329>

- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Brown University Press.
- Latif, M. M. A. (2020). Corpus literacy instruction in language teacher education: Investigating Arab EFL student teachers' immediate beliefs and long-term practices. *ReCALL*, 1–15.
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a Foreign Language*, 22(1), 15–30.
- Le Foll, E. (2021). *Creating Corpus-Informed Materials for the English as a Foreign Language Classroom: A step-by-step guide for (trainee) teachers using online resources* (1st ed.). Open Educational Resource. <https://elenlefol.pressbooks.com>
- Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, 22, 42–53.
<https://doi.org/10.1016/j.jeap.2016.01.008>
- Lessard-Clouston, M. (2013). Word Lists for Vocabulary Learning and Teaching. *CATESOL Journal*, 24(1), 287–304.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newberry Park. Sage Publications.
[https://doi.org/10.1016/0147-1767\(85\)90062-8](https://doi.org/10.1016/0147-1767(85)90062-8)
- Llorens-Vernet, P., & Miró, J. (2020). The Mobile App Development and Assessment Guide (MAG): Delphi-Based Validity Study. *JMIR mHealth and uHealth*, 8(7). <https://doi.org/10.2196/17760>
- Lobe, B., Morgan, D., & Hoffman, K. A. (2020). Qualitative data collection in an era of social distancing. *International Journal of Qualitative Methods*, 19, 1609406920937875.
<https://doi.org/10.1177/1609406920937875>
- Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, 22(3), 319–344. <https://doi.org/10.1075/ijcl.22.3.02lov>

- Lynn, R. W. (1973). Preparing word-lists: a suggested method. *RELC Journal*, 4(1), 25–28.
<https://doi.org/10.1177/003368827300400103>
- Martínez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes*, 28(3), 183–198.
<https://doi.org/10.1016/j.esp.2009.04.003>
- Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33(3), 299–320.
<https://doi.org/10.1093/applin/ams010>
- McArthur, T. (1998). *Living words: language, lexicography, and the knowledge revolution*. University of Exeter Press.
- McLean, S. (2018). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics*, 39(6), 823–845. <https://doi.org/10.1093/applin/amw050>
- Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching*, 13(4), 221–246. <https://doi.org/10.1017/S0261444800008879>
- Miller, D. P. (2012). *The challenge of constructing a reliable word list: An exploratory corpus-based analysis of introductory psychology textbooks* [Doctoral dissertation, Northern Arizona University].
- Miller, D. P., & Biber, D. (2015). Evaluating reliability in quantitative vocabulary studies: The influence of corpus design and composition. *International Journal of Corpus Linguistics*, 20(1), 30–53. <https://doi.org/10.1075/ijcl.20.1.02mil>
- Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System*, 28(2), 291–304. <https://doi.org/10.1002/RRQ.011>
- Mukherjee, J. (2004). Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In *Applied Corpus Linguistics* (pp. 239–250). Brill.
https://doi.org/10.1163/9789004333772_014
- Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, 47(1), 91–108. <https://doi.org/10.1002/RRQ.011>

- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Heinle and Heinle.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139524759>
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59–82. <https://doi.org/10.3138/cmlr.63.1.59>
- Nation, I. S. P. (2007). The four strands. *International Journal of Innovation in Language Learning and Teaching*, 1(1), 2–13. <https://doi.org/10.2167/illt039.0>
- Nation, I. S. P. (2012). The BNC/COCA word family lists. *Document bundled with Range program with BNC/COCA lists*, 25.
- Nation, I. S. P. (2013). *Learning Vocabulary in Another Language* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139858656>
- Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins <https://doi.org/10.1075/z.208>
- Nation, I. S. P., & Sorell, J. (2016). Corpus selection and design. In I. S. P. Nation (Ed.), *Making and Using Word Lists for Language Learning and Testing* (pp. 95–105). John Benjamins <https://doi.org/10.1075/z.208.10ch10>
- Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6–19). Cambridge University Press.
- Nation, I. S. P., & Webb, S. A. (2011). *Researching and analyzing vocabulary*. Heinle, Cengage Learning.
- Nation, I. S. P., & Yamamoto, A. (2012). Applying the four strands. *Nova Science Publishers*.
- Nation, I. S. P., Coxhead, A., Chung, T. M., & Quero, B. (2016). Specilaized word lists. In I. S. P. Nation (Ed.), *Making and Using Word Lists for Language Learning and Testing* (pp. 145–151). John Benjamins. <https://doi.org/10.1075/z.208.14ch14>

- Nation, I. S. P., Shin, D., & Grant, L. (2016). Multiword units. In I. S. P. Nation (Ed.), *Making and Using Word Lists for Language Learning and Testing* (pp. 81–84). John Benjamins
<https://doi.org/10.1075/z.208.06ch6>
- Nesi, H., Vincent, B., Smith, S., & Alsop, S. (2021, April 6–10). *Core corpus skills for academic purposes: A TEAP proposal* [Conference paper]. First Online Biennial BALEAP Conference
 BALEAP, Glasgow, University of Glasgow
<https://www.eventsforce.net/gu/frontend/reg/absViewDocumentFE.csp?documentID=4574&eventID=123>
- Newman, J. A. (2016). *A corpus-based comparison of the Academic Word List and the Academic Vocabulary List* [Master's thesis, Brigham Young University].
<https://scholarsarchive.byu.edu/etd/6080>
- O'Dell, F. (1997). Incorporating vocabulary into the syllabus. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 258–278). Cambridge University Press.
- Ogden, C. K. (1930). *Basic English: A general introduction with rules and grammar*.
- O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511497650>
- Palinkašević, R. (2017). Specialized word lists: Survey of the literature: Research perspective. *Istraživanja u pedagogiji*, 7(2), 221–238. <https://doi.org/10.17810/2015.61>
- Palmer, H. (1931). Second interim report on vocabulary selection submitted to the Eighth Annual Conference of English Teachers under the auspices of the Institute for Research in English Teaching. *IRET*.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and Communication*, 191, 225.

- Pinchbeck, G. (2014, March). Lexical frequency profiling of a large sample of Canadian high school diploma exam expository writing: L1 and L2 academic English [Roundtable presentation]. *American Association of Applied Linguistics, Portland, OR, USA.*
- Poole, R. (2018). *A guide to using corpora for English language learners*. Edinburgh University Press.
- Praninskas, J. (1972). *American university word list*. Longman.
- Qi, H. (2016). *A corpus-based comparison between the Academic Word List and the Academic Vocabulary List* [Master's thesis, The University of Western Ontario].
<http://ir.lib.uwo.ca/etd/3721>
- Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System*, 42, 220–231.
<https://doi.org/10.1016/j.system.2013.12.003>
- QualtricsXM. (2018). (Version November, 2018) [Software program]. <https://www.qualtrics.com>
- Richards, J. C. (1970). A psycholinguistic measure of vocabulary selection. *IRAL-International Review of Applied Linguistics in Language Teaching*, 8(2), 87–102.
<https://doi.org/10.1515/iral.1970.8.2.87>
- Richards, J. C. (1974). Word lists: Problems and prospects. *RELC Journal*, 5(2), 69–84.
<https://doi.org/10.1177/003368827400500207>
- Robbins, N. B., & Heiberger, R. M. (2011). Plotting Likert and other rating scales. *Proceedings of the 2011 Joint Statistical Meeting*, 1,
- Römer, U. (2006). Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future developments. *Zeitschrift für Anglistik und Amerikanistik*, 54(2), 121–134.
<https://doi.org/10.1515/zaa-2006-0204>
- Römer, U. (2009). Corpus research and practice: What help do teachers need and what can we offer? In K. Aijmer (Ed.), *Corpora and Language Teaching* (pp. 83–98). John Benjamins.
<https://doi.org/10.1075/scl.33.09rom>

- Rossiter, M. J., Abbott, M. L., & Kushnir, A. (2016). L2 vocabulary research and instructional practices: where are the gaps? *TESL-EJ*, 20(1).
- Schmitt, N. (2000). Lexical chunks. *ELT Journal*, 54(4), 400–401. <https://doi.org/10.1093/elt/54.4.400>
- Schmitt, D. (2016, September 12–14). *Beyond caveat emptor: Applying validity criteria to word lists* [Conference paper]. Vocab@Tokyo Conference Tokyo: Meiji Gakuin University http://vli-journal.org/vocabattokyo/vocabattokyo_handbook_2016.pdf
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan. <https://doi.org/10.1057/9780230293977>
- Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition*, 19(1), 17–36. <https://doi.org/10.1017/S0272263197001022>
- Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484–503. <https://doi.org/10.1017/S0261444812000018>
- Schmitt, N., & Schmitt, D. (2020). *Vocabulary in language teaching* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/9781108569057>
- Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly*, 36(2), 145–171. <https://doi.org/10.2307/3588328>
- Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2017). How much vocabulary is needed to use English? Replication of Van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching*, 50(2), 212–226. <https://doi.org/10.1017/S0261444815000075>
- Schmitt, N., Dunn, K., O'Sullivan, B., Anthony, L., & Kremmel, B. (2021). Introducing knowledge-based vocabulary lists (KVL). *TESOL Journal*, 12(4). <https://doi.org/10.1002/tesj.622>
- Schools, C. f. (2021). *Teaching English Language with Corpus Linguistics*. <http://wp.lancs.ac.uk/corpusforschools/>

- Schroeter, D. C. (2008). *Sustainability evaluation: Development and validation of an evaluation checklist* [Doctoral dissertation, Western Michigan University].
<https://scholarworks.wmich.edu/dissertations/814>
- Scriven, M. (2007). *The logic and methodology of checklists*. The Evaluation Center.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487–512. <https://doi.org/10.1093/applin/amp058>
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Sinclair, J. M. (1987). *Looking up: An account of the COBUILD project in lexical computing and the development of the Collins COBUILD English language dictionary*. Collins ELT.
<https://www.jstor.org/stable/25469963?seq=1>
- Sorell, C. J. (2013). *A study of issues and techniques for creating core vocabulary lists for English as an international language* [Doctoral dissertation, Victoria University of Wellington].
<http://hdl.handle.net/10063/3016>
- Stoeckel, T., Ishii, T., & Bennett, P. (2018). Is the Lemma More Appropriate than the Flemma as a Word Counting Unit? *Applied Linguistics*. <https://doi.org/10.1093/applin/amy059>
- Stuart, M., Dixon, M., Masterson, J., & Gray, B. (2003). Children's early reading vocabulary: Description and word frequency lists. *British Journal of Educational Psychology*, 73(4), 585–598. <https://doi.org/10.1348/000709903322591253>
- Stufflebeam, D. L. (2000). Guidelines for developing evaluation checklists: The checklists development checklist (CDC).
https://wmich.edu/sites/default/files/attachments/u350/2014/guidelines_cdc.pdf
- Sue, V. M., & Ritter, L. A. (2012). *Conducting online surveys* (2nd ed.). Sage Publication.
<https://doi.org/10.4135/9781506335186>
- Szudarski, P. (2018). *Corpus Linguistics for Vocabulary: A Guide for Research*. Routledge.
<https://doi.org/10.4324/9781315107769>

- Tanaka-Ishii, K., & Terada, H. (2011). Word familiarity and frequency. *Studia Linguistica*, 65(1), 96–116. <https://doi.org/10.1111/j.1467-9582.2010.01176.x>
- Tashakkori, A., & Creswell, J. W. (2007). *Exploring the nature of research questions in mixed methods research*. Sage Publications. <https://doi.org/10.1177/1558689807302814>
- Text Inspector (2021). VocabProfile Home [Online lexis analysis tool]. <https://textinspector.com/>
- Thompson, P., & Alzeer, S. N. (2019). A Survey of issues, practices and views related to corpus-based word lists for English language teaching and learning. *International Journal of Applied Linguistics and English Literature*, 8(6), 43–53. <https://doi.org/10.7575/aiac.ijalel.v.8n.6p.43>
- Thorndike, E. L. (1921). *The teacher's word book*. Teachers College, Columbia University.
- Thorndike, E. L. (1932). *A teacher's word book of 20,000 words*. Teachers College, Columbia University.
- Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. Teachers College, Columbia University.
- Tognini-Bonelli, E. T. (2010). Theoretical overview of the evolution of corpus linguistics. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 14–28). Routledge. <https://doi.org/10.4324/9780203856949>
- Tomlinson, B. (2003). Materials evaluation. In B. Tomlinson (Ed.), *Developing materials for language teaching* (pp. 15–36). Continuum.
- Treffers-Daller, J., & Milton, J. (2013). Vocabulary size revisited: The link between vocabulary size and academic achievement. *Applied Linguistics Review*, 4(1), 151–172. <https://doi.org/10.1515/applirev-2013-0007>
- Treffers-Daller, J., & Tidball, F. (2008). Analysing lexical richness in French learner language: What frequency lists and teacher judgements can tell us about basic and advanced words. *Journal of French language studies*, 18(3). <https://doi.org/10.1017/S0959269508003463>

- Valipouri, L., & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes*, 12(4), 248–263.
<https://doi.org/10.1016/j.jeap.2013.07.001>
- van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34(4), 457–479.
<https://doi.org/10.1093/applin/ams074>
- Vilkaitė-Lozdienė, L., & Schmitt, N. (2020). Frequency as a Guide for Vocabulary Usefulness. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 81–96). Routledge.
<https://doi.org/10.4324/9780429291586-6>
- Vongpumivitch, V., Huang, J., & 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), 33–41. <https://doi.org/10.1016/j.esp.2008.08.003>
- Wadden, P., Ferreira, D., & Rush, E. (2016). The global academic vocabulary lexicon: A new ELT resource. *Accents Asia* (8), 135–142.
- Walker-Egea, C. F. (2014). *Design and validation of an evaluation checklist for organizational readiness for evaluation capacity development* [Doctoral dissertation, University of South Florida]. <https://scholarcommons.usf.edu/etd/5399>
- Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*, 28(3), 170–182.
<https://doi.org/10.1016/j.esp.2009.04.001>
- Ward, J., & Chuenjundaeng, J. (2009). Suffix knowledge: Acquisition and applications. *System*, 37(3), 461–469.
- Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.
- Webb, S., Yanagisawa, A. & Uchihara, T. (2020). How effective are intentional vocabulary-learning activities? A meta-analysis. *The Modern Language Journal*, 104(4), 715–738. <https://doi.org/10.1111/modl.12671>

Weisser, M. (2016). *Practical Corpus Linguistics: An Introduction to Corpus-based Language Analysis*.

John Wiley & Sons Ltd. <https://doi.org/10.1002/9781119180180>

West, M. (1927). *The new method readers (new series)*. Longmans, Green & Co.

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

White, R. V. (1988). *The ELT curriculum: Design, innovation and management*. Blackwell Publishers.

Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00.

Behavior research methods, instruments, & computers, 20(1), 6–10.

<https://doi.org/10.3758/BF03202594>

Wray, A. (2009). Identifying formulaic language: Persistent challenges and new opportunities.

Formulaic Language, 1, 27–51. <https://doi.org/10.1075/tsl.82.02ide>

Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication*, 3(2), 215–229.

Yang, H. (1986). A new technique for identifying scientific/technical terms and describing science texts. *Literary and Linguistic Computing*, 1(2), 93–103. <https://doi.org/10.1093/lc/1.2.93>

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

<https://doi.org/10.1016/j.esp.2014.05.003>

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2010). *The program evaluation standards: A guide for evaluators and evaluation users*. Sage Publications.

Zahan, I., & Begum, J. (2013). How to evaluate an EFL/ESL textbook: A problem and a solution. *ASA University Review*, 7(1).

Zimmerman, C. B. (1997). Historical trends in second language vocabulary instruction. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition* (Vol. 6, pp. 5–19). Cambridge University Press. <https://doi.org/10.1017/CBO9781139524643.003>

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Hafner.

APPENDICES

APPENDIX A

Corpus-based Word Lists for English Language Teaching (2000-2020)

General English Word Lists

Name	Author(s), Year	Vocabulary type	Size of word list, word unit	Corpus details	Criteria used	Target audience and purpose	Notes
General Service List (GSL)	West, 1953	General	2,000 word families	5 million word written corpus (sources date from 1920's)	Frequency and subjective criteria (Efficiency criteria)	Vocabulary to be learned by L2 learners in the first year with a focus on reading	Contains archaic forms of some words like shilling and lacks many modern words like plastic, okay, computer, etc. Many revised versions
The BNC 2000	Nation, 2006	General	1,000 word families divided into 14 lists	100 million word BNC corpus	Frequency, range and dispersion in addition to subjective adjustments	High-frequency receptive vocabulary needed for successful comprehension by L2 learners	
Word Frequency List of American English	Davies & Gardner, 2010	General	20,000 lemmas	Corpus of Contemporary American English, 400+ million words			
The BNC/COCA 2000	Nation, 2012	General	1,000 word families	BNC 10 million corpus (40% of written and 60% of spoken)	Range, frequency, dispersion and subjective adjustments	Course design for EFL at secondary school level and receptive vocabulary size testing	Studies (Dang & Web, 2017; Dang, Webb, & Coxhead, 2020) compared the BNC/COCA2000 with the GSL, New GSL and BNC2000 using information from corpora, learners and teachers and found that the BNC/COCA 2000 is likely to be a more suitable list for EFL learner
New General Service List 1 (NGSL 1)	Brezina & Gablasova, 2015	General	2,494 lemmas	12.1 billion words from 4 corpora (LOB, BNC, BE06, and EnTenTen12)	Frequency, dispersion and distribution across corpora	An attempt to update and improve the GSL employing quantitative approach with	Includes 378 "current vocabulary" words

						a focus on the question of core vocabulary existence and stability	
New General Service List 2 (NGSL2)	Browne, Culligan & Phillips, 2013	General	2800 flemma	273 million word subsection of the 2 billion word Cambridge English Corpus	A combination of objective and subjective measures following West	Frequency list for receptive and productive use by L2 beginners; an attempt to update the GSL with a list of fewer items.	Includes 2368 word families
COCA frequency lists	Davies, 2015	General	5000, 20,000, 60,000 and 100,000 lemma	Based on the 520 million word Corpus of Contemporary American English			Based on the largest publically available corpus. Many other resources available.
CEFR- J Word list	Tono, 2016	General (CEFR-based)	7,815 lemmas	Based on the major English textbooks used at primary to secondary schools (Years 3 to 10) in China, Korea, and Taiwan	Frequency	CEFR for Japan	All the words have part-of speech information and corresponding CEFR levels.
Essential Word List	Dang & Webb, 2016	General	800 flemmas	9 spoken and 9 written corpora containing 10 varieties of English	A combination of strongest items from the four lists (GSL, BNC2000, BNC/COCA2000 and New-GSL1)	A starting point for L2 beginners' lexical development	624 lexical words, 176 function words
New General Service List - Spoken (NGSL-S)	Browne & Culligan, 2017	General, spoken	721 flemma	Spoken section of the Cambridge English Corpus			Part of the NGSL list

General Academic Word Lists

Name	Author(s), Year	Vocabulary type	Size of word list, word unit	Corpus details	Criteria used	Target audience and purpose	Notes
Academic Word List (AWL)	Coxhead, 2000	Academic	570 word families	3,500,000 academic word corpus	Specialized occurrence (outside of the GSL), range and frequency	Academic vocabulary for university students in EAP programs, reading	An attempt to replace the UWL. The first corpus-based and most influential general academic word list; widely used in vocabulary learning and research
Vocabulary for Academic Lecture Listening (VALL)	Thompson, 2006	Academic lectures	200 word families	The BASE corpus; 1,644,942 tokens from 160 lectures and 40 seminars	outside the 2000 from the BNC, frequency and range	learners who are preparing for listening to lectures	Focus on Economics lectures
Academic Keyword List	Paquot, 2010	Academic	930 lemmas	3 million word, professional and novice academic corpus		Aimed more at writing than reading	Does not exclude 2000 most frequent words
New Academic Word List (NAWL)	Browne, Culligan & Phillips, 2013	Academic	963 modified lexemes	288 million word academic corpus (mostly from the Cambridge English Corpus - Academic)	Frequency, dispersion and appropriateness in addition to exclusion of New-GSL2 items	An attempt to update Coxhead's AWL which was based on West's GSL to fit with the NGSL2, developed by the same researchers	Excludes the NGSL; available in headwords and lemma with English and Japanese meanings
Lecture Introduction Word list	Yaqoob, 2013	Academic-lecture introductions	200 word types	45,305 tokens from 89 lectures			Includes some formulaic language

Academic Vocabulary List (AVL)	Davies & Gardner, 2014	Academic,	3015 lemmas	120 million COCA Academic subset	Frequency, range, dispersion and discipline-specificity measure	An attempt to update the AVL and address some of its methodological problems; target at L2 learners in EAP programs	The frequency of words must be at least 50% higher in the academic corpus than in a general corpus. Word family version also available.
TOEIC Service List 1.1	Browne & Culligan, 2016	Academic - TOEIC exam	1200 modified lexemes	1,5 million word corpus			Excludes the NGSL
Academic Spoken Word List	Dang, Coxhead & Webb, 2017	Academic, spoken	1,741 word families	13 million word academic spoken corpus	Frequency, range, and dispersion	L2 learners in EAP/EGAP programs regardless of discipline and proficiency level; academic speech (lectures)	Vocabulary from 24 subjects across 4 disciplinary sub corpora, has four levels which are adaptable to learners' current knowledge of general vocabulary

Specialised Academic Word Lists

Name	Author(s), Year	Vocabulary type	Size of word list, word unit	Corpus details	Criteria used	Target audience and purpose	Notes
Medical Research Article Word Lists	Chen & Ge, 2007	Academic, Medical	292 word families	90,425 tokens from 50 medical Research articles covering 25 medical subject categories			Designed as a preliminary study into a medical academic word list. Word families are included in the AWL.
Academic Vocabulary in Business News	Boonyapapong, 2007	Academic, Business news	100 word families	859,890 tokens from a Thai English newspaper	Frequency	Identify the 100 most frequent academic vocabulary in business news	List only includes AWL families
Business Word List (BWL)	Konstantakis, 2007	Academic, business	480 word families	600,000 token corpus from 33 business English textbooks (designed for L2 learners)		L2 undergraduate Business students	Excludes the GSL and AWL; small corpus of simplified texts
Pilot Science Word list for EAP	Coxhead & Hirsh, 2007	Academic, Science	318 word families	1,761,380 tokens across 14 science subject areas from course books for first year Science students at a university, New Zealand	Range, frequency and dispersion	A pilot list to discover science vocabulary outside the GSL and AWL	
Medical Academic Word List (MAWL)	Wang, Liang & Ge, 2008	Academic, Medical	623 word families	Medical research articles (RAs)	Specialised occurrence (outside of the GSL), Range and Frequency		Excludes items from the GSL, contains 342 words from the AWL
Agricultural Word List (Agrocorpus List)	Martinez, Beck & Panza, 2009	Academic, Agriculture	92 word families	826, 416 token AgroCorpus based on 218 articles from journals from 2000-2003	Frequency, coverage and distribution of AWL in the AgroCorpus plus subjective criteria	Reading and writing in agriculture	Reduced list including only AWL families
Business Word List (2)	Konstantakis, 2010	Academic, business	1,613 word families	1 million tokens			

Business Word List for postgraduates	Hsu, 2011	Academic, business	426 word families	7.62 million tokens from 2,200 business research articles across 20 business areas		Graduate business majors who must read business textbooks designed for native speakers.	Excludes the BNC 3000; includes 151 word families from the AWL
Applied Linguistics AWL	Khani & Tazik, 2013	Academic, Applied Linguistics	773 word types	1,553,450 tokens from 240 articles from 12 applied linguistics journals (20 from each).	Frequency and range	Academic vocabulary in the field of applied linguistics	Excludes GSL; 573 of the 773 word types are in the AWL
Medical Academic Word List for clinical case histories (MAWLcc)	Mungra & Canzianni, 2013	Academic	241 word families	246,907 token corpus from 200 case studies from 72 medical journals	Specialised occurrence, range and frequency, plus subjective criteria to exclude items familiar to the target physicians	Vocabulary development of L2 medical students	Excludes the first 2000 words from GSL. Range of at least 50% coverage of 24 areas. Occurrence of at least 30 times in corpus.
Chemistry Academic Word List (CAWL)	Valipouri & Nassaji, 2013	Academic, Chemistry	1400 word families	4 million word chemistry research articles corpus from four sub-subjects	Corpus comparison based on frequency	Help EFL chemistry students with low proficiency levels understand research articles	Also identifies 390 non-GSL/AWL word families
Environmental Academic Word List (EAWL)	Liu & Han, 2015	Academic, Environmental	458 word families	862,242 words from 200 texts from 10 subject areas	Exclusion of the first GSL 2000 families, range, frequency and optimised usefulness	Establishing an Environmental Academic Word List due to lack of coverage of the AWL in the specialised corpus	Shares 318 words with the AWL
Nursing Academic Word List	Yang, 2015	Academic, technical-nursing	676 word families	1 million words from 252 nursing articles; The Nursing Research Articles Corpus		Help Nursing graduate students read and write academic papers	Excludes the GSL (first 2000 words); 378 word families overlap with the AWL
Academic Business English List (ABEL)	Stella, 2015	Academic, Business	840 word families	15 textbooks used in core courses in an undergraduate business program	Exclude words from BNC/COCA 3000 then frequency and range	The technical vocabulary needed by undergraduate students in business in order to understand their textbooks	

Food Science and Technology Academic Word List (FSTAWL)	Esfandiari & Moein, 2015	Academic, technical-food science and technology	1090 word families	4,652,444 running words from 1421 research articles from 38 journals across 5 sub-disciplines			prepositions, pronouns, determiners, conjunctions, auxiliaries, particles, proper names, and acronyms removed
New Medical Academic Word List	Lei & Liu, 2016	Academic, technical-medical	819 lemmas	2.7 million words from medical journal articles (760 articles from 38 journals) and 3.5 million words from medical English textbooks	Combination of AWL and AVL measures (minimum frequency, ratio frequency, range, dispersion, discipline measure and special meaning criteria for general high-frequency) corpus comparison	Help medical students, as well as non-native professionals and researchers, read research articles	146 General English lemmas with no medical meaning removed
Linguistics Academic Word List (LAWL)	Moini & Islamizadeh, 2016	Academic, Linguistics	1263 word families	4 million words from 700 linguistics research articles covering four main linguistics sub disciplines			Includes 224 words not in the GSL and AWL
Business Service List 1.01 (BSL)	Browne & Culligan, 2016	Business	1700 modified lexemes	64 million word corpus	Frequency and exclusion of items in NGSL		
Science Textbook Word List (STWL)	Veenstra & Sato, 2018	Academic, Science	309 word families	700,000 word academic corpus from 12 textbooks on biology, chemistry, physics and engineering	Specialised occurrence (outside of the GSL), range and frequency	Help Japanese student in their reading comprehension of English science textbooks	GSL excluded; 127 word families overlap with the AWL

Name	Author(s), Year	Vocabulary type	Size of word list, word unit	Corpus details	Criteria used	Target audience and purpose	Notes
Technical Vocabulary in discipline-related movies and TV shows	Csomay & Petrovic, 2012	Technical, legal	1,124 word types	130,000 words from legal subject based movies and TV shows	Distribution the RANGE program then checking specialised dictionary	Legal content	To investigate the extent to which watching discipline-related movies and TV shows in L2 may facilitate incidental learning of technical vocabulary
Basic Engineering List (BEL)	Ward, 2009	Technical, engineering	299 word types	Corpus from 3rd and 4th year undergraduate textbooks		Foundation engineering students with little lexical or grammatical knowledge; applies to all engineering disciplines	
Computer Science Word List	Minshall, 2013	Technical, Computer Science	433 word families	3,661,337 tokens compiled from journal articles and conference proceedings			Covers 10 sub disciplines of computer Science; Excludes the GSL and the AWL
Theological Word List	Lessard-Clouston, 2010	Technical, Theology	100 items	23 (90 minutes) academic theology lectures	Inclusion of words from handouts used during class and those written on the board. The final list is divided into two groups based on frequency		Excludes GSL and AWL
Essential Pharmacology Word List (EPWL)	Fraser, 2012	Technical, pharmacology	570 word families	369,000 words from 100 pharmacology articles			411 unproblematic and function words removed

Integrated Pharmacology Word List (IPWL)	Fraser, 2009	Technical, pharmacology	2000 word families	369,000 words from 100 pharmacology articles			
Pharmacology Word List (PWL)	Fraser, 2007	Technical, pharmacology	601 word families	180,000 words from 50 research articles			Excludes GSL and AWL
OPEC Word List	Aluthman, 2017	Technical, oil marketing	255 word types	1,004,542 words from 40 OPEC monthly reports released between 2003 - 2015	Frequency and range	Lexicographic research and L2 vocabulary learning in oil marketing and oil industries	Excludes AWL and GSL
Newspaper Word List (NWL)	Chung, 2009	Technical, newspapers	588 word families	579,849 words from 12 sections in three Newspapers published online Feb.- March 2006	Frequency	Reading newspaper articles	Excludes the GSL
Insurance Research Articles Word List	Khamphairoh & Tangpijaikul, 2012	Technical, insurance	100 types and collocations	980,121 tokens from 155 research articles from two insurance journals from between 2007 - 2010			Only keywords with an insurance specific meaning were selected. Two and three word collocations for the first 10 keywords also provided.
Education and Training Program Word List (ETPWL)	Freund, 2014	Technical, grant guidelines	604 word types (not including plural forms)	252,599 running words from 5 sub grant calls in 2013		Vocabulary for those who are applying for grants	Excludes GSL

Technical Words in Finance Word List	Tangpoorn-Patanasorn, 2018	Technical, finance	979 lemmas	2,004,964 running words from 4 finance-related books, journals, websites and newspapers			Includes 569 word families. Includes 413 words from GSL and 291 words from AWL
Student Engineering Word List (SEWL)	Mudraya, 2006	Technical, engineering	1200 word families	2 million tokens from 13 complete textbooks from undergraduate engineering courses			
Engineering English Word List (EEWL)	Hsu, 2014	Technical, engineering	729 word families	4.57 million words from 100 college textbooks across 20 engineering subject areas	Frequency and exclusion of the BNC/ COCA 2000		
Opaque Engineering Word List	Todd, 2017	Technical, engineering	186 word types	1.15 million tokens from 27 engineering course books	Frequency and dispersion and meanings should be included on the final list is opacity		Focus on opaque vocabulary (words that cannot easily be understood).
Anatomy Word List	Fraser, Davies & Tatsukawa, 2015	Technical, anatomy	500 word types, 353 two word terms, 100 three word terms	361,097 tokens from Grey's Anatomy for students textbook			Also has 100 top multiword units

Engineering Technology Word List (ETWL)	Jin et al., 2012	Technical, engineering	313 word types	124, 581 word engineering corpus (Malaysian engineering technology textbooks)			Excludes GSL and AWL
First-year Engineering Word List (FEWL)	Murphy, 2015	Technical, engineering	570 word families	All textbooks used in first year engineering courses for 2014-15 academic year			Excludes the GSL; 295 word families overlap with the AWL
Law Word List (LWL)	Aichah, 2012	Technical, Law	373 word families	3,843,107 tokens			Unpublished MA thesis (Swansea University); Excludes GSL and AWL. Also includes technical multiword list for law
Medical Word List	Hsu, 2013	Technical, medical	595 word families	155 medical textbooks across 31 subject areas; approximately 15 million tokens	range, frequency,	Aimed at medical novices, the MWL includes both sub-technical and lay-technical vocabulary, to bridge the gap before technical terms	Excludes the most frequent 3000 word families from the BNC
Music Word List	Wang & Picard. 2016	Specialised, music	1725 word families	1,601, 876 tokens from 5 music course books			1,314 word families included in the NGSL and 167 word families included in the NAWL

EU Word List (EUWL)	Jablonkai, 2017		513 word families	About 1 million running words from official EU texts from 40 different genres	Specialised occurrence (outside of the GSL), range and Frequency	Basis for course and materials design for English for the EU language programmes	
Hard Science Spoken Word List	Dang, 2018a	Specialised vocabulary in hard science (spoken)	1,595 word families	6.5 million running words from lectures, seminars, labs and tutorials in hard sciences	Frequency, range, and dispersion	L2 learners in ESAP programs	Vocabulary from 12 subjects, has four levels which are adaptable to learners' current knowledge of general vocabulary
Soft Science Spoken Word list	Dang, 2018b	Specialized vocabulary in soft science (spoken)	1,964 word families	6.5 million running words from lectures, seminars, labs and tutorials in soft sciences	Frequency, range, and dispersion	L2 learners in ESAP programs	vocabulary from 12 subjects, has four levels which are adaptable to learners' current knowledge of general vocabulary
Medical Spoken Word List	Dang, 2020	Specialized vocabulary in medicine (spoken)	895 word types	556,074-word corpus of medical lectures and seminars	Objective criteria (range, frequency, keyness) and subjective criteria (dictionary checking, expert judgment)	L2 learners in ESP programs	

APPENDIX B

Survey of the Uses of Word Lists for Language Teaching and Learning (Phase 1)

Participant Information

My name is Sarah Alzeer and I am a PhD student at The University of Birmingham, United Kingdom. I am currently surveying practitioners' and learners' views and uses of word lists for language teaching and learning related purposes. The survey is intended for individuals aged 18 years or more, who have been teaching and/or learning English as a second/foreign language, or have been engaged in any practice or research related to English language teaching and learning. Word lists are understood here to mean lists of words that are ranked by frequency and/or other measures, and grouped as single words and may include multi-word units. The results of the survey will be used to develop a framework for evaluating the usefulness of word lists for different pedagogical purposes. Please answer this online questionnaire which should take approximately 10 minutes to complete.

Please note that your email address will be collected with your original response which will be used for identifying you if you agree to be contacted for participation in further stages of the study and/or request a summary of the findings, or if you decide to withdraw your data from the study. Your email address will not be published as part of the research, and will not be used for any purpose other than communication regarding this project if you consent. Email addresses will be removed from the database and email program at the end of my PhD which are both securely stored on The University of Birmingham's server.

The information you would provide will be used solely for the purpose of this research project, and data will be kept confidential and will be anonymised in any publications. No persons other than my supervisors and I will have access to the information you would provide. Upon completion of the research, all questionnaires will be securely stored for ten years according to The University of Birmingham policy.

Participation is entirely voluntary and you are free to withdraw from the survey at any time you feel uncomfortable or unwilling to participate, and you do not have to specify a reason. Any in-part or total contribution can be withdrawn before 1/2/2019 Midnight GMT. After 1/2/2019, it will not be possible to withdraw your contribution from the results of the research. If you wish to withdraw please contact me (contact details below).

Please note that the survey will be open until 18/1/2019 Midnight GMT. If you have any questions regarding this project, please feel free to contact me.

Thank you for your time and assistance.

Please move to the next page for the consent form.

Student Contact Details

Sarah Alzeer

[Redacted]

Supervisors Contact Details

Dr. Paul Thompson

Department of English Language and Linguistics

University of Birmingham

[Redacted]

Dr. Amanda Patten

Department of English Language and Linguistics

University of Birmingham

[Redacted]

Consent form

I confirm that the purpose of the study has been explained and that I have understood it.
 I have had the opportunity to ask questions (by email), if necessary.
 I understand that data will be collected in an identifiable form, but that it will be treated as confidential and anonymised in any research outputs.
 I confirm that I have read and understood the above information and that I agree to participate in this study.

Please press "next" if you agree to this information.

Next

End

Skip To: End of Survey If - I confirm that the purpose of the study has been explained and that I have understood it. - I... = End

Email Address

Please enter your email address.

End of Block: Participant Information

Start of Block: General Background

Display This Question:

If - I confirm that the purpose of the study has been explained and that I have understood it. - I... = Next

Q1 Which of the following describes your current status? (more than one option can be selected)

- Language Teacher
- Language Learner
- Researcher
- Language Tester
- Materials developer
- Course/program coordinator
- Other (please specify) _____

Q2 In which context of English language teaching/learning have you been mainly involved in? (more than one option can be selected)

- General English at the school level
- General English for adults
- Academic English at the university level
- Specialized English at the university level
- Other (please specify) _____
- Further comments _____

End of Block: General Background

Start of Block: Directing questionnaire

Q3 Have you ever used a word list(s) for second language teaching/learning purposes?

- Yes
- No

Display This Question:

If Have you ever used a word list(s) for second language teaching/learning purposes? = No

Q3a Why have not you used a word list for second language teaching/learning?

- I have not thought of using word lists.
- I do not believe word lists are useful for language teaching and learning.
- I could not use word lists.
- I do not know what a word list is.
- Other (please explain) _____
- Further comments _____

End of Block: Directing questionnaire

Start of Block: Using Word Lists

Q4 Please choose a definition below that best describes what a word list is (to you).

1. A list of words or phrases ranked according to their frequency.
2. A list of the most important words in a particular area.
3. A group of words categorized according to certain criteria and for specific purposes.
4. Other (please provide your own definition) _____

Display This Question:

If Have you ever used a word list(s) for second language teaching/learning purposes? = Yes

Q5 Please select the purpose for which you used a word list(s). Answer the following questions accordingly. If you used a word list(s) for more than one purpose, please select the most recent one. If you wish, you can reflect on your other experience in the comments box.

- Second language teaching
- Second language learning
- Second language testing
- Materials development (including computer programs and graded readers)
- Syllabus designing or course coordination
- Research purposes (please explain) _____
- Comments _____

Display This Question:

If Have you ever used a word list(s) for second language teaching/learning purposes? = Yes

Q6 Please indicate which type of word list(s) you used.

1. Self-made word list
2. Ready-made published word list
3. A combination of published and self-made word lists

Display This Question:

If Please indicate which type of word list(s) you used. = Self-made word list

Q6a.1 What were your reasons for making your own word list rather than using a ready-made published one? (more than one option can be selected)

- I could not find a published word list suitable for my teaching situation.
- I did not have access to published word lists.
- I was concerned about the validity and reliability of the ready-made word lists.
- I did not like the fact that there was no explicit description of how some published word lists were made.
- Other (please explain) _____
- Further comments _____

Display This Question:

If Please indicate which type of word list(s) you used. = Self-made word list

Q6a.2 Rank the importance of the following criteria for selecting and ordering words in a word list. If you have not used any, please select "Low importance"

	High importance	Medium importance	Low importance
Frequency (the number of word occurrences in a text)	•	•	•
Range (the number of different texts a word occurs in)	•	•	•
Dispersion (the number of word occurrences across a range of different texts)	•	•	•
Intuitive judgments	•	•	•
Surveys, questionnaires and/or interviews	•	•	•
Combination of published word lists	•	•	•
Other (please specify)	•	•	•

Display This Question:

If Please indicate which type of word list(s) you used. = Self-made word list

Q6a.3 Why did you follow the above criteria?

- I followed the criteria of a published researcher.
- I combined the criteria used by a number of researchers.
- I used the criteria a particular software provided me with.
- Other (please explain) _____

Display This Question:

If Please indicate which type of word list(s) you used. = Ready-made published word list

Q6b.1 Why did you select this word list(s)? (more than one option can be selected)

- It was suitable for my purpose and context.
- It was the only list I had access to.
- It was available in electronic form.
- It was the only list available.
- It was a widely used word list.
- It was the only list I knew.
- It was imposed by someone with authority.
- It was recommended by a colleague.
- Other (please explain) _____
- Further comments _____

Display This Question:

If Please indicate which type of word list(s) you used. = Ready-made published word list

Q6b.2 Did you make changes to the list?

- Yes
- No

Display This Question:

If Did you make changes to the list? = Yes

Q6b.2.1 What type of change?

- Addition
- Deletion
- Change of order
- Other (please explain) _____
- Further comments _____

Display This Question:

If Did you make changes to the list? = Yes

Q6b.2.2 Why did you make those changes? (please elaborate on your answer if you wish)

- To make the list suitable for my purpose and context. _____
- To make the list more practical. _____
- Other (please explain) _____

Display This Question:

If Please indicate which type of word list(s) you used. = A combination of published and self-made word lists

Q6c.1 What were the reasons for combining a ready-made word list with a self-made one? (more than one option can be selected)

- The ready-made word list could not be used alone for my purpose and context. (Please elaborate) _____
- The self-made word list could not be used alone for my purpose and context. (Please elaborate) _____
- There were different word lists for my purpose and context and combining them would be efficient.
- Other (please elaborate) _____
- Further comments _____

Display This Question:

If Please indicate which type of word list(s) you used. = A combination of published and self-made word lists

Q6c.2 How did you combine the word lists?

- The ready-made word list was the main list and words were added. (Please elaborate) _____
- The self-made word list was the main lists and words from a ready-made list were added. (Please elaborate) _____
- Other (please explain) _____

Q7 What is the name of the word list(s)? Please provide a brief description of the list if it is a self-made one.

Q8 Did you use other resources in addition to word lists to achieve your purpose? If yes, please explain.

End of Block: Using Word Lists

Start of Block: Evaluating Word Lists

Q9 Overall, how would you rate the usefulness of using word lists for your purpose?

- Useful
- Neither useful nor useless
- Useless

Q10 Which of the following characteristics do you focus on when making/selecting a word list? It should be ... (more than one option can be selected, please justify your choice)

- Based on corpus evidence.
- Based on statistical measures for selection and ranking.
- Based on subjective judgment for selection and ranking.
- Small in size and/or divided into parts.
- Topic related.
- Other (please specify) _____
- Further Comments _____

Q11 Select the problems you faced when you used a word list(s) for your purpose. (more than one option can be selected)

- Access
- Size of the word list
- Cost
- Format
- Lack of explanation
- No problems
- Other (please specify) _____
- Further comments _____

End of Block: Evaluating Word Lists

Start of Block: Block 3

Q12 Please select whether you would like to be contacted for further stages of the research or receive a summary of the findings. Further stages will consist of interviews, further questionnaires and/or testing the word list evaluation framework which will require another consent. Please note that your agreement to be contacted does not mean that you are as yet consenting to participate.

I agree to be contacted for further stages of the study.

I would like to receive a summary of the findings.

Q13 Do you have further comments on using word lists you would like to report. (optional)

Q14 Do you have comments on the questionnaire you would like to report. (optional)

End of Block: Block 3

APPENDIX C

The Initial Version of the Evaluation Tool (Phase 2)

A Framework for Corpus-based Word List Evaluation for Pedagogical Purposes

Section 1: General attributes of the word list

Provide a brief overview of the main attributes of the corpus-based word list under evaluation. **Table 1** (Excel Doc) summarizes the most popular corpus-based word lists of different types of vocabulary published between 2000 and 2018 (in addition to West's 1953 list). If you cannot find your list in Table 1, use the following table (Table 2) to organize information about your list.

Table 2
The General Attributes of the Word List

Name of the list	Author(s), Year (if published)	Type of vocabulary	Size; Lemma/ Word Family/ word types/ etc.	Corpus details	Criteria	Target audience and purpose	Notes

Section 2: Context Profile

Describe your context in which the list under evaluation will be used. Information about your purpose, audience and target vocabulary would provide a basis for evaluating the list.

A. Which of the following describes your purpose in using the word list? Select **one** of the following uses.

- a. Course/program design
- b. Teaching and learning
- c. Testing (assessment)
- d. Materials development
- e. Research (related to one of the above purposes)
- f. Other

Table 3 explains the uses specified above (adapted from Burkett, 2017, p.195)

B. What is the target type of vocabulary?

- a. General English vocabulary
- b. General Academic vocabulary
- c. Specialized Academic vocabulary
- d. Professional or occupational vocabulary
- e. Other

Table 3

Examples of Word Lists in English Language Teaching and Learning Contexts

Course design	Setting short term (course) learning goals
	Setting long term (program) learning goals
Teaching and learning	Selecting material that includes the target vocabulary
	A resource for intentional vocabulary learning activities
	Selecting vocabulary from texts to focus on
Testing	Analysing the vocabulary load in a text
	Modifying the level of vocabulary in tests
	Selecting words for vocabulary size and proficiency tests
	Testing list vocabulary in context
Materials development	Test results inform teaching and learning
	Setting the vocabulary level of materials
	Creating vocabulary focused class teaching materials
	Creating vocabulary focused class independent study resources

C. Characteristics of the target audience with whom the list will be used.

Now, provide a description of the main characteristics of your audience.

1. What is the age range of the target audience?

- a. Less than 5 years
- b. Between 5 - 15 years
- c. Between 16 -18 years
- d. More than 19 years

2. What is the proficiency level of the target audience?

- a. Proficient
- b. Advanced
- c. Upper-Intermediate
- d. Intermediate
- e. Pre-Intermediate
- f. Elementary
- g. Beginner/False Beginner
- h. Unassessed

Table 4 describes the levels above (adapted from Exam English)

Table 4

Description of the Proficiency Levels

Level	European Framework	Description
Proficient	C2.	The capacity to deal with material which is academic or cognitively demanding, and to use language to good effect at a level of performance which may in certain respects be more advanced than that of an average native speaker.
Advanced	C1	The ability to communicate with the emphasis on how well it is done, in terms of appropriacy, sensitivity and the capacity to deal with unfamiliar topics.
Upper-Intermediate	B2	The capacity to achieve most goals and express oneself on a range of topics.
Intermediate	B1	The ability to express oneself in a limited way in familiar situations and to deal in a general way with non-routine information.
Pre-Intermediate	A2	An ability to deal with simple, straightforward information and begin to express oneself in familiar contexts.
Elementary Beginner	A1 A0-A1	A basic ability to communicate and exchange information in a simple way

3. What is the discipline of the target audience?

- a. Hard sciences (e.g., Maths, Engineering, Medicine)
- b. Soft sciences (e.g., Business, Law, History)
- c. Professional and Occupational fields (e.g., nursing communication)
- d. Mixed
- e. Unspecified

4. Other characteristics relevant to the evaluation

Section 3: The Evaluation

Answer each of the following evaluative questions using the rating scale 1 to 3. Please note 1 indicates a “not at all” response, 2 indicates a “somehow” response, 3 indicates a “most definitely” response. If a question is not applicable or a response is not known, choose N/A.

Where relevant, there are explanations of technical terms and concepts and guidance on how to approach some of the questions in yellow boxes and blue font.

A. How are words grouped as a unit in the list?

Please select the word unit used to group words in the list under evaluation. Refer to Table 5 for an explanation of these word units and Table 5 for their potential uses, then answer the following questions.

- a. Words are group by **Type**
- b. Words are grouped by **Lemma**
- c. Words are grouped by **Flemma**
- d. Words are grouped by **Word Family**

Table 5

The Most Common Word Units Used in Corpus-based Word Lists

Word unit	Definition	Example	Possible purpose and audience	Notes	WL Example
Type	A different form is a different word.	The sentence <i>It is not easy to say it correctly</i> contains 7 types (different words). <i>It</i> is counted once.	Type-based word lists are suitable for productive vocabulary development (i.e., writing or speaking); technical vocabulary learning, because one from of the word may be technical but not the other related forms; Spelling;	Type-based lists are usually large.	Applied Linguistics Academic Word List (Khani & Tazik, 2013)
Lemma	A headword together with its inflected forms that are of the same part of speech (POS). Up to eight inflectional suffixes in English.	The lemma <i>DEVELOP</i> consists of <i>develops</i> , <i>developed</i> , and <i>developing</i> .	Lemma-based lists are suitable for productive vocabulary development while learners have not mastered the derivations. If they know the headword, usually they would know its inflections of the same POS as the headword (McLean, 2017). Beginners with less morphological knowledge.	Too restrictive to very lowest proficiency learners	The Academic Vocabulary List (Davies & Gardner, 2013)
Flemma	A headword is grouped with its inflected forms of different POS.	The verb and noun <i>walk</i> belong to the same flemma.	A Flemma-based list would be suitable for lower-intermediate and intermediate learners.		The Essential Word List for Beginners (Dang & Webb, 2016)
Word family	A headword is grouped together with its inflectional and derivational forms.	The word family <i>develop</i> consists of <i>develops</i> , <i>developed</i> , <i>developing</i> , <i>undeveloped</i> , <i>underdeveloped</i> , <i>development</i> , <i>developments</i> , <i>developer</i> , and <i>developers</i> .	Word-family based lists are suitable for receptive vocabulary development (i.e., listening and reading comprehension) since the meaning of inflected form (e.g., <i>businesslike</i>) can be guessed if the meaning of the base is known <i>business</i> (Nation, 2016, p. 8); Suitable for developing tests, course design and graded readers. Requires morphological knowledge acquired later in the learning. The L1 background of learners affects affix knowledge. It is the most common unit, thus it is used for word list comparison or exclusion of other type of vocabulary.	Word-family lists can be short and eliminate repetition of words. It is widely used by programs like Range and AntWord Profiler.	The Academic Word List (Coxhead, 2000)

Now, rate the suitability of the word unit referring to Table 4		1	2	3	N/A
1.	Is the word unit (<i>Type, Lemma, Flemma, or Word Family</i>) of the word list suitable to your purpose?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.	Is the word unit of the word list suitable for the proficiency level of your audience?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3.	Are issues related to the definition of the “word” unit such the UK and US spelling, alternative spellings, part of speech, abbreviations and numbers dealt with appropriately to your context?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

For questions four and five, the response would be NO (1), YES (2) or not known (N/A).

	1	2	N/A
4. If you think the word unit is not appropriate for your purpose and/or audience, is it well-justified?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Are there different versions of the list with different definitions of the “word”?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

➤ **NOTES ON QUESTIONS 4 AND 5**

Sometimes, the word unit selected when developing a word list is not appropriate for the target purpose and/or audience. It is important to note that the word list developer may have been constrained by practical considerations such as:

- Some non-commercial corpus software packages are limited to a particular word unit.
- The developer might use a published list based on the unit not suitable for their purpose while developing their own list.
- For some purposes like the comparison between lists, the word unit has to be the same even if not appropriate.

➤ **Pedagogical solution:**

Some word list developers have started publishing their lists in different versions with different word units, and teachers can select the version they need.

B. What is included in the main list?

This section evaluates what has been included in the main list and what has been listed in separate lists according to your context and purpose.

1 = “not at all”; 2 = “somehow”; 3 = “most definitely”

	1	2	3	N/A
6. Is the description of what forms are counted as words suitable to the purpose?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Are homonyms and homographs dealt with according to the purpose?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. Are proper names dealt with, including proper name homoforms, according to the purpose?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. Are content bearing proper names distinguished?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. Are hyphenated words dealt with in a consistent well-justified way?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11. Are transparent compounds dealt with in a way consistent with hyphenated words?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12. Are acronyms dealt with, including acronym homoforms?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13. Are the proper name lists and other lists revised on the basis of initial output?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
14. How satisfactorily are marginal words dealt with?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

➤ NOTES ON QUESTION SEVEN:

- Homonyms are two or more completely different meanings of the same spoken and written forms as in *kind* (generous) and *kind* (type).
- Homographs are two or more completely different meanings of the same written forms (different spoken form) as in *present* (current) and *present* (gift).

It is important to distinguish members of homonyms and homographs in separate word families if the word list is for vocabulary learning purposes or to identify vocabulary levels and sizes since different meanings need to be learned. This is not necessary if the word list is for spelling purposes.

One meaning more frequent than the other.

➤ NOTES ON QUESTIONS EIGHT AND NINE:

Proper nouns should be counted as a separate word list as they are semantically different from other words.

Content bearing proper nouns like *Gandhi* and *Einstein* need to be listed separately as they require previous knowledge to successfully understand a text.

Transparent proper names like *Quality* or *Meat* [capitalised] should be counted as occurrences of the common noun, especially if the purpose of the list is vocabulary learning.

➤ **NOTES ON QUESTION TEN:**

Hyphenated words are simply words with a hyphen. The item before or after the hyphen could be an affix, part of a word or a word that can stand alone. There are many reasons for using hyphens, but generally, the overarching reason is to show a close relationship between the hyphenated items.

Hyphenated words, transparent compounds and two separate words with space need to be dealt with in a consistent, well-justified way.

In dealing with compound words, a decision needs to be made whether hyphens are seen as letters and thus considered part of the word, or the opposite and thus replacing the hyphen with space while editing the corpus. The latter decision shows grammatical relationships and compounding and is used in duplicated words, yet some compound words exist as both hyphenated and not. It would also mean that two words are counted which can have a negative effect in syllabification and with prefixes and suffixes. Counting a hyphen as part of the word, on the other hand, would result in fewer tokens and more types.

Transparent compounds: these are words written as one word (with no hyphens or spaces within) where the meaning is connected to the meaning of each part (which are existing words).

They should be counted separately since their meaning is known by knowing its parts. Even the hyphenated ones are counted as two words. On the other hand, the argument against this decision is that the meaning of the compound is lost in the split, but this is the case of counting inflected and derived forms under families.

If counted as one family, then under which part of the word? The solution to this problem should be related to the purpose of the count.

Marginal words are words, commonly found in spoken discourse, to fill pauses and express emotions but could not be considered as words in their own as they have little referential meaning. For example,

Erm

Arrgh

C. What material (corpus) is the list based on?

Corpus-based word lists are based on a collection of texts from which words are selected and organized. The quality of the resulting list depends on the corpus.

To evaluate the corpus, you need to have information about the kind of texts collected, and their representativeness, the size of corpus, the balance of the corpus, the length of texts and how many corpora were used (Coxhead, 2000).

- Do you have access to information about the corpus the list is based on?

If YES (you have access to information about the corpus, but not necessarily all aspects of the corpus), then answer the following questions.

If NO, move to the next section.

	1	2	3	N/A
15. Does the <i>content</i> of the corpus represent the actual or potential language your audience is likely to be exposed to?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
16. Is the <i>content</i> of the corpus suitable to the proficiency level and age of your audience?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
17. Are factors like text type, geographical division, the age of the audience and learning situation considered according to your purpose?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
18. Is the <i>size</i> of the corpus suitable to get reliable results for the intended purpose?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
19. Is the corpus divided into sub-corpora? If yes,	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
20. Are the sub-corpora large enough?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
21. Are the sub-corpora of equal size?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
22. Are the sub-corpora coherent?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

➤ NOTES

Content: A word list can be said to be good if it is based on a corpus that represents that actual or potential language the target audience will likely encounter. However, it is not possible to collect *all* instances of the target language, e.g. academic English. Consequently, we must rely on a **sample** that represents, as closely as possible, the language used by the target population.

A corpus could include textbooks or textbook chapters, journal articles, literary texts, graded readers, teaching materials, audio-recorded conversations or lectures or movies, among other texts. To evaluate the suitability of the corpus, you need to think about your purpose and context. For example, if the intended audience for the list is young children, then the list should be based on a corpus of texts that those children are likely to encounter.

If the proficiency level of the intended audience is low, and they are learning English as a second language, then the corpus should include everyday conversation.

Size: There is little agreement on the size of the corpus or what comprises a large or small corpus. The size of the corpus depends very much on the purpose of the resulting list and other practical considerations. As a rule of thumb, a large corpus is generally considered to be better, yet it is possible to get much useful data from a small corpus, particularly when investigating high-frequency items that occur often or when investigating the vocabulary which would occur frequently in a specialised text (Coxhead, 2018).

Another example is when developing a list based on the teaching materials in a particular teaching context. In this case, the small pedagogic corpus built based on the teaching material would result in a reliable and useful list (Coxhead, 2018).

It is also worth bearing in mind that there are practical issues that may constrain the size of the corpus, such as:

- Size limits set by some commercial corpus software.
- Spoken corpora tend to be smaller than written corpora since collecting and transcribing spoken texts is time-consuming. According to O'Keeffe et al. (2007, p. 4), over a million words of speech is considered a large corpus, whereas anything under five million words of written text is quite small.
- Published texts may be limited by what you can obtain permission for from the copyright holder.

Proportion of sub-corpus: Dividing a corpus into sub-corpora creates range and dispersion. There are three principal ways to divide a corpus: (1) the sub-corpora should be large enough to allow the target words to occur, (2) the sub-corpora should be of equal sizes to give an accurate range, and (3) each of the sub-corpora should be coherent. That is, of the same type for range to be accurate.

D. How are words selected from the corpus and ordered into the list?

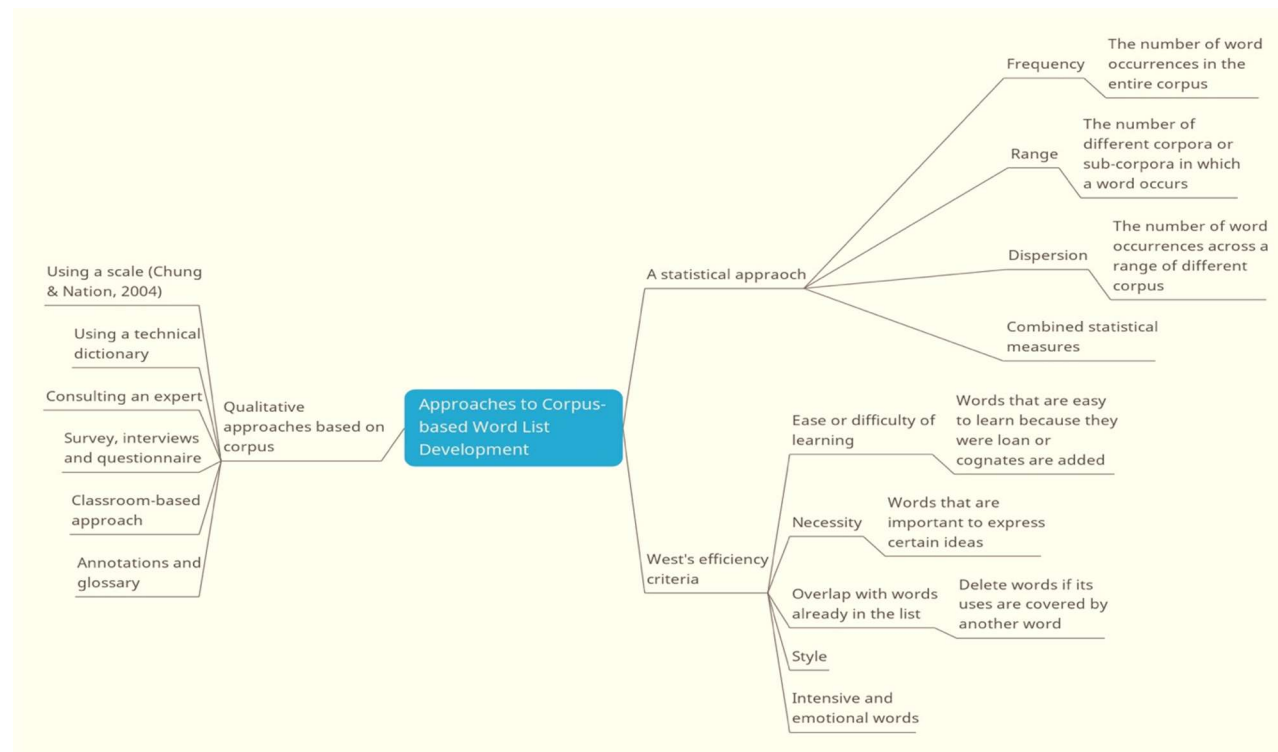
Corpus-based word lists are developed by identifying and ordering words from a corpus following specific criteria.

Diagram 1 shows the common criteria employed to develop corpus-based word lists: (a) strictly using a statistical criterion or a combination; (b) using statistical criteria then adjusting the results subjectively (e.g., West's efficiency criteria); or (c) using qualitative data/subjective judgements. It is important to note that those criteria are sometimes combined and new criteria may be employed to serve the purpose of the list which makes replication of the criteria impossible.

Please select the criteria used to develop the list under evaluation from the following options. Refer to Table 4 for an explanation of these criteria. If you cannot find information about the development of your list and it is not one of the top word lists in Table 1 (section one), then skip these questions.

Describe how the list is made.

- A single or a combination of statistical criterion (For example, frequency, range, dispersion)
- Statistical results tailored subjectively
- A purely subjective
- Other



D. How appropriate are the criteria used to develop the word list to your purpose?	1	2	3	N/A
23. How suitable are the criteria for inclusion and ordering in the list (e.g., frequency, range, dispersion, or some composite measure) for getting reliable results?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
24. If subjective criteria are used, how well are they formalized and applied?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
25. Are the criteria for making the sub-list clearly described and justified, if any?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
26. Are the lists checked against competing lists for coverage and overlapping and non-overlapping words?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
27. Are other supplementary lists used to exclude or add item from the list? How?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

➤ NOTES

There are advantages and disadvantages to each approach to word list development, but the criteria have to be evaluated against the purpose and context.

Word lists based on statistical measures are claimed to be more reliable and accurate than those based on intuition (Szudarski, 2018). Nation notes that a large truly representative corpus is often impossible, and that most corpus software tools are still limited, and therefore to produce useful pedagogical word lists, it might be necessary to use subjective criteria combined with statistical measures (Nation, 2016) taking into account the target audience, context and purpose. This corresponds with the respondents' opinion in the survey of making word lists. In fact, most respondents argue that teacher should be acquainted with word lists development so that they are able to tailor the lists to their context and purposes. It has often been assumed that quantitative analysis of corpus is more valid than the intuition of the expert native speakers. However, many corpus linguists (e.g., O'Keeffe et al., 2007) highlight the equal importance of quantitative and qualitative approaches (Szudarski, 2018, p. 29) especially if the purpose is to make a useful word list suitable to the learners. The GSL is the most common example of adjusting statistical results subjectively in an attempt to make the list as efficient and complete as possible to the target users. Despite its age, the GSL still achieves high lexical coverage in different corpora (Dang & Webb, 2016a).

Coxhead used the GSL to exclude general high-frequency word from her corpus to make a list of academic words. Others have used this step to develop word lists for different purposes. The disadvantage of this step is that it takes the weakness of the supplementary word list. So the quality of the resulting list will be affected by the

quality of the supplementary list. On the other hand, Gardner and Davies did not use other lists and used statistical criteria to exclude general high-frequency list to make their academic word list.

It is important to note that identifying different types of vocabulary may require different criteria. For example, specialized word lists may require different criteria such as the judgment of a specialist informant.

The most important point is that the criteria are formalized and applied consistently.

E. How can the list be implemented in context?

Answering those questions is not a simple yes-no. It is important to think how those elements will serve your purpose and whether their lack of existence would lead to problems.

It is really important to have access to information about the list and its development, particularly if the list has been developed for a specific purpose.

How appropriate the criteria used to develop the word list are to the purpose of the list?	1	2	3	N/A
28. Is the size of the list manageable for your intended purpose?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
29. Is the context for each word provided (e.g., through concordance lines or example sentences)?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
30. Are there supplementary materials based on the list to be used in class?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
31. Do you have access to an electronic version of the list so that you can edit it according to your purpose?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
32. Is there guidance on how to use the list?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
33. Are the weaknesses of the lists acknowledged and justified?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

➤ NOTES ON QUESTION TWENTY-NINE:

It has been argued that vocabulary can be learned if provided in context. To decide on the importance of context, you need to consider if it would be important to your context and whether the context would be useful and accessible to your audience.

if “yes”, the context is provided, how good is it to your context and purpose, and on what basis is it selected. If not, does it really matter? If your purpose and context requires presenting word in context, do you have access to the corpus the list is based on?

For teaching and learning purposes, the context would be more important than using a word list for materials development or testing. However, the form of context might not be suitable to the proficiency level of your audience or their needs which means that the teacher providing the context might be more useful. Kury (2017) provides a step-by-step guide on making word lists from the reading material and adding the senses and patterns of the words.

➤ **NOTES ON QUESTION THIRTY:**

Similar to question 29, deciding on the importance of the supplementary materials depends on your purpose and context and whether those material would be useful in your case. The availability of materials based on a word list might be useful for replication or preparing materials of similar kind.

Comments:

APPENDIX D

Interview Questions (Phase 3)

Warm-up questions: Professional background

- Could you talk about your experience in using word lists? What was your role at the time?
- Have you ever made an evaluation of a word list?

If yes, can you describe the process of the evaluation? Was it part of the list's development? Was it to make a decision?

(They are expected to talk about why there was an evaluation; what word list; who decided to make the evaluation; tools used, if any).

Now, let us have a look at this copy of the proposed framework:

Clarity

If changes are made and received before the interview:

Discuss them.

If NOT:

- Are the items and guidelines clearly worded? *If the items are not clearly worded, what needs to be changed to improve them?*
- Is the wording of the response scale clear? If no, what needs to be changed?

Relevance

- How helpful do you think that framework is for determining the usefulness of a corpus-based word list? For specific pedagogical purposes you can think of e.g., conduct vocabulary size test.
- How applicable is this framework to all the different types of word lists you think of? Do you think this framework is applicable to evaluate general vocabulary word lists, academic word list and specialised word lists?
- Does it help you to assess the data (corpus) on which the word list is based?
- Does the framework appear to be a good measure of the elements necessary for the development of a good corpus-based word lists developed for a specific pedagogical purpose? Please explain.
- Do you think the framework would guide you to develop a corpus-based word list?

Accessibility

- How helpful do you think the guidelines are to guide the non-expert evaluator in making a decision/interpreting the item?

How helpful do you think the guidelines would guide you in making a decision/interpreting the item?

- Can you suggest ways to improve the guidelines?

Practicality/organization?

- Do you think the framework is practical in terms of length, logical flow, style and format?
- Is the interpretation scale appropriate? If no, what needs to be changed?

Conclusion

- Are there any items that should be added and/or deleted to the framework? If yes, please explain.
- What changes, if any, do you consider necessary to improve the framework?
- Is there anything you want to say?

Note: The grey coloured questions were not discussed with all participants

APPENDIX E

Coding Scheme (Phase 3)

Theme	Definition	Code
WORDING	Responses related to word choice/vocabulary/terminology in the questions, notes and instructions.	<p>Clear wording; positive comments about the wording.</p> <p>Reword with better (more common) terms for clarity.</p> <p>Revise technical terms for ACCESSIBILITY; recommendations to revise technical terms that would be difficult for non-experts to understand.</p> <p>Explain technical terms for ACCESSIBILITY; recommendations to provide a definition, example or explanation.</p>
USABILITY	Responses related to how easy and practical the framework is to use. These are related to the questions about the practicality and organisation of the framework.	<p>Long</p> <p>Complicated logical flow; the organisation of the notes and table is complicated.</p> <p>Limitation refers to problems with the framework in terms of how it could be used or what contexts or word lists could be evaluated when using it.</p> <p>Scale appropriateness</p> <p>Impractical</p> <p>Give instructions on how to use the document.</p>
ACCESSIBILITY	Responses related to how accessible the framework is to non-expert practitioners. These are related to what information was included (content), how that information was presented (usability and wording) and whether the finished framework met expectations (usefulness)	<p>Important content</p> <p>Give examples</p> <p>Accessible for the non-expert</p> <p>Not relevant to practitioners; non-experts may not be interested.</p> <p>Improve accessibility</p>
CONTENT	Responses related to the information presented in the framework.	<p>Component is important to a specific audience and should be moved. Refers to components that were important but should be presented in a separate reference document to improve the logical flow and length of the framework.</p> <p>Irrelevant items; component items whose relocation or deletion were suggested because they were either irrelevant or difficult.</p> <p>Add items</p> <p>Component fits but revise; the items is appropriate but needs revision.</p>

Theme	Definition	Code
		Component fits but explain Too detailed; the level of details is high. Overlap Add references for authorship acknowledgement and further reading.
USEFULNESS	Responses related to the framework's usefulness as a tool	Useful Not sure; review is reluctant to account for usefulness. Limitation
IMPLEMENTING WORD LISTS	Responses in which participants reported how word lists are used in the classroom, what problems word lists trigger and their attitudes towards word lists.	Issues in using word lists Negative attitude towards word lists Practices Awareness

APPENDIX F

Revisions to the Proposed Tool Based on the Interviews (Phase 3)

The table below summarises the main changes to the tool's content and wording. The first column presents the proposed tool components, the second one presents the revisions/changes made and the last one shows the type of revision, important notes, or clarifications. Within the last column, deletions are denoted by *D*, relocations to a separate document by *S*, item modifications (to add content or examples or reword the item) by *M* and retention without change by *R*.

Proposed Tool	Revised Tool	Notes on Revision
	Introduction to Word lists Document (A separate PDF document)	Added (from the yellow boxes and literature)
	Introduction to the Evaluation Tool How the evaluation too is organised	Added
Section 1: General attributes of the word list	Section 1: General attributes of the word list	R
Section 2: Context profile Which of the following describes your purpose in using the word list? Select one of the following uses.	What is the purpose for using the examined word list? Describe the purpose(s) for using the word list. In doing so, think about the most important aim to be accomplished by using the list.	(M) Selecting more than one option is now allowed.
Table 3 Examples of the uses	Table 2	(R)
What is the target type of vocabulary? a. General English vocabulary b. General Academic vocabulary c. Specialized Academic vocabulary d. Professional or occupational vocabulary e. Other	What is the target vocabulary type in the context? a. English for General Purposes b. English for General Academic Purposes c. English for Specialised Academic Purposes d. English for Professional or Occupational Purposes e. Other (please specify)	(M) The wording of the vocabulary type was changed to use more common terms.
Characteristics of the target audience	Characteristics of the target learners	
What is the age range of the target audience? a. Less than 5 years b. Between 5 - 15 years c. Between 16 -18 years d. More than 19 years		(D) This question was deleted in full.
What is the proficiency level of the target audience? a. Proficient b. Advanced	What is the proficiency level of the learners? a. Proficient b. Independent	(M) This question's position was moved forward.

Proposed Tool	Revised Tool	Notes on Revision
c. Upper-Intermediate d. Intermediate e. Pre-Intermediate f. Elementary g. Beginner/False Beginner h. Unassessed	c. Basic d. Beginner or False Beginner e. Unassessed or Mixed	The classification of the proficiency levels was limited to five options. Explanations of the levels were provided.
Table 4 <i>Description of the Proficiency Levels</i>	Table 3 Descriptions of proficiency level based on CEFR	(M) A note was added on how those classifications are intended to be useful guidelines but not strict delineations.
What is the discipline of the target audience? a. Hard sciences (e.g., Maths, Engineering, Medicine) b. Soft sciences (e.g., Business, Law, History) c. Professional and Occupational fields (e.g., nursing communication) d. Mixed e. Unspecified	What is the academic discipline of the learners? a. Hard sciences (e.g., Maths, Engineering, Medicine) b. Soft sciences (e.g., Business, Law, History) c. Professional and Occupational fields (e.g., nursing, communication) d. Mixed e. Other (please specify) f. Unspecified g. Not relevant	(M)
Other characteristics relevant to the evaluation: ...	Specify other characteristics relevant to the evaluation, such as cultural background and first language.	(M)
Section 3:	Rubric deleted	
How are words grouped as a unit in the list? Please select the word unit used to group words in the list under evaluation. Refer to Table 4 for an explanation of these word units and their potential uses, then answer the following questions. A. Words are grouped by Type B. Words are grouped by Lemma C. Words are grouped by Flemma D. Words are grouped by Word Family	Brief definitions and examples were given. The explanatory table <i>The Most Common Word Units Used in Corpus-Based Word lists</i> was moved to the notes section.	

Question 4 was modified.

Proposed Tool	Revised Tool	Notes on Revision
		Question 5 was deleted.
Section B: what is included in the main list?		The whole section was deleted
C. What material (corpus) is the list based on?		
15. Does the <i>content</i> of the corpus represent the actual or potential language your audience is likely to be exposed to?	1. Does the <i>content</i> of the corpus represent the actual or potential language that the target learners will likely be exposed to?	(M)
16. Is the <i>content</i> of the corpus suitable to the proficiency level and age of your audience?	2. Is the <i>content</i> of the corpus suitable for the proficiency level and age of the target learners?	
17. Are factors like text type, geographical division, the age of the audience and learning situation considered according to your purpose?	3. Is the <i>size</i> of the corpus suitable to obtain reliable results for the intended purpose?	
18. Is the <i>size</i> of the corpus suitable to get reliable results for the intended purpose?	4. Is the <i>age</i> of the corpus suitable for the intended purpose?	
19. Is the corpus divided into sub-corpora? If yes,	5. Is the <i>location</i> of the corpus suitable for the intended purpose?	
20. Are the sub-corpora large enough?		
21. Are the sub-corpora of equal size?		
22. Are the sub-corpora coherent?		
How suitable are the criteria for inclusion and ordering in the list (e.g., frequency, range, dispersion, or some composite measure) for getting reliable results?	6. How suitable are the criteria for inclusion and ordering in the list from the viewpoint of obtaining reliable results?	(M)
If subjective criteria are used, how well are they formalised and applied?	7. If subjective criteria have been used, how well have they been formalised and applied?	
Are the criteria for making the sub-list clearly described and justified, if any?	8. Have the lists been checked against similar lists for coverage and overlapping and non-overlapping words?	
Are the lists checked against competing lists for coverage and overlapping and non-overlapping words?		
Are other supplementary lists used to exclude or add item from the list? How?		
Is the size of the list manageable for your intended purpose?	9. Is the list size manageable for the intended purpose?	(M)
Is the context for each word provided (e.g., through concordance lines or example sentences)?	10. Is the context for each word provided (e.g., through concordance lines or example sentences)?	
Are there supplementary materials based on the list to be used in class?		

Proposed Tool	Revised Tool	Notes on Revision
Do you have access to an electronic version of the list so that you can edit it according to your purpose?	11.Are there supplementary materials based on the list for use in the class?	
Is there guidance on how to use the list?	12.How useful is each entry in the list? Are there items that need to be deleted or added?	
Are the weaknesses of the lists acknowledged and justified?	13.Is there guidance on how to use the list?	
	14.Are the weaknesses of the list acknowledged and justified? How problematic are they to your context?	
	15.Are factors such as text type, geographical location, cultural consideration and learning situation considered based on the intended purpose?	

APPENDIX G

Introduction to Corpus-based Word Lists (Phase 3)

© 2020 Sarah Alzeer

WHAT IS A CORPUS-BASED WORDLIST?

A **wordlist** is a sorted list of single words or multi-word units grouped by their frequency counts in a given **corpus** (a large collection of naturally occurring language), sometimes combined with other measures. It may be called a vocabulary list or frequency list.

The main **purpose** of making wordlists is to select vocabulary to teach based on the assumption that the more frequent a word is, the more useful it is. Focusing on those words will enable learners to understand much of what they encounter and produce what they know.

There are different **types of wordlists** based on the vocabulary they present

General vocabulary or high-frequency wordlists
such as West's (1953) General Service List (GSL)

General academic wordlists
such as Coxhead's (2000) Academic Word List (AWL)

Specialised, subject-specific or technical wordlists
such as Lei and Liu's list of specialised terms (2016) in medicine

HOW ARE WORDLISTS ORGANIZED?

Wordlists can be organised by different levels of word units, namely, **type**, **lemma**, **flemma**, or **word family**.

TYPE

Each sequence of letters separated by space or punctuation counted once. *Counting1 words2 is3 difficult4 but5 it6 is fun7.* The sentence contains eight word forms but seven types. The type (is) is counted once.

LEMMA

A lemma consists of a stem *develop*, together with its inflected forms *develops*, *developed*, and *developing* belonging to the same part of speech.

FLEMMMA

A flemma is a more inclusive version of the lemma. It includes different parts of speech so that *developing* in "*developing (adj) countries*" and "*They are developing (v) a wordlist*" are one flemma.

WORD FAMILY

A stem grouped together with its **inflectional** and **derivational** forms. The word family *develop* consists of *develops*, *developed*, *developing*, *undeveloped*, *underdeveloped*, *development*, *developments*, *developer*, and *developers* belonging to different parts of speech.

HOW ARE WORDLISTS MADE?

Wordlists are typically made by choosing or building a corpus of appropriate texts and then selecting words from that corpus employing specified criteria. The common criteria include one or more of:

- **Frequency**: the number of occurrences of a word in the entire corpus.
- **Range**: the number of different texts or sub-corpora that a word occurs in.
- **Dispersion**: the number of a word's occurrences across different texts or sub-corpora.

Statistical results are often adjusted to make the list more suitable and useful for a particular group of learners.

In some cases practitioners combine ready-made wordlists to create a wordlist for their own purposes.

WHAT WORDLISTS ARE USED FOR?



Course Design

Setting short and long term learning goals. Which and how many words to be learned at each learning stage?



Classroom Teaching and Learning

A resource for structured vocabulary instruction and independent learning.



Materials Development

A resource for creating vocabulary focused materials and assessing vocabulary level in teaching materials.



Testing

A resource for developing vocabulary size and learning tests.



What is a good wordlist?

A one-size-fits-all wordlist does not exist, but the best wordlist is one that takes into account the purpose and context of use.

APPENDIX H

The First Iteration of the Evaluation Tool

Corpus-Based Word List Evaluation Tool for Pedagogical Purposes

Introduction to the *Evaluation Tool*

The *Evaluation Tool* is designed to help you evaluate the suitability of corpus-based word lists to be used for pedagogical purposes. It is based on the observation that a one-size-fits-all word list does not exist. Contextual factors are likely to differ from case to case. Therefore, the tool considers these factors in determining the usefulness of a word list and how it can be adapted for maximum utility.

The tool is developed based on and guided primarily by (a) Nation's (2016) critique Framework, (b) a review of the relevant literature on word list production and use for pedagogical purposes, and (c) a survey of practices and views related to the word list usage in English language teaching (ELT).

The target users of the tool are, broadly speaking, ELT practitioners, such as teachers, curriculum or assessment coordinators, or materials developers involved in directing vocabulary acquisition. They may range from (a) practitioners deeply involved in curriculum development and coordination, often in a management position at an institution and who may or may not be involved in the implementation of word lists, to (b) teachers with varying levels of teaching experience who implement word lists in their teaching and may discuss and help with materials development but have no management role. The tool aims to cater to practitioners with different levels of expertise and knowledge, and especially those who are unfamiliar with the intricacies of developing corpus-based word lists. It summarises the debates in the field to help an evaluator consider the most important elements associated with word list usage.

The tool is designed for the evaluation of word lists containing single-word units.

How the *Evaluation Tool* is organised

The tool is divided into three sections:

Section 1: General Attributes of the Word List

In this section, information about the word list under examination is elicited in order to prepare you for the evaluation.

Section 2: Context Profile

In this section, you should input the purpose of using the examined list, the type of vocabulary targeted in the learning setting, and the learners' characteristics. This information will serve as the basis for evaluating the list and tailoring it for maximal usefulness.

Section 3: Evaluation

This is the core of the tool, and it includes 19 evaluation questions adapted from Nation's tool and categorised into four groups:

- Word unit
- Corpus
- Criteria for word list development
- Pedagogical considerations

At the end of the document, you will find a set of Notes that provides key information about word lists, guided answers, debates, and arguments. A brief summary of the information is provided for the benefit of practitioners who are unfamiliar with the intricacies of developing corpus-based word lists. An example is also presented at the end to help you use the *Evaluation Tool*. In this example, an evaluator uses the tool to evaluate an actual word list, click [here](#). A list of references is also provided.

Section 1: General Attributes of the Word List

This section prompts you to create a brief overview of the main attributes of the examined word list. You can either locate the information in the attached Excel file (adapted from Burkett, 2017) which summarises the most popular word lists of different types of vocabulary published between 2000 and 2018 (in addition to West's 1953 list), or, if the target list is not included, then you should fill out the following table.

Table 1: The examined word list's attributes

Name of the list	Author(s), Year (if published)	Vocabulary type	Size of word list and word unit	Corpus details	Criteria used	Target audience and purpose

Section 2: Context Profile

In this section, you should describe the context in which the examined list will be used. Explaining the purpose(s), the learners, and the target type of vocabulary is crucial for determining the usefulness of the list. Analysing the context will help you make the necessary decisions when using the list.

2.1 What is the purpose for using the examined word list?

Describe the purpose(s) for using the word list. In doing so, think about the most important aim to be accomplished by using the list. Table 2 explains the most common purposes. If the intended purpose is not included in the table, please choose 'Other' and then detail what it is.

- a. Course or program design
- b. Classroom teaching and learning
- c. Testing (assessment)
- d. Materials development
- e. Research (related to one of the above purposes)
- f. Other

Table 2 explains the use cases listed above (adapted from Burkett, 2017)

Course design	Setting short term (course) learning goals
	Setting long term (program) learning goals
Teaching and learning	Selecting material that includes the target vocabulary
	A resource for intentional vocabulary learning activities
	Selecting vocabulary from texts to focus on
	Analysing the vocabulary load in a text
Testing	Modifying the level of vocabulary in tests
	Selecting words for vocabulary size and proficiency tests
	Testing list vocabulary in context
	Test results inform teaching and learning
Materials development	Setting the vocabulary level of materials
	Creating vocabulary focused class teaching materials
	Creating vocabulary focused class independent study resources

2.2. What is the target vocabulary type in the context?

- a. English for General Purposes
- b. English for General Academic Purposes
- c. English for Specialised Academic Purposes
- d. English for Professional or Occupational Purposes
- e. Other (please specify)

2.3 What are the characteristics of the learners with whom the list will be used?

Understanding the characteristics of the target learners of the list is crucial for assessing its suitability for the learners and tailoring the list to their needs.

2.3.1. What is the proficiency level of the learners?

- a. Proficient
- b. Independent
- c. Basic
- d. Beginner or False Beginner
- e. Unassessed or Mixed

Table 3 below (adapted from Council of Europe, 2001) describes these levels based on the Common European Framework of Reference (CEFR). In this *Tool*, CEFR is used as a reference to describe the proficiency levels of learners. If you find it difficult to assign one of the levels specified in the *Tool* to your learners, especially if they belong to more than one levels or if CEFR has not been used as a reference in your context, a rough description of the learners is adequate.

Table 3 Descriptions of proficiency level based on CEFR

Level		Description
Proficient	C2	The capacity to deal with material that is academic or cognitively demanding and to use language good fluently, spontaneously and precisely in complex situations.
Advanced	C1	The ability to deal with long and complex unfamiliar topics, to communicate fluently and spontaneously, and to produce clear, well-structured, detailed text on complex subjects.
Upper-Intermediate	B2	The ability to understand the main ideas of concrete and abstract topics and express oneself with a degree of fluency and spontaneity on a range of topics.
Intermediate	B1	The ability to express oneself in a limited manner in familiar situations and to deal in a general manner with non-routine information.
Pre-Intermediate	A2	The ability to understand frequently used language and begin to express oneself in familiar contexts.
Elementary Beginner	A1 A0-A1	The basic ability to understand and use familiar expressions and basic phrases in a simple manner.

2.3.2. What is the academic discipline of the learners?

- a. Hard sciences (e.g., Maths, Engineering, Medicine)
- b. Soft sciences (e.g., Business, Law, History)
- c. Professional and Occupational fields (e.g., nursing, communication)
- d. Mixed
- e. Other (please specify)
- f. Unspecified
- g. Not relevant

2.3.3. Specify other characteristics relevant to the evaluation, such as cultural background and first language.

.....

.....

.....

Section 3: The Evaluation

This is the central section of the tool. Answer each of the following evaluative questions to determine the usefulness or suitability of the element under assessment. Refer to the notes section at the end of this document for guidance and explanations.

3.1. How are words grouped as a unit in the list?

The principle used for grouping words as a unit in a list is crucial for evaluating the list because it determines what learners are expected to learn (Nation, 2016).

Please select the word unit used to group words in the examined list and then answer questions 1–4.

- a. Words are grouped by **Type** (a different form is a different word counted once, e.g., *develop* is counted on its first appearance in a corpus but *develops* is another type counted separately).
- b. Words are grouped by **Lemma** (a stem with inflected forms, e.g., *develop* includes *develops*, *developed*, and *developing*).
- c. Words are grouped by **Flemma** (a stem and inflected forms of different parts of speech, so the forms *developing* (adj.) and *developing* (v) belong to one flemma).
- d. Words are grouped by **Word Family** (a stem with inflected and derived forms, e.g., *develop* includes *develops*, *developed*, *developing*, *undeveloped*, *underdeveloped*, *development*, *developments*, *developer*, and *developers*).

Now, rate the suitability of the word unit by answering questions 1-4 referring to the notes section, click [here](#).

1. Is the word unit (*Type*, *Lemma*, *Flemma*, or *Word Family*) of the word list suitable for the intended purpose? [See Table 4](#)
2. Is the word unit suitable for the target learners? [See Table 4](#)
3. Have any issues related to the definition of the word" unit, such as UK and US spelling, alternative spellings, part of speech, abbreviations, and numbers been dealt with appropriately with respect to the context?
4. If the word unit is deemed inappropriate for the intended purpose or target learners, how should the list be adapted to make it suitable?

3.2. What material (corpus) is the list based on?

To evaluate the corpus, you need information about the types of texts collected, their representativeness, size of the corpus, balance of the corpus, lengths of the texts, and the number of corpora used (Coxhead, 2000).

Do you have access to information about the corpus on which the list is based?

If YES (*you have access to information about the corpus but not necessarily all aspects of the corpus*), answer questions 5-9. Refer to the notes section (b) if needed.

If NO, *move to the section 3.3*. Understanding of the principles underlying the use of a corpus for word list production is essential, refer to the notes.

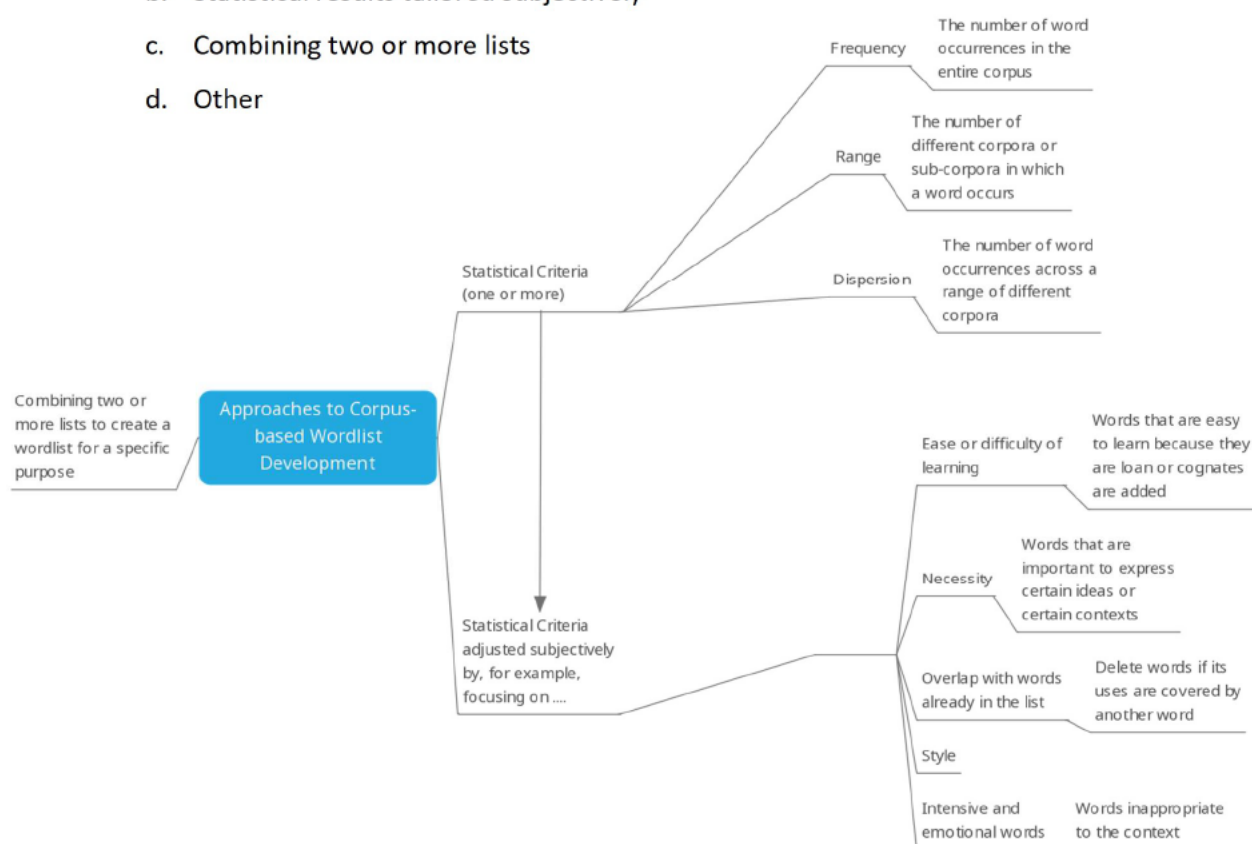
5. Does the *content* of the corpus represent the actual or potential language that the target learners will likely be exposed to?
6. Is the *content* of the corpus suitable for the proficiency level and age of the target learners?
7. Is the *size* of the corpus suitable to obtain reliable results for the intended purpose?
8. Is the *age* of the corpus suitable for the intended purpose?
9. Is the *location* of the corpus suitable for the intended purpose?

3.3. How are words selected from the corpus and ordered into the list?

Select the criteria used to develop the examined list and then answer questions 10–12. Refer to [Table 5](#) for explanations of these criteria. If information about the development of the examined list is not available, and the list is not one of the top word lists in the Excel file (section one), skip these questions.

Describe how the list is made.

- One or more statistical criterion such as frequency, range and dispersion
- Statistical results tailored subjectively
- Combining two or more lists
- Other



The diagram shows the common approaches to corpus-based word lists development.

How appropriate are the criteria used to develop the word list for the intended purpose?

- How suitable are the criteria for inclusion and ordering in the list from the viewpoint of obtaining reliable results?
- If subjective criteria have been used, how well have they been formalized and applied?
- Have the lists been checked against similar lists for coverage and overlapping and non-overlapping words?

3.4. How can the list be implemented in context?

It is really important to have access to information about the list and its development, especially if the list has been developed for a specific purpose.

13. Is the list size manageable for the intended purpose? See note.6
14. Is the context for each word provided (e.g., through concordance lines or example sentences)?
See note.5
15. Are there supplementary materials (e.g., activates, exercises, tests ... etc.) based on the list for use in the class? See note.6
16. How useful is each entry in the list? Are there items that need to be deleted or added?
17. Is there guidance on how to use the list?
18. Are the weaknesses of the list acknowledged and justified? How problematic are they to your context?
19. Are factors such as text type, geographical location, cultural consideration and learning situation considered based on the intended purpose?

This is the end of the *Evaluation Tool*. It is not possible to give a score to the word list based on your evaluation but you should now review your comments on the word list and judge whether the list is suitable to your immediate purpose, whether it may fit other purposes that you may have in the future and how you may adapt it to your teaching context.

Notes

a. Notes on questions 1-4

Questions 1 and 2 address the main factors to be considered when assessing how words are grouped into a word list: (a) the purpose of using the list in terms of whether the focus is productive or receptive knowledge, (b) target vocabulary, and (c) the characteristics of the target learners. Table 3 summarises which word unit is most suitable for each purpose and target learners, as described in the literature.

Table 4 Uses and issues of the common word units

Word unit	Possible purpose and target learners	Drawbacks
Type	Type-based word lists are suitable for (a) productive vocabulary development (i.e., writing or speaking); (b) technical vocabulary learning, because one form of a word may be technical but not the other related forms; and as (c) spelling lists to ensure the correct form is learned (Nation, 2016).	Type-based lists are usually large.
Lemma	Lemma-based lists are suitable for (a) productive vocabulary development; and (b) assessment purposes aimed at beginner to intermediate learners (Kremmel, 2016; McLean, 2018). It is assumed that those learners have not mastered the derivations of English, but they would usually know the inflections of a known stem.	Restrictive as most derivations in English are not difficult.
Flemma	Using this word unit is based on the assumption that learners can make “conceptual links between the use of flemma constituents to express, for example, a nominal entity, a verbal process, or an adjectival attribute” (Stoeckel et al., 2018, p. 1), thus lowering the learning burden (Dang & Webb, 2016b). Learning burden refers to the effort needed to learn (Nation, 2013). Flemma-based lists are more suitable than the word family for low- to intermediate-level proficiency learners.	Not many lists are available.
Word family	Word-family-based lists are suitable for (a) receptive vocabulary development (i.e., listening and reading comprehension) because the meaning of inflected forms (e.g., <i>business-like</i>) can be guessed if the meaning of the base <i>business</i> is known, thus lowering the learning burden (e.g., Schmitt & Zimmerman, 2002) and for (b) developing tests, course design, and graded readers. This grouping assumes that the learners have mastered the morphology of English, commonly acquired later in the learning phase. The L1 background of learners affects affix knowledge (Ward, 2009). Word family is the most common word unit. It has been used widely for word list comparison and development, and in programs such as Range and AntWord Profiler.	Word family-based lists can be short and devoid of word repetition. It may not be pedagogically suitable for most learners

Question 4

Sometimes, the word unit selected for developing a word list is not appropriate for the target purpose or learners. This could be attributed to practical constraints such as the ones listed below:

- Some non-commercial corpus software packages limit themselves to a particular word unit.
- The word list developer might have used a published list based on a unit that is unsuitable for their actual purpose when developing their own list.
- For some purposes, such as comparison between lists, the word unit must be the same across lists even if it is not appropriate.

Recently, some word list developers such as have started publishing alternative versions of their lists with different word units, and teachers can select the version they need. For example, the Academic Vocabulary List (Gardner & Davies, 2014) is available in the word family and lemma format.

b. Notes on questions 5-9:

Content

Although it might be difficult to evaluate a corpus owing to insufficient information about it or lack of knowledge, an understanding of the principles underlying the use of a corpus for word list production is essential.

A corpus can include, for example, textbooks or textbook chapters, journal articles, literary texts, graded readers, teaching materials, audio-recorded conversations or lectures, and movies. A good word list is based on a corpus that represents the actual or potential language that the target audience will likely encounter. However, it is impossible to collect *all* instances of the target language. Consequently, one must rely on a **sample** that closely represents the language used by the target population. To evaluate the suitability of a corpus, the purpose, context and the characteristics of the learners must be considered. For instance, if the target learners are young children, the list should be based on a corpus of texts that those children will likely encounter. The geographical location and cultural background of the source texts used to compile the list must be considered because these will affect what will appear in the list. Words such as *palm* (as in *palm tree*) will be relevant to learners in the Middle East, for example, but they might not appear in texts collected from the UK.

Size

There is little agreement on corpus size, and it depends on the purpose of the resulting list and other practical considerations. There has been a strong preference for large corpora collected to achieve representativeness (e.g., Sinclair, 1991), yet the size of the corpus depends very much on the purpose

of the resulting list and other practical considerations. A small corpus can yield considerable amounts of useful information, especially on high-frequency items that appear frequently in specialised texts (Coxhead, 2018). For example, when developing a list based on the teaching materials for a particular teaching context, a small pedagogic corpus built based on the teaching material would yield a reliable and useful list.

Practical issues that may constrain corpus size include the following:

- Size limits set by some commercial corpus software.
- The fact that spoken corpora tend to be smaller than written corpora because collecting and transcribing spoken texts are time-consuming tasks. According to (O'Keeffe et al., 2007, p. 4), over a million words of speech is considered a large corpus, but fewer than five million words of written text is considered small.
- The words that can be used from published texts may be limited by the permissions provided by the copyright holders.

Age

New or modern words are unlikely to appear in an old corpus. The major criticism of the GSL despite its high lexical coverage is that it was based on old texts from the 1920s and thus does not include modern words such as *email*, *online* etc. (Nation, 2016, p. 124).

c. Notes on questions 10-12: How was the list made?

Each approach to word list development has its advantages and disadvantages, but the criteria used must be evaluated against the purpose and context.

Table 5	
Statistical criteria	The higher the frequency, the wider the range and the distribution of a word, the more likely is for a learner to encounter the word. This rule has been claimed to be more reliable and accurate than those based on intuition (Szudarski, 2018). Statistically based lists are replicable allowing for comparison. However, the development of a large, truly representative corpus is often impossible, and most corpus software tools remain limited (Nation, 2016, p. 119).

Statistical results that have been tailored subjectively	<p>More likely to be pedagogically useful because they consider the needs of the target learners and context. The General Service List (GSL; West, 1953) is the most common example of statistical results that have been adjusted subjectively to make the list as efficient and complete as possible for the target users. Despite its age, the GSL achieves high lexical coverage across different corpora.</p> <p>Drawbacks: (a) there is a subjective element in those lists, and so we might have different lists for the same purpose, (b) it is difficult to replicate the list, though not important in pedagogical lists because situational factors are more likely to differ from one context to another (Nation, 2016, p. 119).</p>
Combining ready-made word lists	<p>The disadvantage of this step is that it incorporates the weaknesses of the supplementary word lists. Therefore, the quality of the resulting list is affected by the quality of the original lists.</p>

Notably, the identification of different types of vocabulary may require different criteria. For example, specialised word lists may require different criteria, such as the judgment of a specialist informant.

It is important that the criteria are formalized and applied consistently.

d. Notes on questions 13-19

Question 14: Fitting a long list of words into a language course is among the most common problems triggered by the use of word lists. It often entails deciding filtering the most of relevant words or asking learners to learn the word list independently.

Question 15: It has been argued that vocabulary can be learned if provided in context. To determine the importance of context, one must consider whether context is important for the intended purpose and whether context would be useful and accessible to the target learners.

If “yes”, state its relevance to the intended purpose, and list the basis for its selection. What meaning was presented?

If “no”, does it really matter? If the intended purpose and context require presenting words in context, is the corpus on which the list is based accessible? Can the teaching materials provide a more appropriate context where the relevant meaning is addressed?

For teaching and learning purposes, the context is more important than the use of a word list, from the viewpoint of materials development and testing. However, the type of context provided in authentic language data, such as corpus examples, might not be suitable for the proficiency level of the target learners or their needs, meaning that it might be more useful if the teacher provides the context.

Question 16: Decisions related to the importance of the supplementary materials depend on the intended purpose and the context, and on whether those materials would be useful in a given case. The availability of materials based on a word list might be useful for replication or preparing materials of a similar type.

Question 17: This corresponds with the respondents' opinion in the survey on word lists generation. Most respondents argued that a teacher should be acquainted with word list development so that they can tailor the lists to their contexts and purposes.

The Evaluation Tool Example

Below is an example to help you use the **Corpus-Based Word list Evaluation Tool**. In this example, an English language teacher at a university foundation program uses it to evaluate the Academic Word List (AWL)—a widely used corpus-based word list—developed by Averil Coxhead (2000). The AWL has been selected by the course coordinators for use alongside the teaching materials; thus, the purpose of this evaluation is not to select a word list but to determine the extent to which the AWL is suitable to the target teaching context and how it can be implemented successfully.

Key of Fonts in the evaluation tool example:

Calibri (body): Evaluation tool

Times new roman (italicized): Teacher's responses

Researcher's commentaries are in yellow boxes.

Section 1: General Attributes of the Word List

This section prompts you to create a brief overview of the main attributes of the examined word list. You can either locate the information in the attached Excel file (adapted from Burkett, 2017) which summarises the most popular word lists of different types of vocabulary published between 2000 and 2018 (in addition to West's 1953 list), or, if the target list is not included, then you should fill out the following table.

Table 1: The examined word list's attributes

Name of the list	Author(s), Year (if published)	Vocabulary type	Size of word list and word unit	Corpus details	Criteria used	Target audience and purpose
<i>Academic Word List (AWL)</i>	<i>Averil Coxhead, 2000</i>	<i>General Academic</i>	<i>570 words families</i>	<i>3,500,000 academic words</i>	<i>Range and frequency (outside of the GSL)</i>	<i>Academic vocabulary for university students in EAP programs</i>

Table 1 shows the teacher filling out information about the examined word list.

Section 2: Context Profile

In this section, you should describe the context in which the examined list will be used. Explaining the purpose(s), the learners, and the target type of vocabulary is crucial for determining the usefulness of the list. Analysing the context will help you make the necessary decisions when using the list.

2.1 What is the purpose for using the examined word list?

Describe the purpose(s) for using the word list. In doing so, think about the most important aim to be accomplished by using the list. Table 2 explains the most common purposes. If the intended purpose is not included in the table, please choose 'Other' and then detail what it is.

- a. Course or program design
- b. Classroom teaching and learning
- c. Testing (assessment)
- d. Materials development
- e. Research (related to one of the above purposes)
- f. Other

Table 2 explains the use cases listed above (adapted from Burkett, 2017)

Course design	Setting short term (course) learning goals
	Setting long term (program) learning goals
Teaching and learning	Selecting material that includes the target vocabulary
	A resource for intentional vocabulary learning activities
	Selecting vocabulary from texts to focus on
	Analysing the vocabulary load in a text
Testing	Modifying the level of vocabulary in tests
	Selecting words for vocabulary size and proficiency tests
	Testing list vocabulary in context
	Test results inform teaching and learning
Materials development	Setting the vocabulary level of materials
	Creating vocabulary focused class teaching materials
	Creating vocabulary focused class independent study resources

Comments (Optional):

As a teacher, I will use the AWL for several purposes, such as classroom teaching. I use the list to make vocabulary-teaching materials and tasks for the class, select a writing topic about which students must write while incorporating words from the list, and test the students' progress. Additionally, I use the list for individualised learning based on students' individual needs. I give each student level-specific words that they must learn, memorise, and put in sentences.

2.3. What is the target vocabulary type in the context?

- a. English for General Purposes
- b. *English for General Academic Purposes*
- c. English for Specialised Academic Purposes
- d. English for Professional or Occupational Purposes
- e. Other (please specify)

2.3 What are the characteristics of the learners with whom the list will be used?

Understanding the characteristics of the target learners of the list is crucial for assessing its suitability for the learners and tailoring the list to their needs.

2.3.1. What is the proficiency level of the learners?

- a. Proficient
- b. Advanced
- c. *Upper-Intermediate*
- d. *Intermediate*
- e. Pre-intermediate
- f. Elementary/ Beginner
- g. Unassessed or Mixed

Table 3 below (adapted from Council of Europe, 2001) describes these levels based on the Common European Framework of Reference (CEFR). In this *Tool*, CEFR is used as a reference to describe the proficiency levels of learners. If you find it difficult to assign one of the levels specified in the *Tool* to your learners, especially if they belong to more than one levels or if CEFR has not been used as a reference in your context, a rough description of the learners is adequate.

Table 3 Descriptions of proficiency level based on CEFR

Level		Description
Proficient	C2	The capacity to deal with material that is academic or cognitively demanding and to use language good fluently, spontaneously and precisely in complex situations.
Advanced	C1	The ability to deal with long and complex unfamiliar topics, to communicate fluently and spontaneously, and to produce clear, well-structured, detailed text on complex subjects.
Upper-Intermediate	B2	The ability to understand the main ideas of concrete and abstract topics and express oneself with a degree of fluency and spontaneity on a range of topics.

Intermediate	B1	The ability to express oneself in a limited manner in familiar situations and to deal in a general manner with non-routine information.
Pre-Intermediate	A2	The ability to understand frequently used language and begin to express oneself in familiar contexts.
Elementary Beginner	A1 A0-A1	The basic ability to understand and use familiar expressions and basic phrases in a simple manner.

2.3.4. What is the academic discipline of the learners?

- a. Hard sciences (e.g., Maths, Engineering, Medicine)
- b. Soft sciences (e.g., Business, Law, History)
- c. Professional and Occupational fields (e.g., nursing, communication)
- d. Mixed Learners from mixed disciplines
- e. Other (please specify)
- f. Unspecified
- g. Not relevant

2.3.5. Specify other characteristics relevant to the evaluation, such as cultural background, age and first language.

The list will be used with first year university intermediate and upper-intermediate learners in the English for Academic Purposes program of a business school. Their first language is Arabic. They constitute a homogeneous L1 cultural and age group, meaning that I must make cultural considerations when considering textbook contents and taught vocabulary. I must also think about the morphology of Arabic relative to that of English.

The responses to section 2 shows that the teacher is attempting to describe how the word list will be used and the teaching context.

Section 3: The Evaluation

This is the central section of the tool. Answer each of the following evaluative questions to determine the usefulness or suitability of the element under assessment. Refer to the notes section at the end of this document for guidance and explanations.

The following pages illustrate how the evaluator (the teacher) answered the 19 evaluative questions about the AWL. Please note that there is no rating scale; these answers are relative, meaning that some questions may not be applicable to your case. In the present example, it seems that the teacher could not answer all the questions.

3.1. How are words grouped as a unit in the list?

The principle used for grouping words as a unit in a list is crucial for evaluating the list because it determines what learners are expected to learn (Nation, 2016).

Please select the word unit used to group words in the examined list and then answer questions 1–4.

- a. Words are grouped by **Type** (a different form is a different word counted once, e.g., *develop* is counted on its first appearance in a corpus but *develops* is another type counted separately).
- b. Words are grouped by **Lemma** (a stem with inflected forms, e.g., *develop* includes *develops*, *developed*, and *developing*).
- c. Words are grouped by **Flemma** (a stem and inflected forms of different parts of speech, so the forms *developing* _(adj.) and *developing* _(v) belong to one flemma).
- d. Words are grouped by **Word Family** (a stem with inflected and derived forms, e.g., *develop* includes *develops*, *developed*, *developing*, *undeveloped*, *underdeveloped*, *development*, *developments*, *developer*, and *developers*).

Now, rate the suitability of the word unit by answering questions 1-4 referring to the notes section, click [here](#).

1. Is the word unit (*Type, Lemma, Flemma, or Word Family*) of the word list suitable for the intended purpose? See Table 4

This unit is not suitable for the current purpose because the goal is productive vocabulary development; students are expected to learn and use the target vocabulary in their speaking and writing. Thus, we must provide suitable forms. However, as a source of academic vocabulary, the list is useful.

2. Is the word unit suitable for the target learners? See Table 4

This unit may not be suitable for the proficiency level of the students as some words of the word family are not relevant to the teaching goal or are difficult for the students.

3. Have any issues related to the definition of the word unit, such as UK and US spelling, alternative spellings, part of speech, abbreviations, and numbers been dealt with appropriately with respect to the context?

The distinction between US and UK spelling always constitutes a problem; usage depends on where the teacher learned English.

4. If the word unit is deemed inappropriate for the intended purpose or target learners, how should the list be adapted to make it suitable?

A solution to the problem of the word family is looking up a dictionary, such as the Oxford Online Dictionary to evaluate the difficulty of new words based on the proficiency level of the word; difficult words are not introduced. We also use the textbook to focus on the important words. We ask the students to use highlighters; the yellow highlighters, for example, could be used for the classroom vocabulary words, which everyone must learn, while the pink highlighters could be used for words on the individualized lists, which are specific to each student. We have a classroom board on which we add new difficult words. Under this board, there are big sheets on which each small group can add their own words.

3.2. What material (corpus) is the list based on?

To evaluate the corpus, you need information about the types of texts collected, their representativeness, size of the corpus, balance of the corpus, lengths of the texts, and the number of corpora used (Coxhead, 2000).

Do you have access to information about the corpus on which the list is based?

If YES (you have access to information about the corpus but not necessarily all aspects of the corpus), answer questions 5-9. Refer to the notes section (b) if needed.

If NO, move to the section 3.3. Understanding of the principles underlying the use of a corpus for word list production is essential, refer to the notes.

5. Does the *content* of the corpus represent the actual or potential language that the target learners will likely be exposed to?
6. Is the *content* of the corpus suitable for the proficiency level and age of the target learners?
5 & 6. The content of the corpus reflects the language university L2 students are likely to be exposed to and it is suitable to their proficiency level and age, but we must still use the textbook to match the target words with the AWL words.
7. Is the *size* of the corpus suitable to obtain reliable results for the intended purpose?
8. Is the *age* of the corpus suitable for the intended purpose?
9. Is the *language variety or geographical location* of the corpus suitable for the intended purpose?

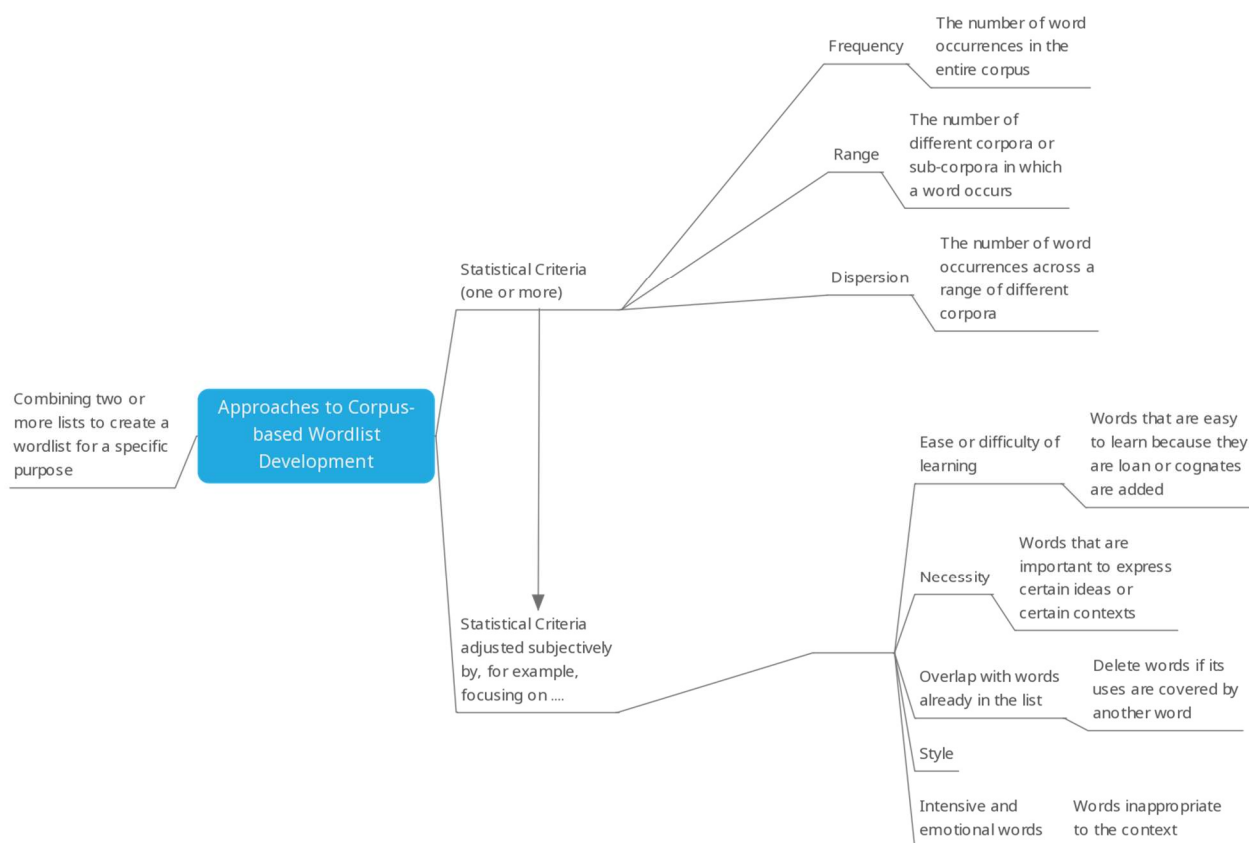
The responses to sections 3.1 and 3.2 suggest that teacher may not be familiar with the concepts of corpus is, though reading the notes was helpful in understanding the source from which words are selected. Access to information about the corpus from which the AWL was derived is accessible online because it is a popular list. The above are the teacher's responses to questions 5 and 6 about the content of the corpus. It seems that the teacher found questions 7 -9 about the size, age and location irrelevant to the teaching context, hence, did not respond to them though information about those elements was accessible.

3.3.How are words selected from the corpus and ordered into the list?

Select the criteria used to develop the examined list and then answer questions 10–12. Refer to [Table 5](#) for explanations of these criteria. If information about the development of the examined list is not available, and the list is not one of the top word lists in the Excel file (section one), skip these questions.

Describe how the list is made.

- a. One or more statistical criterion such as frequency, range and dispersion
- b. Statistical results tailored subjectively
- c. Combining two or more lists
- d. Other



The diagram shows the common approaches to corpus-based word lists development.

3.4. How appropriate are the criteria used to develop the word list for the intended purpose?

10. How suitable are the criteria for inclusion and ordering in the list from the viewpoint of obtaining reliable results?

The AWL was developed by Coxhead using statistical criteria. While I have never thought about the approaches to making word lists, I think that Coxhead's approach has produced a useful list.

11. If subjective criteria have been used, how well have they been formalized and applied?

The AWL was made relying solely on numbers. Nevertheless, we use our subjective judgement of what is useful or difficult to our students to adapt the list to our context in order to make it more relevant and useful for both classroom and learning.

12. Have the lists been checked against similar lists for coverage and overlapping and non-overlapping words? If evidence is provided that the list has been evaluated/validated, list users can likely use the list with greater confidence.

Research shows that the AWL's coverage of academic text is good and that it has been investigated extensively, which confirms the usefulness of the AWL.

3.5. How can the list be implemented in context?

It is really important to have access to information about the list and its development, especially if the list has been developed for a specific purpose.

13. Is the list size manageable for the intended purpose? See note.6

The size of the list may be appropriate for the whole program (two EAP courses and one English for Specialised Purposes course) but not the current EAP course (14 weeks). The coordinators typically manage the list to the course time period.

14. Is the context for each word provided (e.g., through concordance lines or example sentences)? See note.5

No context for the words is provided in the AWL, but we use the teaching textbook to provide example sentences and explain the taught vocabulary.

15. Are there supplementary materials (e.g., activities, exercises, tests ... etc.) based on the list for use in the class? See note.6

We use flashcards, videos and various other tools provided by the course coordinators.

16. How useful is each entry in the list? Are there items that need to be deleted or added?

The words in the AWL are relevant to and useful for our purpose and context despite the problem with the word unit, which can be solved by consulting dictionaries.

17. Is there guidance on how to use the list?

While there is no guidance on using the list, there are teaching goals and a course syllabus to follow.

18. How problematic the weaknesses of the list, if any, to your context?

The acknowledged weaknesses of the AWL are not problematic to current teaching context.

19. Are factors such as text type, geographical location, cultural consideration and learning situation considered based on the intended purpose?

I think the AWL is suitable in terms of text type, geographical location, cultural consideration and the teaching context. As always, we can omit any inappropriate or difficult words, as we do not provide the students with the actual list.

This is the end of the *Evaluation Tool*. It is not possible to give a score to the word list based on your evaluation but you should now review your comments on the word list and judge whether the list is suitable to your immediate purpose, whether it may fit other purposes that you may have in the future and how you may adapt it to your teaching context.

References

- Burkett, T. H. (2017). *An Investigation into the use of word lists in university foundation programs in the United Arab Emirates* [Doctoral dissertation, University of Exeter].
<https://ore.exeter.ac.uk/repository/handle/10871/33151?show=full>
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. University of Cambridge.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
<https://doi.org/10.2307/3587951>
- Coxhead, A. (2018). *Vocabulary and English for Specific Purposes research: Quantitative and qualitative perspectives*. Routledge. <https://doi.org/10.4324/9781315146478>
- Dang, T. N. Y., & Webb, S. (2016a). Evaluating lists of high-frequency words. *ITL-International Journal of Applied Linguistics*, 167(2), 132–158. <https://doi.org/10.1075/itl.167.2.02dan>
- Dang, T. N. Y., & Webb, S. (2016b). Making an essential word list. In I. S. P. Nation (Ed.), *Making and using word lists for language learning and testing* (pp. 153–167). John Benjamins.
- Kremmel, B. (2016). Word families and frequency bands in vocabulary tests: Challenging conventions. *TESOL Quarterly*, 50(4), 976–987. <https://doi.org/10.1002/tesq.329>
- McLean, S. (2018). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics*, 39(6), 823–845. <https://doi.org/10.1093/applin/amw050>
- Nation, I. S. P. (2013). *Learning Vocabulary in Another Language* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139858656>
- Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins
<https://doi.org/10.1075/z.208>
- O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511497650>
- Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly*, 36(2), 145–171. <https://doi.org/10.2307/3588328>
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Stoeckel, T., Ishii, T., & Bennett, P. (2018). Is the Lemma More Appropriate than the Flemma as a Word Counting Unit? *Applied Linguistics*. <https://doi.org/10.1093/applin/amy059>
- Szudarski, P. (2018). *Corpus Linguistics for Vocabulary: A Guide for Research*. Routledge.
<https://doi.org/10.4324/9781315107769>

Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*, 28(3), 170–182.

<https://doi.org/10.1016/j.esp.2009.04.001>

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

APPENDIX I

Critical Feedback Survey (Phase 4)

For each of the following statements about the **Corpus-based Word list Evaluation Tool**, please select the number that best reflects the extent to which you agree or disagree using the rating 1 – 5 where 1 indicates "strongly disagree" and 5 indicates "strongly agree". Circle only **one** rating for each statement. For elaboration, use the further comments.

a. Effectiveness	Strongly disagree		Neutral		Strongly Agree
1. The tool is useful in determining the suitability of a word list for use in a given teaching context.	1	2	3	4	5
2. The tool can help practitioners in adapting word lists to their teaching context.	1	2	3	4	5
3. The tool is applicable to different ELT contexts.	1	2	3	4	5
4. The introductory document is useful for learning about word lists.	1	2	3	4	5

Further comments:

b. Content and evaluation questions	Strongly disagree		Neutral		Strongly Agree
5. The tool takes into account the different aspects that need to be considered when evaluating word lists.	1	2	3	4	5
6. The evaluation questions are comprehensive.	1	2	3	4	5
7. The evaluation questions are answerable	1	2	3	4	5
8. The supporting notes are comprehensive.	1	2	3	4	5
9. The content is non-overlapping.	1	2	3	4	5

Further comments:

c. Efficiency	Strongly disagree		Neutral		Strongly Agree
10. The tool is well-structured.	1	2	3	4	5
11. The content is concise.	1	2	3	4	5
12. The content is clear.	1	2	3	4	5
13. The tool is easy to use.	1	2	3	4	5
14. For practitioners, the tool is easy to understand.	1	2	3	4	5

Further comments:

2. Please answer the following questions regarding the tool:

1. What do you think the key strengths of the tool are, if any?
2. What do you think the key weaknesses of the tool are, if any?
3. Is there any component that should be added, reviewed and/or deleted? If yes, please explain.
4. What changes, if any, do you consider necessary to improve the tool?

Comments:

3. Please answer the following questions regarding your experience:

1. **Which of the following describes your current status?** (more than one option can be selected)
 - Language Teacher
 - Researcher
 - Language Tester
 - Materials developer
 - Course/program coordinator
 - Other (please specify) _____

2. **In which context of English language teaching have you been mainly involved?** (more than one option can be selected)
 - English for General Purposes
 - English for Academic Purposes
 - English for Specialized Purposes
 - English for Professional or Occupational Purposes
 - Other (please specify) _____
 - Further comments _____

3. **In general, which of the following best characterizes the extent of your experience of corpus-based word lists, if any?**
 - None
 - Very little
 - Some
 - A great deal
 - Extensive expert

4. **Taking into account your expertise and familiarity with the relevant literature on word lists, how confident are you about your review and your recommendations, if any?**
 - Absolutely confident
 - Confident
 - Fairly confident
 - Confident in certain in areas
 - Not at all

APPENDIX J

Interview Protocol – Study 2 (Phase 4)

During the interview, the three main points that will be discussed are the effectiveness of the tool, efficiency and comprehensiveness and relevance of the tool's content.

a. Efficiency

1. Can you briefly describe the process you followed while using the tool? Did you follow all the steps in the order they appeared in? Did you read the introductory document ... etc. or did you skip?
2. Did you experience any difficulties when using the tool? In terms of clarity, structure, understanding technical terms?
3. How long did it take you to finish the evaluation using the tool?
4. How helpful was the physical format of the tool?

b. Effectiveness

5. Do you foresee any benefits for implementing word lists in your pedagogical context as a result of using this tool? Why or why not?
6. Which aspects of the tool are most beneficial?
7. To what extent do you think these benefits are likely to be realized?
8. Do you think that the tool will work in different contexts?
9. Which aspects of the tool are most problematic/difficult to understand or irrelevant?

c. Content

10. Do you think the tool covers the right amount of information needed?
11. Were the questions clear and answerable?
12. Is there something else that you would like to add about your experience in using the tool?
13. Are there any further recommendations for improvement of the tool?

APPENDIX K

Coding Scheme (Phase 4)

Theme	Definition	Code
EFFECTIVENESS of the tool and supplementary materials	Positive and negative comments on the effectiveness of the tool, and the supplementary materials in generating an intended result. In this case, tool's effectiveness in evaluating one or more corpus-based wordlists for pedagogical use	Potential benefits of the tool. Uncertainly in the usefulness of the tool Helpfulness of supplementary materials. Change practice after using the tool. Awareness of corpus-based wordlists
CONTENT of the tool and supplementary materials.	Responses related to the completeness and relevance of the content (i.e., the questions, the supplementary notes and the introductory document) in regard to evaluating wordlists for ELT use.	Relevance to teaching contexts and different wordlists Clear content Comprehensive content Useful/important Answerability of the questions Wording/Rephrase/Structure
EFFICIENCY of the tool in practice	Responses concerned with the physical and practical aspects of the tool – its clarity, accessibility, structure, time spent on it and ease of use	Complicated presentation Difficult to use Interface Time Length Format
RECOMMENDATION for improvement	Suggestions to improve the tool including but not exclusively, effectiveness, efficiency and content.	Improve Add Edit Change Paraphrase Issue
USING wordlists in ELT	Responses related to how wordlists are utilised in different ELT contexts and how this may influence the usefulness of the tool and wordlists in general.	Negative attitudes Difficult to use Choosing wordlist Restrictive use

APPENDIX L

The Final Iteration of the Evaluation Tool

Corpus-Based Word list Evaluation Tool for Pedagogical Purposes

Introduction to the *Evaluation Tool*

The *Evaluation Tool* is designed to help you evaluate the suitability of corpus-based word lists to be used for pedagogical purposes. It is based on the observation that a one-size-fits-all word list does not exist. Contextual factors are likely to differ from case to case. Therefore, the tool considers these factors in determining the usefulness of a word list and how it can be adapted for maximum utility.

The tool is developed based on and guided primarily by (a) Nation's (2016) critique Framework, (b) a review of the relevant literature on word list production and use for pedagogical purposes, and (c) a survey of practices and views related to the word list usage in English language teaching (ELT).

The target users of the tool are, broadly speaking, ELT practitioners, such as teachers, curriculum or assessment coordinators, or materials developers involved in directing vocabulary acquisition. They may range from (a) practitioners deeply involved in curriculum development and coordination, often in a management position at an institution and who may or may not be involved in the implementation of word lists, to (b) teachers with varying levels of teaching experience who implement word lists in their teaching and may discuss and help with materials development but have no management role. The tool aims to cater to practitioners with different levels of expertise and knowledge, and especially those who are unfamiliar with the intricacies of developing corpus-based word lists. It summarises the debates in the field to help an evaluator consider the most important elements associated with word list usage.

The tool is designed for the evaluation of word lists containing single-word units, but can be used to reflect on the suitability of multi-word units lists. You are expected to spend approximately 50-60 minutes to respond to the questions in the tool. Before using this Evaluation Tool, it might be helpful to orient yourself with word lists. Read an introductory document **Introduction to Word Lists** OR **The Guide to Corpus-based Word Lists** that was designed to orient practitioners on the basics of word lists.

How the *Evaluation Tool* is organised

The tool is divided into three sections:

Section 1: General Attributes of the Word List

In this section, information about the word list under examination is elicited in order to prepare you for the evaluation.

Section 2: Context Profile

In this section, you should input the purpose of using the examined list, the type of vocabulary targeted in the learning setting, and the learners' characteristics. This information will serve as the basis for evaluating the list and tailoring it for maximal usefulness.

Section 3: Evaluation

This is the core of the tool, and it includes 19 evaluation questions adapted from Nation's (2016) tool and categorised into four groups:

- Word unit
- Corpus
- Criteria for word list development
- Pedagogical considerations

At the end of the document, you will find a set of Notes that provides key information about word lists, guided answers, debates, and arguments. A brief summary of the information is provided for the benefit of practitioners who are unfamiliar with the intricacies of developing corpus-based word lists. An example is also presented next to help you use the *Evaluation Tool*. In this example, an evaluator uses the tool to evaluate an actual word list. A list of references is also provided.

The Evaluation Tool Example

Below is an example to help you use the **Corpus-Based Word list Evaluation Tool**. In this example, an English language teacher at a university foundation program uses it to evaluate the Academic Word List (AWL)—a widely used corpus-based word list—developed by Averil Coxhead (2000). The AWL has been selected by the course coordinators for use alongside the teaching materials; thus, the purpose of this evaluation is not to select a word list but to determine the extent to which the AWL is suitable to the target teaching context and how it can be implemented successfully.

Key of Fonts in the evaluation tool example:

Calibri (body): Evaluation tool

Times new roman (italicized): Teacher's responses

Researcher's commentaries are in yellow boxes.

Section 1: General Attributes of the Word List

This section prompts you to create a brief overview of the main attributes of the examined word list. You can either locate the information in the attached Excel file (adapted from Burkett, 2017) which summarises the most popular word lists of different types of vocabulary published between 2000 and 2018 (in addition to West's 1953 list), or, if the target list is not included, then you should fill out the following table.

Table 1: The examined word list's attributes

Name of the list	Author(s), Year (if published)	Vocabulary type	Size of word list and word unit	Corpus details	Criteria used	Target audience and purpose
<i>Academic Word List (AWL)</i>	<i>Averil Coxhead, 2000</i>	<i>General Academic</i>	<i>570 words families</i>	<i>3,500,000 academic words</i>	<i>Range and frequency (outside of the GSL)</i>	<i>Academic vocabulary for university students in EAP programs</i>

Table 1 shows the teacher filling out information about the examined word list.

Section 2: Context Profile

In this section, you should describe the context in which the examined list will be used. Explaining the purpose(s), the learners, and the target type of vocabulary is crucial for determining the usefulness of the list. Analysing the context will help you make the necessary decisions when using the list.

2.1 What is the purpose for using the examined word list?

Describe the purpose(s) for using the word list. In doing so, think about the most important aim to be accomplished by using the list. Table 2 explains the most common purposes. If the intended purpose is not included in the table, please choose 'Other' and then detail what it is.

- a. Course or program design
- b. Classroom teaching and learning
- c. Testing (assessment)
- d. Materials development
- e. Research (related to one of the above purposes)
- f. Other

Table 2 explains the use cases listed above (adapted from Burkett, 2017)

Course design	Setting short term (course) learning goals
	Setting long term (program) learning goals
Teaching and learning	Selecting material that includes the target vocabulary
	A resource for intentional vocabulary learning activities
	Selecting vocabulary from texts to focus on
	Analysing the vocabulary load in a text
Testing	Modifying the level of vocabulary in tests
	Selecting words for vocabulary size and proficiency tests
	Testing list vocabulary in context
	Test results inform teaching and learning
Materials development	Setting the vocabulary level of materials
	Creating vocabulary focused class teaching materials
	Creating vocabulary focused class independent study resources

Comments (Optional):

As a teacher, I will use the AWL for several purposes, such as classroom teaching. I use the list to make vocabulary-teaching materials and tasks for the class, select a writing topic about which students must write while incorporating words from the list, and test the students' progress. Additionally, I use the list for individualised learning based on students' individual needs. I give each student level-specific words that they must learn, memorise, and put in sentences.

2.4. What is the target vocabulary type in the context?

- a. English for General Purposes
- b. *English for General Academic Purposes*
- c. English for Specialised Academic Purposes
- d. English for Professional or Occupational Purposes
- e. Other (please specify)

2.3 What are the characteristics of the learners with whom the list will be used?

Understanding the characteristics of the target learners of the list is crucial for assessing its suitability for the learners and tailoring the list to their needs.

2.3.1. What is the proficiency level of the learners?

- a. Proficient
- b. Advanced
- c. *Upper-Intermediate*
- d. *Intermediate*
- e. Pre-intermediate
- f. Elementary/ Beginner
- g. Unassessed or Mixed

Table 3 below (adapted from Council of Europe, 2001) describes these levels based on the Common European Framework of Reference (CEFR). In this *Tool*, CEFR is used as a reference to describe the proficiency levels of learners. If you find it difficult to assign one of the levels specified in the *Tool* to your learners, especially if they belong to more than one levels or if CEFR has not been used as a reference in your context, a rough description of the learners is adequate.

Table 3 Descriptions of proficiency level based on CEFR

Level		Description
Proficient	C2	The capacity to deal with material that is academic or cognitively demanding and to use language good fluently, spontaneously and precisely in complex situations.
Advanced	C1	The ability to deal with long and complex unfamiliar topics, to communicate fluently and spontaneously, and to produce clear, well-structured, detailed text on complex subjects.
Upper-Intermediate	B2	The ability to understand the main ideas of concrete and abstract topics and express oneself with a degree of fluency and spontaneity on a range of topics.

Intermediate	B1	The ability to express oneself in a limited manner in familiar situations and to deal in a general manner with non-routine information.
Pre-Intermediate	A2	The ability to understand frequently used language and begin to express oneself in familiar contexts.
Elementary Beginner	A1 A0-A1	The basic ability to understand and use familiar expressions and basic phrases in a simple manner.

2.3.2. What is the academic discipline of the learners?

- a. Hard sciences (e.g., Maths, Engineering, Medicine)
- b. Soft sciences (e.g., Business, Law, History)
- c. Professional and Occupational fields (e.g., nursing, communication)
- d. Mixed Learners from mixed disciplines
- e. Other (please specify)
- f. Unspecified
- g. Not relevant

2.3.3. Specify other characteristics relevant to the evaluation, such as cultural background, age and first language.

The list will be used with first year university intermediate and upper-intermediate learners in the English for Academic Purposes program of a business school. Their first language is Arabic. They constitute a homogeneous L1 cultural and age group, meaning that I must make cultural considerations when considering textbook contents and taught vocabulary. I must also think about the morphology of Arabic relative to that of English.

The responses to section 2 shows that the teacher is attempting to describe how the word list will be used and the teaching context.

Section 3: The Evaluation

This is the central section of the tool. Answer each of the following evaluative questions to determine the usefulness or suitability of the element under assessment. Refer to the notes section at the end of this document for guidance and explanations.

The following pages illustrate how the evaluator (the teacher) answered the 19 evaluative questions about the AWL. Please note that there is no rating scale; these answers are relative, meaning that some questions may not be applicable to your case. In the present example, it seems that the teacher could not answer all the questions.

3.1. How are words grouped as a unit in the list?

The principle used for grouping words as a unit in a list is crucial for evaluating the list because it determines what learners are expected to learn (Nation, 2016).

Please select the word unit used to group words in the examined list and then answer questions 1–4.

- a. Words are grouped by **Type** (a different form is a different word counted once, e.g., *develop* is counted on its first appearance in a corpus but *develops* is another type counted separately).
- b. Words are grouped by **Lemma** (a stem with inflected forms, e.g., *develop* includes *develops*, *developed*, and *developing*).
- c. Words are grouped by **Flemma** (a stem and inflected forms of different parts of speech, so the forms *developing* _(adj.) and *developing* _(v) belong to one flemma).
- d. Words are grouped by **Word Family** (a stem with inflected and derived forms, e.g., *develop* includes *develops*, *developed*, *developing*, *undeveloped*, *underdeveloped*, *development*, *developments*, *developer*, and *developers*).

Now, rate the suitability of the word unit by answering questions 1-4 referring to the notes section, click [here](#).

1. Is the word unit (*Type, Lemma, Flemma, or Word Family*) of the word list suitable for the intended purpose? See Table 4

This unit is not suitable for the current purpose because the goal is productive vocabulary development; students are expected to learn and use the target vocabulary in their speaking and writing. Thus, we must provide suitable forms. However, as a source of academic vocabulary, the list is useful.

2. Is the word unit suitable for the target learners? See Table 4

This unit may not be suitable for the proficiency level of the students as some words of the word family are not relevant to the teaching goal or are difficult for the students.

3. Have any issues related to the definition of the word unit, such as UK and US spelling, alternative spellings, part of speech, abbreviations, and numbers been dealt with appropriately with respect to the context?

The distinction between US and UK spelling always constitutes a problem; usage depends on where the teacher learned English.

4. If the word unit is deemed inappropriate for the intended purpose or target learners, how should the list be adapted to make it suitable?

A solution to the problem of the word family is looking up a dictionary, such as the Oxford Online Dictionary to evaluate, the difficulty of new words based on the proficiency level of the word; difficult words are not introduced. We also use the textbook to focus on the important words. We ask the students to use highlighters; the yellow highlighters, for example, could be used for the classroom vocabulary words, which everyone must learn, while the pink highlighters could be used for words on the individualized lists, which are specific to each student. We have a classroom board on which we add new difficult words. Under this board, there are big sheets on which each small group can add their own words.

3.2. What material (corpus) is the list based on?

To evaluate the corpus, you need information about the types of texts collected, their representativeness, size of the corpus, balance of the corpus, lengths of the texts, and the number of corpora used (Coxhead, 2000).

Do you have access to information about the corpus on which the list is based?

If YES (you have access to information about the corpus but not necessarily all aspects of the corpus), answer questions 5-9. Refer to the notes section (b) if needed.

If NO, move to the section 3.3. Understanding of the principles underlying the use of a corpus for word list production is essential, refer to the notes.

5. Does the *content* of the corpus represent the actual or potential language that the target learners will likely be exposed to?
6. Is the *content* of the corpus suitable for the proficiency level and age of the target learners?
5 & 6. The content of the corpus reflects the language university L2 students are likely to be exposed to and it is suitable to their proficiency level and age, but we must still use the textbook to match the target words with the AWL words.
7. Is the *size* of the corpus suitable to obtain reliable results for the intended purpose?
8. Is the *age* of the corpus suitable for the intended purpose?
9. Is the *language variety or geographical location* of the corpus suitable for the intended purpose?

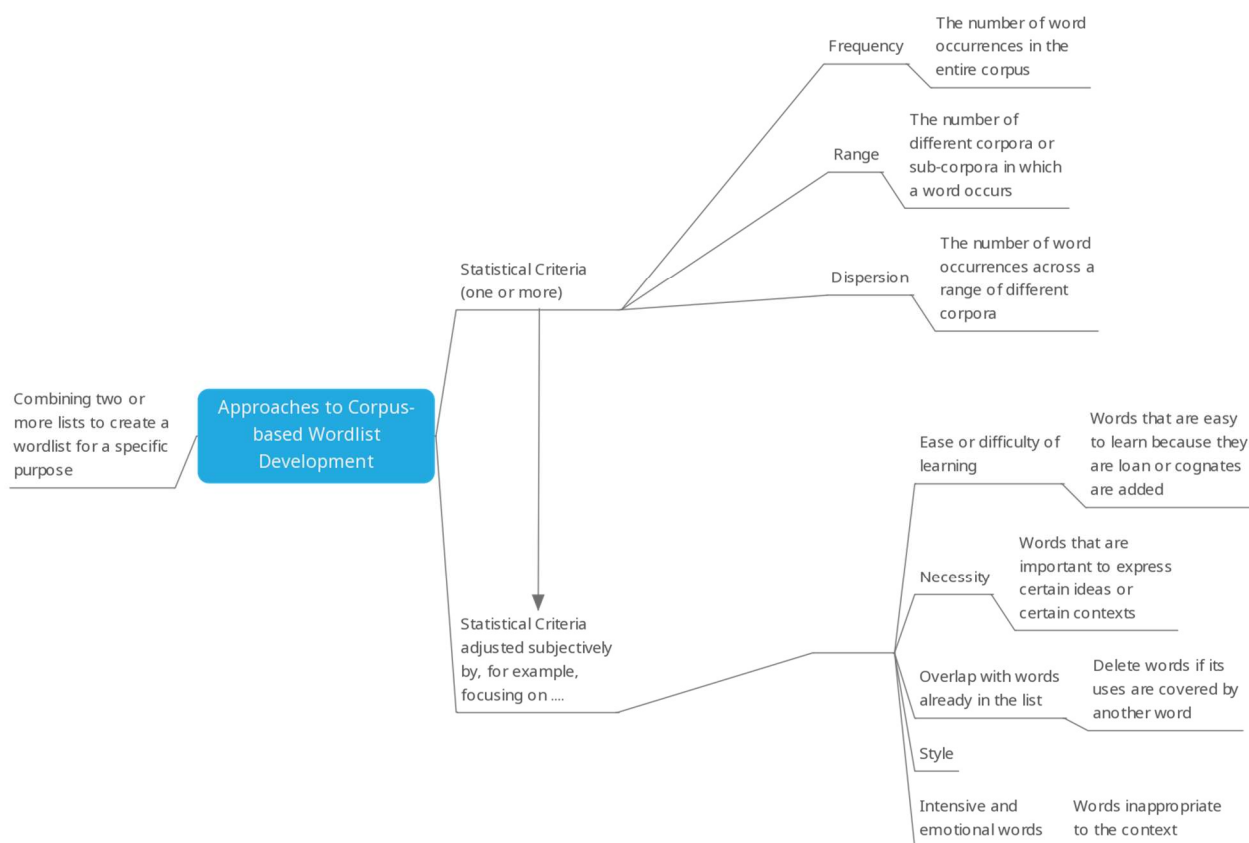
The responses to sections 3.1 and 3.2 suggest that teacher may not be familiar with the concepts of corpus is, though reading the notes was helpful in understanding the source from which words are selected. Access to information about the corpus from which the AWL was derived is accessible online because it is a popular list. The above are the teacher's responses to questions 5 and 6 about the content of the corpus. It seems that the teacher found questions 7 -9 about the size, age and location irrelevant to the teaching context, hence, did not respond to them though information about those elements was accessible.

3.3.How are words selected from the corpus and ordered into the list?

Select the criteria used to develop the examined list and then answer questions 10–12. Refer to [Table 5](#) for explanations of these criteria. If information about the development of the examined list is not available, and the list is not one of the top word lists in the Excel file (section one), skip these questions.

Describe how the list is made.

- a. One or more statistical criterion such as frequency, range and dispersion
- b. Statistical results tailored subjectively
- c. Combining two or more lists
- d. Other



The diagram shows the common approaches to corpus-based word lists development.

3.4. How appropriate are the criteria used to develop the word list for the intended purpose?

10. How suitable are the criteria for inclusion and ordering in the list from the viewpoint of obtaining reliable results?

The AWL was developed by Coxhead using statistical criteria. While I have never thought about the approaches to making word lists, I think that Coxhead's approach has produced a useful list.

11. If subjective criteria have been used, how well have they been formalized and applied?

The AWL was made relying solely on numbers. Nevertheless, we use our subjective judgement of what is useful or difficult to our students to adapt the list to our context in order to make it more relevant and useful for both classroom and learning.

12. Have the lists been checked against similar lists for coverage and overlapping and non-overlapping words? If evidence is provided that the list has been evaluated/validated, list users can likely use the list with greater confidence.

Research shows that the AWL's coverage of academic text is good and that it has been investigated extensively, which confirms the usefulness of the AWL.

3.5. How can the list be implemented in context?

It is really important to have access to information about the list and its development, especially if the list has been developed for a specific purpose.

13. Is the list size manageable for the intended purpose? See note.6

The size of the list may be appropriate for the whole program (two EAP courses and one English for Specialised Purposes course) but not the current EAP course (14 weeks). The coordinators typically manage the list to the course time period.

14. Is the context for each word provided (e.g., through concordance lines or example sentences)? See note.5

No context for the words is provided in the AWL, but we use the teaching textbook to provide example sentences and explain the taught vocabulary.

15. Are there supplementary materials (e.g., activities, exercises, tests ... etc.) based on the list for use in the class? See note.6

We use flashcards, videos and various other tools provided by the course coordinators.

16. How useful is each entry in the list? Are there items that need to be deleted or added?

The words in the AWL are relevant to and useful for our purpose and context despite the problem with the word unit, which can be solved by consulting dictionaries.

17. Is there guidance on how to use the list?

While there is no guidance on using the list, there are teaching goals and a course syllabus to follow.

18. How problematic the weaknesses of the list, if any, to your context?

The acknowledged weaknesses of the AWL are not problematic to current teaching context.

19. Are factors such as text type, geographical location, cultural consideration and learning situation considered based on the intended purpose?

I think the AWL is suitable in terms of text type, geographical location, cultural consideration and the teaching context. As always, we can omit any inappropriate or difficult words, as we do not provide the students with the actual list.

This is the end of the *Evaluation Tool*. It is not possible to give a score to the word list based on your evaluation but you should now review your comments on the word list and judge whether the list is suitable to your immediate purpose, whether it may fit other purposes that you may have in the future and how you may adapt it to your teaching context.

The Evaluation Tool

Section 1: General Attributes of the Word List

This section prompts you to create a brief overview of the main attributes of the examined word list. You can either locate the information in the attached Excel file (adapted from Burkett, 2017) which summarises the most popular word lists of different types of vocabulary published between 2000 and 2018 (in addition to West's 1953 list), or, if the target list is not included, then you should fill out the following table.

Table 1: The examined word list's attributes

Name of the list	Author(s), Year (if published)	Vocabulary type	Size of word list and word unit	Corpus details	Criteria used	Target audience and purpose

Section 2: Context Profile

In this section, you should describe the context in which the examined list will be used. Explaining the purpose(s), the learners, and the target type of vocabulary is crucial for determining the usefulness of the list. Analysing the context will help you make the necessary decisions when using the list.

2.1 What is the purpose for using the examined word list?

Describe the purpose(s) for using the word list. In doing so, think about the most important aim to be accomplished by using the list. Table 2 explains the most common purposes. If the intended purpose is not included in the table, please choose 'Other' and then detail what it is.

- a. Course or program design
- b. Classroom teaching and learning
- c. Testing (assessment)
- d. Materials development
- e. Research (related to one of the above purposes)
- f. Other

Table 2 explains the use cases listed above (adapted from Burkett, 2017)

Course design	Setting short term (course) learning goals
	Setting long term (program) learning goals
Teaching and learning	Selecting material that includes the target vocabulary
	A resource for intentional vocabulary learning activities
	Selecting vocabulary from texts to focus on
	Analysing the vocabulary load in a text
Testing	Modifying the level of vocabulary in tests
	Selecting words for vocabulary size and proficiency tests
	Testing list vocabulary in context
	Test results inform teaching and learning
Materials development	Setting the vocabulary level of materials
	Creating vocabulary focused class teaching materials
	Creating vocabulary focused class independent study resources

2.5. What is the target vocabulary type in the context?

- a. English for General Purposes
- b. English for General Academic Purposes
- c. English for Specialised Academic Purposes
- d. English for Professional or Occupational Purposes
- e. Other (please specify)

2.3 What are the characteristics of the learners with whom the list will be used?

Understanding the characteristics of the target learners of the list is crucial for assessing its suitability for the learners and tailoring the list to their needs.

2.3.1. What is the proficiency level of the learners?

- a. Proficient
- b. Advanced
- c. Upper-Intermediate
- d. Intermediate
- e. Pre-intermediate
- f. Elementary/ Beginner
- g. Unassessed or Mixed

Table 3 below (adapted from Council of Europe, 2001) describes these levels based on the Common European Framework of Reference (CEFR). In this *Tool*, CEFR is used as a reference to describe the proficiency levels of learners. If you find it difficult to assign one of the levels specified in the *Tool* to your learners, especially if they belong to more than one levels or if CEFR has not been used as a reference in your context, a rough description of the learners is adequate.

Table 3 Descriptions of proficiency level based on CEFR

Level		Description
Proficient	C2	The capacity to deal with material that is academic or cognitively demanding and to use language good fluently, spontaneously and precisely in complex situations.
Advanced	C1	The ability to deal with long and complex unfamiliar topics, to communicate fluently and spontaneously, and to produce clear, well-structured, detailed text on complex subjects.
Upper-Intermediate	B2	The ability to understand the main ideas of concrete and abstract topics and express oneself with a degree of fluency and spontaneity on a range of topics.
Intermediate	B1	The ability to express oneself in a limited manner in familiar situations and to deal in a general manner with non-routine information.
Pre-Intermediate	A2	The ability to understand frequently used language and begin to express oneself in familiar contexts.
Elementary Beginner	A1 A0-A1	The basic ability to understand and use familiar expressions and basic phrases in a simple manner.

2.3.2. What is the academic discipline of the learners?

- a. Hard sciences (e.g., Maths, Engineering, Medicine)
- a. Soft sciences (e.g., Business, Law, History)
- b. Professional and Occupational fields (e.g., nursing, communication)
- c. Mixed Learners from mixed disciplines
- d. Other (please specify)
- e. Unspecified
- f. Not relevant

2.3.3. Specify other characteristics relevant to the evaluation, such as cultural background, age and first language.

.....

.....

.....

Section 3: The Evaluation

This is the central section of the tool. Answer each of the following evaluative questions to determine the usefulness or suitability of the element under assessment. Refer to the notes section at the end of this document for guidance and explanations.

3.1. How are words grouped as a unit in the list?

The principle used for grouping words as a unit in a list is crucial for evaluating the list because it determines what learners are expected to learn (Nation, 2016).

Please select the word unit used to group words in the examined list and then answer questions 1–4.

- a. Words are grouped by **Type** (a different form is a different word counted once, e.g., *develop* is counted on its first appearance in a corpus but *develops* is another type counted separately).
- b. Words are grouped by **Lemma** (a stem with inflected forms, e.g., *develop* includes *develops*, *developed*, and *developing*).
- c. Words are grouped by **Flemma** (a stem and inflected forms of different parts of speech, so the forms *developing* _(adj.) and *developing* _(v) belong to one flemma).
- d. Words are grouped by **Word Family** (a stem with inflected and derived forms, e.g., *develop* includes *develops*, *developed*, *developing*, *undeveloped*, *underdeveloped*, *development*, *developments*, *developer*, and *developers*).

Now, rate the suitability of the word unit by answering questions 1-4 referring to the notes section, click [here](#).

1. Is the word unit (*Type*, *Lemma*, *Flemma*, or *Word Family*) of the word list suitable for the intended purpose? [See Table 4](#)
2. Is the word unit suitable for the target learners? [See Table 4](#)
3. Have any issues related to the definition of the word unit, such as UK and US spelling, alternative spellings, part of speech, abbreviations, and numbers been dealt with appropriately with respect to the context?
4. If the word unit is deemed inappropriate for the intended purpose or target learners, how should the list be adapted to make it suitable?

3.2. What material (corpus) is the list based on?

To evaluate the corpus, you need information about the types of texts collected, their representativeness, size of the corpus, balance of the corpus, lengths of the texts, and the number of corpora used (Coxhead, 2000).

Do you have access to information about the corpus on which the list is based?

If YES (*you have access to information about the corpus but not necessarily all aspects of the corpus*), answer questions 5-9. Refer to the notes section (b) if needed.

If NO, *move to the section 3.3. Understanding of the principles underlying the use of a corpus for word list production is essential*, refer to the notes.

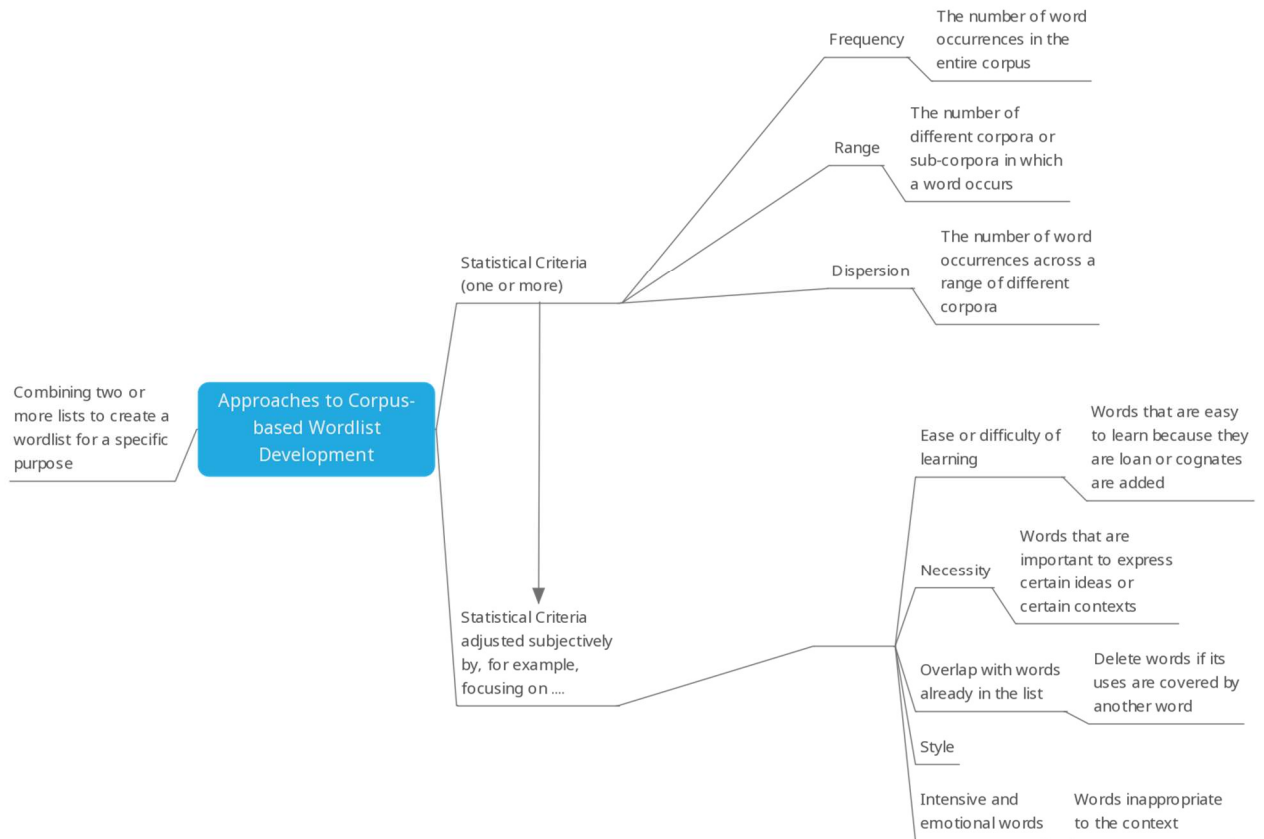
5. Does the *content* of the corpus represent the actual or potential language that the target learners will likely be exposed to?
6. Is the *content* of the corpus suitable for the proficiency level and age of the target learners?
7. Is the *size* of the corpus suitable to obtain reliable results for the intended purpose?
8. Is the *age* of the corpus suitable for the intended purpose?
9. Is the *language variety or geographical location* of the corpus suitable for the intended purpose?

3.3. How are words selected from the corpus and ordered into the list?

Select the criteria used to develop the examined list and then answer questions 10–12. Refer to Table 5 for explanations of these criteria. If information about the development of the examined list is not available, and the list is not one of the top word lists in the Excel file (section one), skip these questions.

Describe how the list is made.

- a. One or more statistical criterion such as frequency, range and dispersion
- b. Statistical results tailored subjectively
- c. Combining two or more lists
- d. Other



The diagram shows the common approaches to corpus-based word lists development.

3.5. How appropriate are the criteria used to develop the word list for the intended purpose?

10. How suitable are the criteria for inclusion and ordering in the list from the viewpoint of obtaining reliable results?
11. If subjective criteria have been used, how well have they been formalized and applied?
12. Have the lists been checked against similar lists for coverage and overlapping and non-overlapping words? If evidence is provided that the list has been evaluated/validated, list users can likely use the list with greater confidence.

3.4. How can the list be implemented in context?

It is really important to have access to information about the list and its development, especially if the list has been developed for a specific purpose.

13. Is the list size manageable for the intended purpose? See note.6
14. Is the context for each word provided (e.g., through concordance lines or example sentences)?
See note.5
15. Are there supplementary materials (e.g., activities, exercises, tests ... etc.) based on the list for use in the class? See note.6
16. How useful is each entry in the list? Are there items that need to be deleted or added?
17. Is there guidance on how to use the list?
18. How problematic the weaknesses of the list, if any, to your context?
19. Are factors such as text type, geographical location, cultural consideration and learning situation considered based on the intended purpose?

This is the end of the *Evaluation Tool*. It is not possible to give a score to the word list based on your evaluation but you should now review your comments on the word list and judge whether the list is suitable to your immediate purpose, whether it may fit other purposes that you may have in the future and how you may adapt it to your teaching context.

Notes

Notes on questions 1-4

Questions 1 and 2 address the main factors to be considered when assessing how words are grouped into a word list: (a) the purpose of using the list in terms of whether the focus is productive or receptive knowledge, (b) target vocabulary, and (c) the characteristics of the target learners. Table 4 summarises which word unit is most suitable for each purpose and target learners, as described in the literature. [Back to questions.](#)

Table 4 Uses and issues of the common word units

Word unit	Possible purpose and target learners	Drawbacks
Type	Type-based word lists are suitable for (a) productive vocabulary development (i.e., writing or speaking); (b) technical vocabulary learning, because one form of a word may be technical but not the other related forms; and as (c) spelling lists to ensure the correct form is learned (Nation, 2016).	Type-based lists are usually large.
Lemma	Lemma-based lists are suitable for (a) productive vocabulary development; and (b) assessment purposes aimed at beginner to intermediate learners (Kremmel, 2016; McLean, 2018). It is assumed that those learners have not mastered the derivations of English, but they would usually know the inflections of a known stem.	Restrictive as most derivations in English are not difficult.
Flemma	Using this word unit is based on the assumption that learners can make “conceptual links between the use of flemma constituents to express, for example, a nominal entity, a verbal process, or an adjectival attribute” (Stoeckel et al., 2018, p. 1), thus lowering the learning burden (Dang & Webb, 2016b). Learning burden refers to the effort needed to learn (Nation, 2013). Flemma-based lists are more suitable than the word family for low- to intermediate-level proficiency learners.	Not many lists are available.
Word family	Word-family-based lists are suitable for (a) receptive vocabulary development (i.e., listening and reading comprehension) because the meaning of inflected forms (e.g., <i>business-like</i>) can be guessed if the meaning of the base <i>business</i> is known, thus lowering the learning burden (e.g., Schmitt & Zimmerman, 2002) and for (b) developing tests, course design, and graded readers. This grouping assumes that the learners have mastered the morphology of English, commonly acquired later in the learning phase. The L1 background of learners affects affix knowledge (Ward, 2009). Word family is the most common word unit. It has been used widely for word list comparison and development, and in programs such as Range and AntWord Profiler.	Word family-based lists can be short and devoid of word repetition. It may not be pedagogically suitable for most learners

Question 4

Sometimes, the word unit selected for developing a word list is not appropriate for the target purpose or learners. This could be attributed to practical constraints such as the ones listed below:

- Some non-commercial corpus software packages limit themselves to a particular word unit.
- The word list developer might have used a published list based on a unit that is unsuitable for their actual purpose when developing their own list.
- For some purposes, such as comparison between lists, the word unit must be the same across lists even if it is not appropriate.

Recently, some word list developers have started publishing alternative versions of their lists with different word units, and teachers can select the version they need. For example, the Academic Vocabulary List (Gardner & Davies, 2014) is available in the word family and lemma format.

[Back to question 4.](#)

Notes on questions 5-9:

Content

Although it might be difficult to evaluate a corpus owing to insufficient information about it or lack of knowledge, an understanding of the principles underlying the use of a corpus for word list production is essential.

A corpus can include, for example, textbooks or textbook chapters, journal articles, literary texts, graded readers, teaching materials, audio-recorded conversations or lectures, and movies. A good word list is based on a corpus that represents the actual or potential language that the target audience will likely encounter. However, it is impossible to collect *all* instances of the target language. Consequently, one must rely on a **sample** that closely represents the language used by the target population. To evaluate the suitability of a corpus, the purpose, context and the characteristics of the learners must be considered. For instance, if the target learners are young children, the list should be based on a corpus of texts that those children will likely encounter. The geographical location and cultural background of the source texts used to compile the list must be considered because these will affect what will appear in the list. Words such as *palm* (as in *palm tree*) will be relevant to learners in the Middle East, for example, but they might not appear in texts collected from the UK. [Back to questions 5 and 6.](#)

Size

There is little agreement on corpus size, and it depends on the purpose of the resulting list and other practical considerations. There has been a strong preference for large corpora collected to achieve

representativeness (e.g., Sinclair, 1991), yet the size of the corpus depends very much on the purpose of the resulting list and other practical considerations. A small corpus can yield considerable amounts of useful information, especially on high-frequency items that appear frequently in specialised texts (Coxhead, 2018). For example, when developing a list based on the teaching materials for a particular teaching context, a small pedagogic corpus built based on the teaching material would yield a reliable and useful list.

Practical issues that may constrain corpus size include the following:

- Size limits set by some commercial corpus software.
- The fact that spoken corpora tend to be smaller than written corpora because collecting and transcribing spoken texts are time-consuming tasks. According to (O'Keeffe et al., 2007, p. 4), over a million words of speech is considered a large corpus, but fewer than five million words of written text is considered small.
- The words that can be used from published texts may be limited by the permissions provided by the copyright holders. [Back to question 7.](#)

Age

New or modern words are unlikely to appear in an old corpus. The major criticism of the GSL despite its high lexical coverage is that it was based on old texts from the 1920s and thus does not include modern words such as *email*, *online* etc. (Nation, 2016, p. 124). [Back to question 8.](#)

Notes on questions 10-12: How was the list made?

Each approach to word list development has its advantages and disadvantages, but the criteria used must be evaluated against the purpose and context.

Table 5

Statistical criteria	The higher the frequency, the wider the range and the distribution of a word, the more likely is for a learner to encounter the word. This rule has been claimed to be more reliable and accurate than those based on intuition (Szudarski, 2018). Statistically based lists are replicable allowing for comparison. However, the development of a large, truly representative corpus is often impossible, and most corpus software tools remain limited (Nation, 2016, p. 119).
-----------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Statistical results that have been tailored subjectively	<p>More likely to be pedagogically useful because they consider the needs of the target learners and context. The General Service List (GSL; West, 1953) is the most common example of statistical results that have been adjusted subjectively to make the list as efficient and complete as possible for the target users. Despite its age, the GSL achieves high lexical coverage across different corpora.</p> <p>Drawbacks: (a) there is a subjective element in those lists, and so we might have different lists for the same purpose, (b) it is difficult to replicate the list, though not important in pedagogical lists because situational factors are more likely to differ from one context to another (Nation, 2016, p. 119).</p>
Combining ready-made word lists	<p>The disadvantage of this step is that it incorporates the weaknesses of the supplementary word lists. Therefore, the quality of the resulting list is affected by the quality of the original lists.</p>

Notably, the identification of different types of vocabulary may require different criteria. For example, specialised word lists may require different criteria, such as the judgment of a specialist informant.

It is important that the criteria are formalized and applied consistently. [Back to questions.](#)

Notes on questions 13-19

Question 14: Fitting a long list of words into a language course is among the most common problems triggered by the use of word lists. It often entails filtering the most of relevant words or asking learners to learn the word list independently.

Question 15: It has been argued that vocabulary can be learned if provided in context. To determine the importance of context, one must consider whether context is important for the intended purpose and whether context would be useful and accessible to the target learners.

If “yes”, state its relevance to the intended purpose, and list the basis for its selection. What meaning was presented?

If “no”, does it really matter? If the intended purpose and context require presenting words in context, is the corpus on which the list is based accessible? Can the teaching materials provide a more appropriate context where the relevant meaning is addressed?

For teaching and learning purposes, the context is more important than the use of a word list, from the viewpoint of materials development and testing. However, the type of context provided in authentic language data, such as corpus examples, might not be suitable for the proficiency level of the target learners or their needs, meaning that it might be more useful if the teacher provides the context.

Question 16: Decisions related to the importance of the supplementary materials depend on the intended purpose and the context, and on whether those materials would be useful in a given case. The availability of materials based on a word list might be useful for replication or preparing materials of a similar type.

Question 17: This corresponds with the respondents' opinion in the survey on word lists generation. Most respondents argued that a teacher should be acquainted with word list development so that they can tailor the lists to their contexts and purposes.

[Back to questions.](#)

References

- Burkett, T. H. (2017). *An Investigation into the use of word lists in university foundation programs in the United Arab Emirates* [Doctoral dissertation, University of Exeter].
<https://ore.exeter.ac.uk/repository/handle/10871/33151?show=full>
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. University of Cambridge.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
<https://doi.org/10.2307/3587951>
- Coxhead, A. (2018). *Vocabulary and English for Specific Purposes research: Quantitative and qualitative perspectives*. Routledge. <https://doi.org/10.4324/9781315146478>
- Dang, T. N. Y., & Webb, S. (2016a). Evaluating lists of high-frequency words. *ITL-International Journal of Applied Linguistics*, 167(2), 132–158. <https://doi.org/10.1075/itl.167.2.02dan>
- Dang, T. N. Y., & Webb, S. (2016b). Making an essential word list. In I. S. P. Nation (Ed.), *Making and using word lists for language learning and testing* (pp. 153–167). John Benjamins.
- Kremmel, B. (2016). Word families and frequency bands in vocabulary tests: Challenging conventions. *TESOL Quarterly*, 50(4), 976–987. <https://doi.org/10.1002/tesq.329>
- McLean, S. (2018). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics*, 39(6), 823–845. <https://doi.org/10.1093/applin/amw050>
- Nation, I. S. P. (2013). *Learning Vocabulary in Another Language* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139858656>
- Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins
<https://doi.org/10.1075/z.208>
- O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511497650>
- Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly*, 36(2), 145–171. <https://doi.org/10.2307/3588328>
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Stoeckel, T., Ishii, T., & Bennett, P. (2018). Is the Lemma More Appropriate than the Flemma as a Word Counting Unit? *Applied Linguistics*. <https://doi.org/10.1093/applin/amy059>
- Szudarski, P. (2018). *Corpus Linguistics for Vocabulary: A Guide for Research*. Routledge.
<https://doi.org/10.4324/9781315107769>

Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*, 28(3), 170–182.

<https://doi.org/10.1016/j.esp.2009.04.001>

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

APPENDIX M

A Guide for Corpus-based Word Lists in the Language Classroom



A GUIDE FOR CORPUS-BASED WORDLISTS IN THE ENGLISH-LANGUAGE CLASSROOM

What is a wordlist? What is a wordlist used for? What is a good wordlist?



A wordlist is a sorted list of single words or multiword units grouped by their frequency counts, sometimes combined with other measures, in a given corpus*. Alternatively, a wordlist may be called a vocabulary list or a frequency list.



THIS GUIDE INCLUDES:

- An introduction to the basics of corpus-based wordlists.
- A set of questions to analyse your wordlist and context (your purpose of using the list, the target vocabulary and the characteristics of the target learners).
- A set of questions related to using wordlists to inform your judgement and future lesson planning.

Wordlists are a useful resource for guiding vocabulary development towards the most important words. Different wordlists have been compiled in different ways and for different purposes. To decide whether to use a wordlist, which wordlist to use and how to use it efficiently, you need to know the basics of wordlists and how you can adapt them for your context and purpose. This is because contextual factors such as learners' proficiency level and target vocabulary are likely to differ from case to case, and a one-size-fits-all wordlist does not exist.

Who is this Guide for?

If you are a language teacher and you use (or plan to use) wordlists for language teaching, this Guide will introduce you to the basics of corpus-based wordlists and how you can use them in your teaching context.

The target users include English language teaching (ELT) practitioners who are either involved or interested in directing vocabulary acquisition and who are not familiar with corpus-based wordlists.

If you already have experience with corpus-based wordlists, you may want to use the *Corpus-based Evaluation Tool*. Contact the researcher if you are interested.

The main **purpose** of making wordlists is to select vocabulary for teaching based on the assumption that the more frequent a word is, the more useful it is. Researchers claim that focus on those words will enable learners to understand much of what they encounter and reproduce what they know.

- There are different types of wordlists based on the vocabulary they present:

General vocabulary or high-frequency wordlists, such as West's (1953) General Service List (GSL), present the basic vocabulary that language learners should learn first.

General academic wordlists, such as Coxhead's (2000) Academic Word List (AWL), focus on academic vocabulary that exists across a wide range of academic disciplines.

Specialised, subject-specific or technical wordlists, such as Lei and Liu's list of specialised terms (2016) in medicine, focus on words for a particular subject area.

How Are Wordlists Made?

List compilers typically make wordlists by choosing or building a corpus of appropriate texts and then selecting words from that corpus by employing specified criteria. The common criteria for selecting and ordering words include one or more of the following:

Frequency: the number of occurrences of a word in the entire corpus, e.g., a word is included in the AWL if it occurs 100 times in the corpus.

Range: the number of different texts or sub-corpora in which a word occurs, e.g., a word is included in the AWL if it occurs at least 10 times in each sub-corpus.

Dispersion: a number calculated by looking at the frequency of a word in each of the sub-corpora or text, i.e., how evenly a word occurs across different texts or sub-corpora.

Statistical results are often **adjusted** by deleting or adding words for example to make the list more suitable and useful for a particular group of learners. In some cases, practitioners combine ready-made wordlists to create a wordlist for their own purposes.



”

*A **corpus** (plural: **corpora**) is a large collection of texts stored in a computer for analysis. A text can include textbooks or textbook chapters, journal articles, literary texts, graded readers, teaching materials, audio-recorded conversations or lectures, and movies.

“

How Are Words Organised into a Wordlist?

Wordlists are lists of words, but **what counts as words?**

Words are commonly known as sequences of letters separated by space or punctuation. This simple criterion, however, would lead to a large number of words in the list. Wordlist compilers typically organise words that share a classical root or additional morphemes as members of a one-word unit. Words are:

- Grouped by **Type** (a different form is a different word counted once. e.g. *develop* is counted on its first appearance in a corpus, but *develops* is another type that is counted separately).
- Grouped by **Lemma** (a root with inflected forms; e.g., *develop* includes *develops*, *developed* and *developing*).
- Grouped by **Flemma** (a root and inflected forms of different parts of speech; e.g. the forms *developing* (adj.) and *developing* (v) belong to one flemma).
- Grouped by **Word Family** (a root with inflected and derived forms; e.g., *develop* includes *develops*, *developed*, *developing*, *undeveloped*, *underdeveloped*, *development*, *developments*, *developer*, and *developers*).

What Are Some Common Applications of Corpus-Based Wordlists?



Course Design: Setting short and long term learning goals. What words and how many words will the students learn at each learning stage?



Classroom Teaching and Learning: A resource for structured vocabulary instruction and independent learning, and selecting materials that include the target vocabulary.



Materials Development: A resource for creating vocabulary-focused materials for classroom and independent learning, and assessing vocabulary level in teaching materials.



Testing: A resource for developing vocabulary size and learning tests.



Wordlists are NOT intended for learners' direct use.



What Is a Good Wordlist?

The best wordlist is one that is suitable for the purpose and context of use. Thus, you as a teacher need to analyse your context, think of your teaching goals, and then plan how you can use a wordlist effectively. This guide walks you through this process.



1/ What is the wordlist you are planning to use?

You need to layout the wordlist's information, and then see how relevant it is for your purpose and context. Click on the link for a table of the most popular wordlists with related information, [Click here](#).

2/ What are you going to use the wordlist for and in what context?

Decide on the **purpose(s) of using** the wordlist or what you expect from the list and describe both the **learners** and the target **type of vocabulary**. Analysis of the context will help you make the necessary decisions when using the list.

What will you use the wordlist for?

- Course design
- Classroom teaching and learning
- Testing
- Materials development

What is the target vocabulary type you are teaching?

- English for general purposes
- English for general academic purposes
- English for specialised academic purposes
- English for professional or occupational purposes
- Other (please specify): -----

3/ Who are the target learners with whom the list will be used?

Understanding the characteristics of the target learners of the list is crucial for assessing its suitability for the learners and tailoring the list to their needs.

What is the proficiency level of the learners?

- Proficient
- Advanced
- Upper-Intermediate
- Intermediate
- Pre-Intermediate
- Elementary
- Beginner or False Beginner
- Unassessed or Mixed

Refer to the Common European Framework for description of each level.

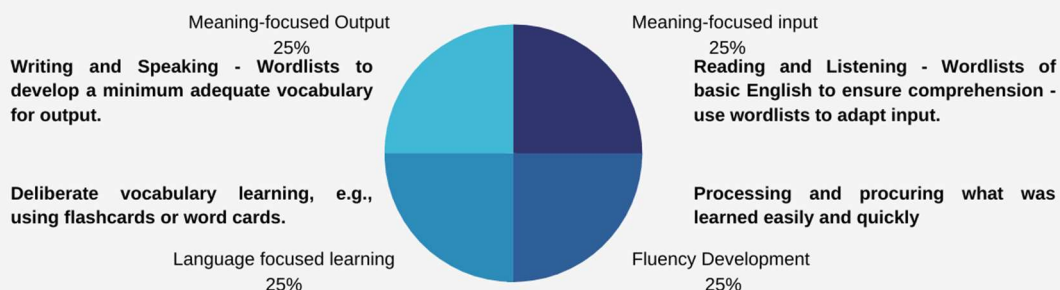
What is the academic discipline of the learners?

- Hard sciences (e.g., Maths, Engineering)
- Soft sciences (e.g., Business, Law, History)
- Professional and occupational fields (e.g., nursing, communication)
- Mixed
- Unspecified
- Not relevant

Specify other characteristics relevant to the evaluation, such as cultural background and first language.

How to implement the list for use?

Vocabulary instruction need to provide balanced opportunities of the four strands of learning from input, learning from output, deliberate learning, and fluency development (Nation, 2016). The chart below shows how wordlists can inform each of the strands.



Now, consider the following points, think about how suitable the list is to your purpose and context, and then plan the next step.



The size of the wordlist.

How many words are there, and can you fit the list into the program?

The inclusion of a long list of words into a language course is among the most common problems that the use of wordlists triggers. Some wordlists are divided into sublists that can be used separately. If the whole list cannot be used:

- Focus on the most relevant or difficult words.
- Ask learners to learn specified words independently.



Supplementary materials.

Are there supplementary materials (e.g. activities, exercises, and tests) based on the list for use in class? To what extent these materials would be useful for you? How good are they?



Contextualised or decontextualised vocabulary learning.

Is the context for each word provided (e.g. through concordance lines or example sentences), and how good is the context?

Most wordlists do not provide context for the words. However, textbook developers often present wordlists in the context of the teaching materials. If you are using a raw wordlist, then you need to determine the importance of the context, how relevant is the context to the intended purpose, and what is the basis for this selection? What meaning is presented? If 'no', does it really matter?



Remove and add words. Review the list and trust your intuition about what is useful to your students.

Consider factors such as text type, geographical location, cultural consideration, and learning situation based on the intended purpose, and decide what needs to be added or deleted.

This guide is an offshoot of an evaluation tool designed for practitioners to evaluate corpus-based wordlists.

For more information about the development and use of corpus-based wordlists, refer to:

- Dang, T. N. Y. (2020). Corpus-based word lists in second language vocabulary research, learning, and teaching. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp.288-304). Routledge.
- Folse, K. S., & Youngblood, A. (2017). Survey of corpus-based vocabulary lists for TESOL classes. *MexTESOL Journal*, 41(1), 1-15.
- Lessard-Clouston, M. (2013). Word Lists for Vocabulary Learning and Teaching. *CATESOL Journal*, 24(1), 287-304.
- Nation, I. S. P. (2016). Making and using word lists for language learning and testing. John Benjamins.
- The [Vocabulary section of the EAP](#) is a useful resource.