Constructing a professional English word list for Field Service Engineers

A corpus-based study

Name: Allysha J. Humphrey

Student number: 4532473

Date: 4 October 2020

Primary supervisor: Dr. S. van Vuuren

Secondary supervisor: Dr. J.K.M. Berns

MA Thesis Linguistics (Language and Communication Coaching

ACKNOWLEDGEMENTS

Provided that I could not have written my master’s thesis without the support and assistance of several people, I would like to extend several words of thanks.

First of all, I would like to express my sincerest gratitude to my supervisor, Sanne van Vuuren, whose expertise and feedback have been invaluable for this thesis. It was her course in the master’s programme that not only inspired me to write this thesis but also to choose the Language and Communication Coaching specialisation in general. Writing a thesis amidst a global pandemic has not been effortless, but Sanne’s flexibility, enthusiasm, and genuine concern were greatly appreciated.

I also wish to thank Marel Boxmeer for welcoming me into their organisation and allowing me to conduct on-site research to critically evaluate their language and communication habits and policies. I hope that my research will provide fruitful and practical directions for a language and communication course in the Marel Academy.

My special thanks are extended to my other half, Ties, for helping me become familiar with Python and Swift coding. This confirmed once more that linguistics comes in all shapes and sizes, also in the form of coding languages with syntax and everything that goes with it. I am, however, positive about this: it is not for me.

I would also like to acknowledge my direct family and closest friends. You were always there for me to listen to my incessant blathering about all things related to my life at university. I know you have always regarded my work and study environment as an ivory tower, but I hope this thesis will give an inkling of my life in academia.

Finally, I am particularly grateful to my father, Edward Humphrey, for inspiring me to conduct such concrete and practical research. You are one of the most hard-working and genuine people I will ever know. One day, I hope I am just as tireless and brave as you. I hope I have made you proud.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AWL:	Academic Word List (Coxhead, 2000)
BEL:	Basic Engineering List (Ward, 2009)
EAP:	English for Academic Purposes
EFL:	English as a Foreign Language
EOP:	English for Occupational Purposes
ESP:	English for Specific Purposes
FSE(s):	Field Service Engineer(s)
GSL:	General Service List (West, 1953)
NA:	Needs Analysis
PEEC:	Professional Engineering English Corpus
PEWL:	Professional Engineering Word List
SEWL:	Student Engineering Word List (Mudraya, 2006)

ABSTRACT

This study reports on the construction of a word list which was developed out of a self-compiled corpus of professional Engineering English. This word list is intended to serve as the basis of a lexical syllabus for an English for Specific Purposes (ESP) course for Field Service Engineers (FSEs), who have expressed the need for such a word list, in their training period at Marel Boxmeer. As such, this study takes up the call for more research into local and discipline-specific language teaching and presents a concrete resource, the Professional Engineering Word List (PEWL), for the target group at focus. This study expounds the compilation of the Professional Engineering English Corpus (PEEC), a corpus of more than 70,000 tokens which is comprised of two sources: customer reports and user manuals. It was analysed quantitatively for word frequency and keyness to develop the PEWL. The PEWL was then compared to other general service word lists, such as West's (1953) GSL and Coxhead's (2000) AWL, as well as other Engineering English word lists, including Mudraya's (2006) and Ward's (2009) word lists. A qualitative analysis was conducted to assess the nature of vocabulary items in the word list, which reveals that most items are highly technical in nature. The PEWL is established as a stand-alone resource in a narrow-angled Engineering English course for the target group in question, given that other word lists do not reach high coverage over the PEWL.

Keywords: Engineering English, ESP, EOP, corpus linguistics, word list, vocabulary

CHAPTER 1: INTRODUCTION

1. Introduction

In an era of globalisation, various companies are reconsidering their language policies or already find themselves in a bilingual working environment, with English being the foreign or second language in most of these companies. This phenomenon has given rise to an increasing need for training staff to be more proficient in English, specifically within their given area of occupation (Martinez, Beck & Panza, 2009). English for Specific Purposes (henceforth ESP) caters for this need, ensuring that learners are presented with precisely that type of language use that they will need in their studies or working environment (Paltridge & Starfield, 2013). To recognise this need for specialised language courses, specialised vocabulary is one of the core elements that has brought about an increasing demand and extensive research. (Ng et al., 2013). To identify specialised vocabulary and investigate which vocabulary is especially frequent, corpora have proven to be useful. A corpus allows for both quantitative and qualitative analysis and may assist language teachers in tailoring ESP materials and syllabi to their learners' needs (Shamsudin, Husin, & Manan, 2013). One concrete form of output of such a lexical corpus analysis is a frequency-based word list. Such a word list can feature as a key component of a lexical syllabus in an ESP course. An example of such a word list is Coxhead's (2000) Academic Word List (henceforth AWL), based on a corpus of texts written by students, with the aim of providing future students with the means to cope with their academic reading. Hyland and Tse (2007), however, argue that word lists like Coxhead's (2000) are too generic since individual words may behave and occur differently in each discipline. To be most valuable, then, they argue that word lists should be "more local and discipline-specific" (Hyland & Tse, 2007 p. 109).

What becomes evident is that there is a need for word lists that are more discipline-specific so that teachers can identify vocabulary patterns in specific disciplines of English language use. The current study sets out to precisely examine this recommendation in that it aims to compile a word list that is relevant for use in an Engineering English course for a specific target group. Several studies have generated word lists from general (West, 1953), academic (Coxhead, 2000) and specialised (Mudraya, 2006; Ward, 2009) corpora. One of these specialised contexts is that of Engineering English, which focuses on the needs of engineering students and practitioners. Several inquiries into Engineering English have revealed the relevance and need for engineering students to develop a repertoire of discipline-specific words

(Mudraya, 2006; Ward, 2009; Watson Todd 2017; Ng et al., 2013). These studies have resided within a EAP context, however, focusing on engineering students' needs rather than engineering professionals' needs.

The need for an Engineering English word list was established in a needs analysis into the language and communication needs of Field Service Engineers at Marel (Humphrey, 2019). Marel is a multinational provider of advanced processing systems, software, machines and services to the poultry, meat, and fish industries (Marel, n.d.). Their employees work in over 30 countries or at one of their 100 partners' on-site locations across six continents (Marel, n.d.). The present study reduces its scope to Marel Boxmeer, which is a branch of the company in the poultry processing industry. To install and maintain machines their customers' machines, Marel employs Field Service Engineers (FSEs). FSEs concern themselves with stakeholder management, preparation, global travel and performing service visits on a daily basis (Marel, n.d.). FSEs are therefore technical specialists who also keep in touch with customers and manage personnel in factories all across the globe. Next to FSEs being specialists within their fields of service, Marel also expects FSEs to be fluent and eloquent in the English language to perform accordingly in stakeholder management and training customers' engineers. Most FSEs, however, have not had any schooling higher than a Dutch mtslevel (i.e. an intermediate technical school). Consequently, they have not received training at a level beyond basic-to-intermediate General English. Though highly technical, FSEs therefore often lack the proficiency in advanced or technical English. As has become clear from the needs analysis of this target group, FSEs especially experience difficulties in terms of productive language skills (Humphrey, 2019).

To prepare starting FSEs for the professional work field, Marel established the Marel Academy as a training programme for technical and practical groundwork for their function as FSE. The purpose of the present study is to contribute to the Marel Academy in the form of a word list for use in a lexical syllabus in a potential ESP course for newly appointed FSEs in their trial time. This way, FSEs are equipped with the lexical means to function in an international engineering environment. In effect, this study reports on the construction of a local and discipline-specific (i.e. Engineering English) word list, the Professional Engineering Word List, that is relevant and useful for the target group in terms of text coverage and general frequency. This proposed word list is developed out of a corpus of Engineering English: The Professional Engineering English Corpus, which consists of user manuals which FSEs refer to daily, and customer reports written by FSEs themselves.

The justification of this study is threefold. Firstly, this study takes up the call in recent ESP literature to conduct empirical corpus research and generate specific word lists. (Hyland & Tse, 2007; Paltridge & Starfield, 2013; McEnery & Xiao, 2011). These word lists can then be used by teachers to corroborate their materials when designing ESP curricula and syllabi. In effect, this method adheres to the lexical approach in language teaching, which is strongly supported in research (Lewis, 1993; McEnery & Xiao, 2011). Secondly, this thesis proposes to apply more focus to the domain of English for Occupational Purposes in corpus design, given that the majority of learner corpora designed for ESP reside in the realm of English for Academic Purposes (Shamsudin et al., 2013). Learners in this specific EOP context (i.e. FSEs at Marel Boxmeer) are homogenous in terms of education level, age, and most straightforwardly their profession. According to Hyland (2002) and Basturkmen (2003), such a homogenous group of learners benefits most from narrow-angled ESP courses based on their specific language needs instead of generic language skills. Similarly, the more experienced the learners are, the more specific the materials ought to be (Basturkmen, 2003). In order to take up the call for more specificity in course design, as argued by Hyland (2002), the concrete word list that is developed out of the corpus in the present study may feature prominently in the lexical syllabus of a future ESP course in the Marel Academy. In that sense, this study also intends to bridge the gap between linguistic research and language pedagogy. Lastly, by relating the present corpus to other general service, academic, and engineering word lists, this study intends to determine the nature and technicalness of the most frequent words in the corpus. That way, this thesis helps gain deeper insight into the exact nature of specialised vocabulary in a specific professional setting. As pointed out by Chung and Nation (2003), the role of technical vocabulary in specialised language is generally underestimated. More information is therefore required on how technical vocabulary relates to general service, academic, and other specialised vocabulary.

In sum, the overarching research objective of this study is to construct an EOP Engineering word list that is representative of technical writing in the target domain as well as sufficiently distinct from existing general service and discipline-specific word lists. In pursuit of meeting this objective, the following three research questions will be addressed:

RQ1: What is the frequency and distribution of specialised vocabulary in the PEEC?

RQ2: What is the nature of the most frequent vocabulary items in the PEEC: highly technical, sub-technical, semi-technical or non-technical?

RQ3: How do vocabulary items found in the PEEC relate to West's (1953) *General Service List*, Coxhead's (2000) *Academic Word List,* Mudraya's (2006) *Student Engineering Word List* and Ward's (2009) *Basic Engineering List*?

To answer these questions, this thesis is organised as follows. Chapter 2 will provide a theoretical framework and an overview of relevant literature in the field of ESP, corpus linguistics, vocabulary, and the development of word lists in relation to the specific target group at focus. The third chapter will outline the methodology of the present study and explain how the corpus was analysed quantitively and qualitatively. The construction of the Professional Engineering Corpus and the Professional Engineering Word List will also be addressed. Chapter 4 will provide a detailed account of the results and a thorough explanation of the findings in relation to the research questions. This chapter is also coupled with a discussion of the findings in relation to the relevant literature outlined in the first chapter. Finally, the last chapter will summarise and synthesise the main findings and draw conclusions. Any limitations and implications of the present study will also be discussed in this chapter.

CHAPTER 2: THEORY AND BACKGROUND

2.1 Introduction

This chapter will explore the theoretical framework in which this thesis is situated. It will provide the characteristics of the field of English for Specific Purposes and consequently apply focus to the discipline of Engineering English. What follows is a background on a needs analysis that was conducted on the target group at focus, which will underscore the necessity of a technical and discipline-specific word list for this specific target group. Next, relevant theory and concepts within corpus linguistics will be addressed. A review of the lexical approach in language learning and teaching is followed by theory about vocabulary knowledge, size, and terminology. Lastly, this chapter unites insights from corpus linguistics with those from vocabulary theory to discuss multiple studies which derived word lists out of corpora. These are the word lists that will ultimately be reviewed and compared to the word list that was constructed for the purpose of this thesis.

2.2 English for Specific Purposes

2.2.1 ESP theory and practice

As Belcher (2009) justly states, English, or any language, is ideally taught with a specific purpose in mind. The reality of this, however, is that language instruction is subjected to larger educational rules and policy which do not always match the learners' purposes. To these learners, language instruction may therefore likely resemble something along the lines of "language for no purpose", or even "language for other people's purposes" (Belcher, 2009, p. 1). What distinguishes an English for Specific Purposes (henceforth: ESP) approach from other approaches of English language teaching, is that the former approach takes the individual learner's goals and purposes in mind. It is especially relevant for the current study to provide a comprehensive definition of ESP for future reference. Anthony (2015) defines ESP as follows:

> English for Specific Purposes (ESP) is an approach to language teaching that targets the current and/or future academic or occupational needs of learners, focuses on the language, skills, discourses, and genres required to address these needs, and assists learners in meeting these needs through general and/or discipline-specific teaching and learning methodologies. (Anthony, 2015, p. 2)

From this definition, it becomes apparent that learners in an ESP course have a particular goal and purpose with regards to English. Courses that adhere to an ESP approach of teaching and learning ought to centralise the learners' linguistic and communicative needs. ESP courses cover exactly those domains of language that learners ultimately need to be able to perform in for their studies or jobs. The group of learners in an ESP course, which are typically (although not always) adults, is generally homogenous in terms of their learning goals yet not always in terms of their language proficiency (Paltridge & Starfield, 2013).

Kırkgöz and Dikilitaş (2018) point out that the field of ESP arose after the Second World War due to the immense transformation of scientific, technical, and economic activity. These developments led to a demand for an international language, which English was able to supply as the world's *lingua franca* – a common language for speakers with varying mother tongues across the world. This created a new kind of learner with individual motives for learning English who require tailored instruction with regards to their specific language goals. The guiding principle of ESP at that time can be summarised via the phrase "tell me what you need English for and I will tell you the English that you need" (Hutchinson & Waters, 1987, pp. 7-8, cited in Kırkgöz & Dikilitaş, 2018, p. 2). Thereafter, ESP evolved as a dynamic, interdisciplinary, and global field of research. ESP research and courses are generally subdivided into English for Academic Purposes (EAP) and English for Occupational Purposes (EOP) (Paltridge & Starfield, 2013). Where EAP is concerned with helping learners study, conduct research and teach in English, EOP refers to English for a variety of professions and vocational purposes in either work or pre-work situations (Flowerdew & Peacock, 2001, p. 8). These two disciplines of ESP have evolved over the course of years, leading to the establishment of ESP courses and subdisciplines such as Aviation English, Maritime English, and Business English, to name just a few. This extensive range of subdisciplines reflect the specific learner needs and target communities.

Whilst an ESP approach may appear to be quite straightforward, it is more difficult to actually meet the goal of specific learner-centred language instruction (Belcher, 2009). For an ESP approach to succeed, an effort on both the part of the teacher as well as the learner is required. The teacher needs to be prepared to immerse themselves in subject matter for academic or occupational purposes that they may not be familiar with. Additionally, the teacher is required to engage in critical reflection as to whether the learners' purposes are served in an ESP course (Belcher, 2009). On the side of the learner, it is essential to be patient for the teacher who may not be familiar with their field of study or expertise. Next to that, learners need to be conscious of the role of their own motivation and effort, so that the ESP course can provide the best returns in terms of language learning.

2.2.2 ESP research

As Anthony (2015) justly points out, ESP is one of the most prominent areas of EFL teaching today. Next to that, ESP has grown into a broad field of research with two international peer-reviewed journals, *English for Specific Purposes* and *Journal of English for Academic Purposes,* which put forward theoretical frameworks and findings that are subsequently put to practice in teaching. Despite being an ESP practitioner, it is also essential to function as a researcher who focuses on the literature and main methodology as a theoretical basis to create courses and develop materials. In that sense, there is an overlap between theory and practice in ESP, as ESP practitioners ought to be informed of state-of-the-art developments in the field. This research field of ESP has become increasingly specific, more narrow-angled, that is, and increasingly research-based from the 1960s onwards (Paltridge & Starfield, 2013).

Paltridge and Starfield (2013) reviewed the theoretical developments in ESP research via a review of the two aforementioned journals. They argue that language and discourse research of specific purposes genres is still as important as it was in the early days of ESP research. Especially genre studies are still largely popular, though the focus has shifted to the socially situated nature of genre in specific contexts as well as to the multimodality of texts in digital genres. Other main trends within ESP research are studies on English as a lingua franca in specific purpose settings, research into advanced academic literacies, identity in teaching and learning, and ethnographic approaches. Belcher (2009) argues that inquiries into disciplinary language and ESP teaching are also attracting ESP researchers' interests.

Another largely prominent development in ESP research which is especially relevant for the scope of this study is the trend of corpus-based studies. As Starfield (2014) argues, corpora have aided in gaining a better understanding of the nature of specific purpose language use by virtue of the large scale in which corpus-research is carried out. Since ESP practitioners are generally not well-versed in their students' professions or disciplines, they may lack intuitive understanding of language use in these domains (Nesi, 2013). A corpus, then, may help ESP practitioners gain an insight into their learners' language use or the language they can expect to come across.

In sum, Paltridge and Starfield (2013) see that the field of ESP has moved away from research on linguistic descriptions in a text and discourse analytic perspective and towards genre-based, corpus-based, and ethnographic research. In that sense, ESP research has become highly diversified and is bound to grow rapidly due to the global reach of English lingua franca speakers who are reshaping English for their own purposes.

2.2.3 Engineering English

In this study, the focus of inquiry is on Engineering English as a subdiscipline of English for Science and Technology, which is in its turn a subdiscipline of English for Occupational Purposes. Naturally, engineering itself is a broad discipline with multiple branches, such as electrical, mechanical, civil, structural, and industrial engineering. Given that engineering is also an academic field of study, there is also a role for engineering English within English for Academic Purposes.

Research within applied linguistics has focused on numerous aspects of Engineering English. With regards to research within an EAP dimension, several studies have considered the role of word lists for engineering students. Ward (2009) concluded that it is complex to design a mutual engineering word list for academic purposes due to the expansive nature of the discipline. Similarly, Hyland and Tse (2007) conducted a multidisciplinary corpus analysis of various engineering texts in which they argued that a single academic engineering word list could not serve the purposes of all engineering disciplines. Despite this, they were able to construct a semi-technical word list for mechanical and electronic engineering (Hyland & Tse, 2007). This conclusion is in line with Mudraya (2006), who argued that sub-technical and non-technical terms from the academic register showed enough overlap in terms of word families so that a word list can be established. Corpus studies are also being conducted within the engineering English subdiscipline for a variety of applications in addition to word lists. Despite the aforementioned research providing an initiative for creating word lists for Engineering English, these endeavours remain within the realm of EAP (see for instance Mudraya, 2006; Ward, 2009; Ng et al., 2013). In a professional or EOP context, research on Engineering English tends to focus on creating a needs analysis (Spence & Liu, 2013). Corpus research within this domain is scarce, therefore creating a demand for research founded on a professional engineering corpus of English, similar to that of Hyland and Tse (2007).

2.3 Needs Analysis Field Service Engineers

In October 2019, a needs analysis (henceforth NA) was conducted on Field Service Engineers (*N* = 34) at Marel Boxmeer to shed light on their English language and communication needs (Humphrey, 2019). This NA consisted of a triangulation of a questionnaire for FSEs, semi-structured interviews with Dutch FSEs, non-Dutch FSEs and managers, and an analysis of reports written by Dutch FSEs and on-site observations. From this NA it became evident that FSEs’ writing is often the target of intercultural and factual miscommunications. Next to that, both managers and FSEs themselves stressed the importance of a course on Engineering English for starting engineers in their training period. They argued that general English courses at Dutch schools for higher professional education do not fully grasp what it means to be a global engineer and how this translates to their English language skills. They felt that the general English courses were a sufficient foundation to build specialised language skills on, but that not enough attention is paid to specific Engineering English in their training period or thereafter. They are often left to their own devices and therefore tend to communicate with colleagues and engineers across the globe via translation tools, hand gestures, other colleagues, or interpreters who translate for them. One reason why these vocabulary items cannot simply be extracted from existing coursebooks on technical vocabulary is because engineering as a discipline has multiple subdisciplines which are markedly different from one another. The target group of the present study, Field Service Engineers, are mostly educated within electrical and mechanical engineering, but are also often mechatronics graduates. Most FSEs are all-round machine engineers who are experienced in a range of engineering disciplines either via education or via experience. Put differently, whatever may be relevant for FSEs may not at all be relevant to civil engineers. Therefore, an Engineering English course needs to be developed with the specific target group's needs in mind, which are in this case the needs of the FSEs.

When considering the course materials that may be designed based on this needs analysis as well as the present study, it is relevant to consider the target group in light of curriculum and syllabus design. As Hyland (2002) points out, English language courses are nowadays threatened by conceptions which argue in favour of wide-angled courses to cater for a larger group of learners with similar, yet far from identical needs and interests. Such courses target generic language skills which can be of use in a broad range of disciplines (Basturkmen, 2003). Although such courses are easier to construct and generally more cost-effective, this wide-angled view is not always as beneficial for a specific target group, however. With regards to a potential ESP course for FSEs, the advice resulting from the NA is to include authentic communicative events that require English language and communication skills. This way, the course may be perceived as a more readily relatable course that covers relevant communicative events instead of their previous general English courses that – in their view – merely laid language foundations that were far from workplace-specific. These courses have generally been disregarded by most FSEs. It is therefore sensible to create an entirely different atmosphere in an ESP course and incorporate authentic content so that they feel the course is relevant. Considering that the target group, starting FSEs, are comparable in education level, age, gender, and profession, the target group can be considered as homogenous. As Basturkmen (2003) argues, this influences the specificity of the materials. A homogenous group of learners with similar goals profits from a narrow-angled, or specific, course design which targets their professional language needs. The reason for this is that narrow-angled course designs are considered to be highly motivating due to the obvious relevance for the target group. Consequently, narrow-angled courses are expected to yield higher returns in terms of learning.

2.4 Corpus linguistics in theory and practice

2.4.1 Corpus linguistics theory

Bennett (2010) defines corpus linguistics as "the study of language in use through corpora" (p. 2). A corpus, then, is defined as by Sinclair (2005), one of the most influential researchers within the field of corpus linguistics and ESP, as

> a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research. (Sinclair, 2005, p. 23)

However, as Nesi (2013) points out, the term *corpus* is also sometimes defined more loosely, for instance as collections of texts regardless of their form (i.e. electronic or not). Tribble (2002) argues for two major basic features, an electronic format, and the fact that it is designed and planned with a purpose in mind, albeit a general or specific one. McEnery, Xiao & Tono (2006) ultimately review that there is increasing consensus in the field that a corpus has four characteristics, in that it consists of:

> (1) machine readable
>
> (2) *authentic* texts (including transcripts of spoken data) which is
>
> (3) *sampled* to be
>
> (4) *representative* of a particular language or language variety. (p. 3)

Despite this, there is still disagreement on the third and fourth characteristic: what counts as representative and what sampling techniques should be employed to achieve this representativeness? (McEnery et al., 2006). The authors of the above definition claim that the definition of the term *corpus*, while useful, is at the same time vague and may sometimes unjustly exclude carefully composed collections of texts simply because the term is imprecisely defined. Their definition is nevertheless in line with the one postulated by Sinclair (2005), which is the definition the present study will employ as well. In the composition of this definition, Sinclair (2005) based himself on a common notion in linguistics which argues that words do not carry meaning in themselves, but that meaning is made through several words in a sequence instead. In corpus linguistics, these sequences of meaning are then reviewed in terms of their lexical and grammatical features to find general patterns. These patterns can, in their turn, provide more information about frequency, register and language use in general (Bennett, 2010).

Similar to the discussion about the definition of corpus linguistics, there is no clear consensus on whether corpus linguistics is a discipline, a methodology, or a theory either (McEnery et al. 2006). Tognini-Bonelli (2001), for instance, argues that corpus linguistics has gone beyond functioning as a method and can be considered as an independent discipline (p. 1). The reasoning for this is that corpus linguistics employs an innovative and philosophical approach to study linguistics. McEnery et al. (2006), however, maintain that corpus linguistics is best described as a methodology, since it is not on par with other independent branches of linguistics such as phonetics, syntax, semantics, or pragmatics (p. 4). Provided that these fields all describe or explain an aspect of language use, corpus linguistics is not limited to particular aspects of language and can instead be employed to explore virtually every area of linguistics (McEnery et al., 2006). Since corpus linguistics indeed employs specific methods and principles, it arguably has a theoretical status (Tognini-Bonelli, 2001). However, this does not mean that it is a theory in and of itself. McEnery et al. (2006) illustrate this by means of pointing to on the qualitative methodology, which has its own set of rules and is yet still labelled as a methodology which can be employed to construct other theories on (p. 5). In sum, corpus linguistics may best be described as McEnery et al. (2006) put it, as “bedeviled with definitional confusion” (p. 5). Many of the above definitions fail to persist when considered in light of specific examples, which is why the present study will acknowledge corpus linguistics is a methodology with many possibilities and applications across multiple fields of linguistics (McEnery et al., 2006).

2.4.2 Corpus building

For the purpose of the present study, which intends to construct a corpus of professional Engineering English from documents the present target group often encounters, it is also relevant to consider what needs to be taken into account when creating a corpus in practice. When reconsidering the definition of a corpus, three traits can be discerned (Bennett, 2010). Firstly, a corpus ought to be principled, in that the language in the corpus should not be chosen at random but instead needs to adhere to certain traits. This is especially crucial for specialised corpora such as the corpus at present scrutiny, given that this corpus is intended to cater for a specific target group in terms of the language they can expect to occur. Although it is important for specialised corpora to be principled, this is equally crucial for larger and general corpora to provide a sound basis for any potential generalizations on basis of the corpus (McEnery et al., 2006). Ultimately, then, these principles can be chosen by the researcher themselves to fit the focus of inquiry both for general and specialised corpora. The second factor that should receive attention is that the texts in a corpus must be authentic in that they serve a natural, general, and genuine communicative purpose. In other words, the texts in the corpus should not be created for the sole purpose of serving its function in a corpus. The last trait that Bennett (2010) discerns is the fact that the texts in a corpus are stored electronically. This way, the corpus can be easily and readily accessed online via a computer. In essence, the corpus approach therefore cannot be effectively employed without the use of a computer (McEnery et al., 2006).

In order to ultimately develop a theoretically sound word list, it is crucial to circumvent the criticism that prior corpus-based studies have received. In terms of representativeness, corpora which consist of less than 200,000 words are considered to be small corpora. The rule of thumb for corpus size is that bigger corpora generally provide a larger opportunity for lower frequency items to occur (Coxhead & Hirsch, 2007). However, the size of any corpus really depends on both practical considerations and research focus (McEnery et al. 2006; Coxhead, 2000). Specialised corpora, for instance, may be smaller than general reference corpora, as they represent a smaller subset of language use and have a smaller research focus. (Hunston, 2002, p. 15). Coxhead (2000) also encourages to focus on the research purpose and use of the corpus instead of constraining a corpus to a fixed number of tokens. Instead of looking at the absolute size of a corpus, McEnery et al. (2006) suggest considering the degree of *closure* or *saturation* of a specialized corpus. Saturation can be measured for a particular linguistic feature of a variety of language to see whether that feature is finite or at least limited in terms of variation beyond a certain point. When applying the theory of saturation to develop a test of saturation, the theory prescribes that adding a section of identical size entails that the number of new tokens of previously identified items in any consecutive section should be similar to that in prior sections (Shams, Elsayed & Akter, 2012). In other words: when adding a new segment yields the same number of new lexical items as the previous segment, the corpus is considered to be saturated (McEnery et al., 2006, p. 16).

Any discussion of corpus size necessarily touches upon questions of corpus representativeness and balance (Nelson, 2010). Representativeness is by default a questionable and major issue when compilating a corpus (McEnery et al. 2006; Coxhead & Hirsch, 2007). The reason for this is that it requires in-depth knowledge of the genre and its speakers or learners. In the context of this study, this knowledge constitutes the field of Engineering English and the daily tasks of a FSE. McEnery et al. (2006) assert that all corpora should naturally be as representative as possible yet argue that the representativeness of a corpus largely depends on the research questions. In other words, they propose to interpret the representativeness of a self-compiled specialised corpora in relative, instead of absolute, terms. They illustrate this by saying that one typically does not know the distribution of language production and genre of a specific target group or text type. In the case of the PEEC, a needs analysis has been conducted prior to the creation of the corpus to chart the language use of FSEs. The genre has therefore been extensively examined, as has been illustrated in the NA (Humphrey, 2019).

2.4.3 Types of corpora

Hajiyeva (2015) points out that corpora have played a large role in critically evaluating syllabuses and teaching materials for EFL. Corpora provide insights into authentic language use as well as frequency data that can be interpreted in terms of prioritisation of curriculum content, since the most frequently used words and structures are often also useful to know (Hajiyeva, 2015). Bennett (2010) explains that there are four types of corpora that are especially relevant when employing the corpus approach to ESP syllabus and course development. A distinction can be made between generalized, specialised, learner, and pedagogic corpora. For the purpose of this thesis, the second type of corpora, specialised corpora, are the most relevant type which will be referred to in detail.

The first type, generalized corpora, are generally large and comprehensive corpora consisting of a variety of language as a whole. Examples of generalized corpora are *The British National Corpus* and *The American National Corpus*, providing both written and spoken texts from a wide variety of genres (Bennett, 2010). If a study aims to draw generalizations about language, it is advisable to consult a generalized corpus. Learner corpora, the third type, contain written and spoken samples of language use by learners of a second or foreign language and provide an insight into the characteristics of the interlanguage of these learners. A learner corpus can be both general and specialised, depending on the target group and genre covered by the corpus. The last type of corpora are pedagogic corpora, which contain language used in classroom settings. Bennett (2010) provides some examples, such as textbooks and transcribed classroom interactions, but really any text in an educational setting can be considered. Pedagogic corpora can be used for a multitude of purposes that all serve to monitor whether the classroom language is useful, pedagogically sound, and self-reflective for teachers (Bennett, 2010).

Specialised corpora, the most relevant type given the scope of this thesis, consist of specific text genres and registers that represent the language of this type. These corpora, regardless whether they are small or large, all aim to answer specific questions. Corpus linguistics can aid in answering questions within various areas of language teaching, especially in an ESP setting. Coxhead (2000), for instance, used a corpus to expound on key lexis within English for Academic Purposes. By means of such a specialised corpus, one can derive conclusions about relevant vocabulary, phraseology, and register, among other things. These conclusions can then serve as an empirical basis for course and syllabus design since they reveal what specific language the target group is likely to encounter in their areas of expertise. This way, teachers can prioritise specific language knowledge in the classroom and supplement course materials. One drawback to specialised corpora is that they are often solely available within the context and institution for which they were created (Nesi, 2013). Most personal ESP corpora were created for very specific settings, therefore being of little use for teachers and learners in other contexts. As Nesi (2013) points out, apart from composing their own corpora for specific purposes or considering findings from existing corpus analyses, ESP practitioners can also turn to ready-made corpora that are directly accessible. Examples of this are sub-corpora from general corpora, which allows for separating patterns in a particular register or genre.

2.4.4 Pitfalls and possibilities of corpus research

Despite the potential of corpus linguistics, several studies have also pointed out several criticisms which are reviewed – and refuted – by Flowerdew (2005). She explains that notable critics argued that corpus studies, and especially concordance output, generally lead to descriptions of language that are both atomised, in that they are rather fragmentised, as well as bottom-up. This is in direct opposition to the top-down approach of genre analysis which starts off with focusing on the macrostructure of larger units of texts before funnelling towards sentence-level patterns (Flowerdew, 2005). Another objection against corpus-based approaches is that they lack descriptions of the contextual features of the text. As corpus data are but a fragment of language use, samples of language are often separate from the communicative context that created it (Flowerdew, 2005). In other words, “reality … does not travel with the text” (Widdowson, 1998). This lack of visual and social contextual features may be especially problematic for specific types of corpus analysis, for instance pragmatic and socio-cultural ones, and is often considered as one of the gravest shortcomings (Flowerdew, 2005). McEnery and Wilson (2001) review Chomsky’s criticism against corpus linguistics, which boils down to two arguments: “using texts as the primary source of linguistic information, and the finite nature of a corpus” (McEnery and Wilson, 2001, p. 51). Using data and their frequency, which is the key element of corpus research, indeed clashes with the Chomskian distinction between competence and performance which prioritises competence (Tribble, 2002). The second point, the fact that a corpus is finite by nature, makes that not even the largest corpus can account for all possibilities in a language (Chomsky, 1962, cited in Tribble, 2002). Any generalisation from a single corpus, then, should be referred to as a deduction instead of a fact (Tribble, 2002).

As Tribble (2002) points out, modern corpus linguists are conscious of the shortcomings of corpora and have found possibilities for counteracting them. In terms of refuting the lack of contextual factors, which is arguably the main point of criticism, he asserts that corpus linguists can resort to interviews and focus group discussions with genre users to corroborate the results of the corpus component of a specific study. Another possibility, which is provided by Paltridge and Starfield (2013), is to read up on subject matter that is relevant within the specific discipline of focus so that the results of a corpus study become more meaningful and can be framed more critically. Lastly, whilst corpus studies can reveal several interesting things about the language use of a specific target group, it is crucial to not lose sight of these individual learners and their language needs (Paltridge & Starfield, 2013).

2.4.5 Corpus linguistics and the lexical approach

As Mudraya (2006) argues, corpus linguistics has recently come together with language teaching since language corpora can provide a theoretical basis for collecting information about the specific language that ought to be acquired. Previously, foreign language teachers have received criticism that they compile language learning materials which use a simplified form of language and do not adequately prepare students for target language use situations (Shamsudin, et al., 2013). Lewis (1993) coined the lexical approach to propose a shift from learning abstract grammar rules to prototypical and common examples of grammar in use. Mudraya (2001) explains the lexical approach as follows:

> The lexical approach argues that language consists of ‘chunks’ which, when combined, produce continuous coherent text, and that only a minority of spoken sentences are entirely novel creations (p. 236).

In doing so, she notes that Lewis (1993) explains that a distinction is made between *vocabulary*, which are individual words with fixed meanings, and *lexis*, which considers the word combinations that are stored in our mental lexicon. Corpora, then, can aid in identifying *collocations* (i.e. word partnerships), which are words that generally co-occur in natural text consistently instead of randomly (Lewis, 1993). Similarly, the lexical approach is directed at teaching collocations, arguing that vocabulary items, or *lexis,* ought to be presented in their context, which can be both grammatical and lexical. Whilst Lewis (1993) coined the term itself, other linguists such as Sinclair (1991) have referred to the existence and importance of multi-word units, stating that "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments" (Sinclair, 1991, p. 110).

The lexical approach has received some criticism, mainly because the chunks that Lewis (1993) coined exist in abundance, and it would be impossible to commit all of them to memory. Opponents argued that it would be easier to provide learners with grammar rules that they can use to construct phrases and sentences with (Mudraya, 2001). However, similar to the fact that there are many grammar rules, it is argued that it pays off in fluency to retrieve phrases from memory instead of composing sentences ad hoc when speaking. Fluency, then, is not at all possible without the ability to access an ample set of prefabricated chunks and expressions. Research has shown that chunks are rooted in our language, which is exemplified by the finding that chunk learning can foreground acquisition of the grammar system in an L2 (Lewis, 1993). Additionally, Mudraya (2001), shows that teaching ESP can be improved by the integration of the lexical approach with corpus linguistics. More specifically, she argues that techniques in corpus linguistics may play a substantial part in data-driven learning as learners knowledge of a language as well as the way they use ti may improve by virtue of focussing on corpus-based and form-focused activities. As an example, Mudraya (2001) points to including concordance lines in course design, which can inform teaching and learning as a whole.

As Mudraya (2006) argues, a corpus linguistics approach can contribute to an effective foreign language learning curve as students are presented with authentic and "real world" texts in an ESP course. When teachers compose materials based on authentic texts as they occur in corpora, their learners' proficiency in ESP may improve (Shamsudin et al., 2013). For any given ESP course, it is beneficial to base these materials on a custom-made corpus of relevant texts that students may come across in target language use situations. Similarly, language corpora may aid in providing a 'chunkier' view of language, as corpora show patterns and examples of words in their respective contexts (Mudraya, 2006).

2.5 Vocabulary

2.5.1 Vocabulary knowledge and size

Previous work on vocabulary learning techniques shows that instruction focused on form is an efficient language learning strategy (Gilner, 2011; Nation & Waring, 2000). Moreover, it has been suggested that focused vocabulary learning, especially focusing on high-frequency words, can provide substantial returns for novice learners of a second language (Durrant, 2009; Coxhead, 2000; Ward, 2009). Before addressing learners' needs in ESP courses, it is firstly important to consider the vocabulary knowledge and size that second language learners actually require to perform at a sufficient level in their target language. As Nation and Waring (2000) point out, there are three questions that together explain how much vocabulary a second language learner needs:

1) how many words are there in the target language,
2) how many words do native speakers know, and
3) how many words are needed to do the things that a language user needs to do? (p. 6)

Answers to these questions may, in their turn, help us to outline clear, sensible goals for vocabulary learning (Nation & Waring, 2000). Estimates of vocabulary knowledge and size may be especially relevant for designing ESP curricula as well, because vocabulary knowledge enables language use.

When answering the first question on the amount of words in the target language, which is English in the scope of this thesis, looking at the number of words in the largest dictionary available seems to be the most plausible and straightforward method. However, language is changing continuously with old words falling into disuse as well as new words being coined all the time. Next to that, dictionary makers are faced with determining whether words ought to count as words when they are compounds, archaic, abbreviated, or proper names.

The second question, how many words do native speakers know, is especially relevant for teachers of English as a second language since this can provide an insight into how large the learning task actually is for second language learners (Nation & Waring, 2000). Naturally, second language learners need not match native speakers in their competence and performance, but these insights can help the teacher to set achievable and measurable goals in terms of vocabulary size. The rule of thumb for vocabulary knowledge is that native speakers have a vocabulary size of around 20,000 word families to which they add roughly 1,000 word families a year (Nation & Waring, 2000). However, these numbers are not robust and vary per individual. Nation & Waring (2000) also show that the statistics vary across research, but that the variations can generally be attributed to which items are counted and how the term ‘word family’ is defined in theory.

The final question, which refers to the number of words learners need to perform accordingly in a language, can be answered by means of looking at word frequency. This measure describes how often a certain word occurs in regular use of a language. In English, and most languages for that matter, learners generally need to know only a small number of high frequency words to be able to understand the majority of a written or spoken text (Nation & Waring, 2000). O’Keeffe, McCarthy, and Carter (2007) for instance, determine that 2,000 of the most frequent words in a corpus “accounted for 80 per cent of all the words present” (p. 5). This information is quite significant because a vocabulary size of 2,000 to 3,000 words is then apparently an ample foundation for language use (Nation & Waring, 2000). It is therefore by all means not necessary for a learner of English to know 20,000 words, which is roughly the amount of words that adult native speakers know.

How much vocabulary does a second language learner require, then? Nation and Waring (2000) show that learners require the 3,000 high frequency words of the language before they can focus on other vocabulary. They argue that the low frequency words should be next focus of inquiry for learners. Teachers are encouraged to teach their students strategies to learn these words, for instance by guessing them from context or using mnemonics as a recall technique (Nation & Waring, 2000). When considering what type of vocabulary second language learners need, the answer differs per target group and situation. Whereas learners intending to pursue an academic degree may have a need for general academic vocabulary, the target group of the current study has indicated requiring specialised technical vocabulary for use in the engineering industry (Humphrey, 2019).

2.5.2 Vocabulary terminology

Nation (2001) explains that four types of vocabulary can be identified in a text: high-frequency words, academic words, technical words, and low-frequency words. High-frequency words are those words that are unmarked and generally consist of function words, though there are also several content words considered to be high-frequency words. Multiple frequency word lists may disagree with one another in terms of whether a particular word is in fact high frequency, depending on the cut-off criteria. However, Nation (2001) notes that in research which retrieves data from a well-designed corpus, there is generally 80% agreement about whether a word should be included. It is important to consider range as well as frequency, since range shows how many different texts or sub-corpora each particular word occurs in (Nation, 2001, p. 16).

In teaching, it is important to refer to high frequency words often and focussing learners' attention on them via direct teaching and incidental learning (Nation, 2001). Academic words are those words that occur in various types of academic texts, which make up roughly 9% of the running words in any given academic text. Low-frequency words are those words that are rarely used in the wider scope of language use, making up the largest group of all the words with thousands of them occurring in each language. Technical words, finally, are those words that are strongly related to the general theme of a text and are more common in this topic area than in other general areas. Whilst these words constitute circa 5% of the running words in a text, they differ vastly per discipline. It is this type of vocabulary that is particularly interesting given the scope of this study. Furthermore, as Chung and Nation (2004) stipulate, technical vocabulary is often an obstacle for learners in an ESP or EOP context, and thus also for the target group at focus.

2.5.3 Technical vocabulary

A technical word is defined as one "that is recognisably specific to a particular topic, field or discipline" (Nation, 2001, p. 198). Nation (2001) explains that the reason why technical vocabulary is distinguished from other vocabulary is because it allows for the identification of words that are relevant to learn for learners with specific language needs and goals. In that sense, technical words are also especially relevant in an ESP context where learners need to acquire a specific type of English. When groups of technical vocabulary are distinguished it becomes salient how these words may affect language learning goals. One of the ways in which this can be determined is by the number of words that are necessary to know in order to effectively use a language for specific purposes.

Some technical words are restricted to a specific area or discipline, whilst others occur across disciplines. Nation (2001) argues that this may cause varying degrees of what he calls 'technicalness', which can be demonstrated by organizing technical vocabulary in four categories. In this classification, the technicalness of vocabulary depends on the criteria of relative frequency of form and meaning. Despite this categorical division, words in all four categories share that their frequent occurrence in specialised texts within specific disciplines. Table 1 provides an overview of these categories as well as some examples.

Table 1. An overview of Nation's (2001) classification of four categories of technical vocabulary (pp. 198-199).

Category	Definition	Example
1: Highly technical words	The word form appears rarely, if at all, outside this particular field.	Applied Linguistics: *morpheme, hapax legomena, lemma* Electronics: *anode, impedance, galvanometer, dielectric*
2: Semi-technical words	The word form is used both inside and outside this particular field but not with the same meaning.	Applied Linguistics: *sense, reference, type, token* Electronics: *induced, flux, terminal, earth*
3: Sub-technical words	The word form is used both inside and outside this particular field, but the majority of its uses with a particular meaning though not all, are in this field. The specialised meaning it has in this field is readily accessible through its meaning outside the field.	Applied Linguistics: *range, frequency* Electronics: *coil, energy, positive, gate, resistance*
4: Non-technical words	The word form is more common in this field than elsewhere. There is little or no specialisation of meaning, though someone knowledgeable in the field would have a more precise idea of its meaning.	Applied Linguistics: *word, meaning* Electronics: *drain, filament, load, plate*

Nation (2001) suggests that words in Category 1, which are highly technical in nature, can be analysed by using frequency and range as criteria. He argues that words of this type are not sensibly pre-taught but are rather learned and understood by study and practice of a specific field. Apart from this definition of highly technical words, which focuses on their range, strictly technical words are also defined as not having exact synonyms and being resistant to semantic change (Mudraya, 2006). Words in Nation's (2001) Category 2 constitute semi-technical words in which the general meaning of a word does not provide ready access to the technical meaning and use of the word. Mudraya (2006) points out that the distinction between technical and non-technical vocabulary remains elusive. Prior studies have therefore also distinguished a third category to bridge the gap between non-technical and technical words: so-called *sub-technical* vocabulary (Category 3). Sub-technical words, then, are those that have both a technical as well as a non-technical meaning. They can also be identified based on their high distribution across all specialized fields (Yang, 1986, cited in Mudraya, 2006). Similar to sub-technical words, non-technical words (Category 4) are not exclusive to a specific field in terms of form or meaning. In that sense, non-technical words are by definition less technical than words in the previous two categories. Nation's (2001) overview makes a case for the argument that range is not enough to sensibly discern whether something is considered a technical word. Instead, the meaning of a word must also be considered.

Despite this classification, little is known about *how* to classify words into these categories. Chung and Nation (2004) argue that this is because any word's degree of technicalness can only be determined when the use and context of that word is considered. Being able to identify technical vocabulary reliably, however, is a crucial step in determining how technical vocabulary ought to be dealt with. In their study, Chung and Nation (2004) discern four methods which assist with the identification of technical words: via a rating scale, by using a technical dictionary, via clues provided in the text, and by using a computer-based approach. Their analysis revealed that using a rating scale was the most reliable and valid approach. This rating scale was based on Nation's (2001) classification of technical terms, which was supplemented with specific descriptions and information on the field of anatomy (i.e. the domain that Chung & Nation (2004) chose as an illustration to their methodology). The dictionary approach was also deemed successful, yet this measure is highly dependent on the availability of a relevant specialist dictionary. The clues-based approach was defined as arduous to apply, while the computer-based approach was evaluated as being more practical and easier to adopt. In sum, Chung and Nation (2004) argue that a combination of the approaches is the most successful method to identify technical vocabulary. Though laborious, a triangulation of the above methods is expected to yield the most reliable and valid results.

2.6 Word lists

To address the most important words in any given specialised discipline, researchers have generated word lists that are often incorporated in ESP course and curriculum design (Ng et al., 2013). Prior research has shown that language learners in courses focusing on language forms via word lists or word cards yield better results than learners in courses which do not cover such a component (Nation & Waring, 2000). In that context, word lists function as building blocks of the most important words for specialised fields, which help aid develop ESP materials (Ng et al., 2013). Nation (2001) explains that there are two systematic ways of developing word lists of technical vocabulary: by using a dictionary or by means of a corpus-based frequency count. This study will refer to prior research that made use of a corpus-based frequency count and will disregard research that compiled a word list out of a dictionary. The reason for this is that there are methodological issues in the use of dictionaries to create a technical word list. Amongst others, using a dictionary involves the problem of sampling and classification in that it is unclear how the compilation has taken place and how decisions to include a word were made (Nation, 2001). There are several well-known general and discipline-specific (i.e. Engineering English) word lists of the most frequently occurring words in English. The following section will review these word lists and research on their adequacy. Table 2 provides a comparative overview.

Table 2. A comparative overview of the GSL (West, 1953), AWL (Coxhead, 2000), the first 100 items from Mudraya's (2006) SEWL and Ward's (2009) BEL.

Word list	General Service List (West, 1953)	Academic Word List (Coxhead, 2000)	Mudraya's (2006) Student Engineering Word List	Ward's (2009) Engineering Word List (Basic Engineering List)
Counting units	Word family	Word family	Word family	Word type
Number of items in word list	2,000 word families	570 word families	1,200 word families	299 word types
Disciplines	General disciplines	Arts, commerce, law, and science	Basic engineering disciplines	Chemical, civil, electrical, industrial, and mechanical engineering
Corpus size	5,000,000 tokens	3,500,000 tokens	2,000,000 tokens	271,000 tokens
Corpus texts	Authentic and frequently occurring text types	Textbooks, laboratory tutorials, lecture notes, journal articles	Basic engineering textbooks (for all disciplines)	Engineering textbooks (for 5 disciplines)
Target audience	General English language learners	EAP learners (especially undergraduates)	Thai undergraduate engineering students	Thai undergraduate engineering students

Word selection principle	Frequency, ease of learning, coverage of useful concepts, and stylistic level	Range, frequency, and specialized occurrence: at least 10 in all disciplines (or at least 100 times in the corpus)	Most frequent word families (sum total of 100 occurrences or 0.005%)	Any word should occur at least 5 times in all engineering subdisciplines (or more than 25 times in the whole corpus)

2.6.1 The General Service List

West (1953) developed the *General Service List* (henceforth: GSL) which consists of 2,000 word families, via the frequency figures of a corpus of 5 million tokens. Due to the lack of computer resources at the time, semantic counts of the GSL were conducted manually (Gilner, 2011). Nevertheless, the GSL still provides roughly 90% to 95% coverage of tokens of colloquial texts and 80% to 85% coverage of common texts in English (Khani & Tazik, 2013). Whilst the GSL was issued more than six decades ago, it remains one of the best available high frequency word lists to date (Nation & Waring, 2000; Khani & Tazik, 2013). One reason for this is that the word list provides an insight into the frequency of the various meanings of each word. Furthermore, the words in the GSL are praised for their universality in that they are unrestricted to a specific time or place and occur across countries (Gilner, 2011). The GSL also considers the criterion of utility in that it includes words which can be used to discuss a broad range of topics and disciplines.

One oft-cited criticism of the GSL is the size of the corpus the GSL was based on (Browne, 2013; Khani & Tazik, 2013; Gilner, 2011). The 2.5 million-word corpus was collected without having the technological assets available today and is therefore considered as small by modern standards (Browne, 2013). Another point of criticism in terms of range is made by Engels (1968), who showed that the second 1,000 words in the GSL cover a mere 4.7% of running words in his corpus of non-fiction texts. He therefore considered these words to be “fallacious”, as they fail to represent truly frequent *general* English words and do not occur enough times across texts (p. 266). However, as Gilner (2011) argues, the methodology of Engels’ (1986) study was flawed in that it reported using a list of 3,372 words instead of the 2,000 figure that West (1953) developed, making it numerically impossible to find 3,372 words in his considerably smaller collection of texts (i.e. ten 1,000-word texts). In sum, if frequency distributions are sloped among frequent words, it is not surprising that his analysis yielded a disproportionate finding in terms of range. However, the issue of inconsistencies in what constitutes a headword or a word family remains, which is precisely what causes the present debate on the exact number of words in the GSL. Coxhead (2000) also criticized the nature of the corpus which the GSL was developed out of. She argues that it seems to have been replete in terms of genre, as the GSL covered a mere 76% of non-fiction texts. Nation and Waring (2000) corroborate by demonstrating that the coverage of the words of the GSL differs per genre, with academic texts yielding less coverage (78.4% in total) than novels or textbooks (90.6% in total). A final point of criticism that was first raised by Richards (1974), but later also by others (see for instance Browne, 2013), is that the GSL is dated. On the one hand, he claims that a lot of words in the GSL have limited utility and are not used as often as they were decades before he published his article. At the same time, he reasoned that words that were common in the 1970s (such as *astronaut, television, rocket,* etc.) were evidently missing from the GSL (p. 71). Crucially, given that Richards' (1974) study was published nearly fifty years ago, his criticism persists since the GSL also does not include any of the technological and globalised advances of the past decades (Watson Todd, 2017). However, Gilner (2011) argues that studies that have found fault with the GSL found these errors not due to its datedness It would therefore be unreasonable to dismiss the GSL solely because of its age.

In sum, whilst the GSL may arguably need a revised version due to its age, the inconsistencies it contains and its sole written focus, it remains the best available (Nation & Waring, 2000) as well as the best researched (Gilner, 2011) frequency-range-based word list to date. Several studies have attempted to update the GSL, the most notable being the New General Service List (NGSL) by Browne (2013). Despite these efforts, the differences in text coverages are not considerable nor do they provide contrasting figures to those of the original GSL. The present study will therefore refer to West's GSL (1953) as its reference word list for general service vocabulary.

2.6.2 The Academic Word List

Coxhead's (2000) *Academic Word List* (henceforth: AWL) is an example of a word list for academic purposes which has been developed from a corpus of 3,500,000 running words. The AWL was developed on the basis of numerous representative sub-corpora, as she chose a wide range of scientific fields and divided them into four disciplines: arts, commerce, law and science with a total of 28 subject areas (Coxhead, 2000). By adopting word selection principles such as frequency, range, uniformity, and excluding general high-frequency words, Coxhead (2000) screened out 570 high-frequency words. These words were then grouped into ten sub-lists according to frequency. The words in the first three sub-lists occurred at least once in every twelve pages of academic text on average (Coxhead, 2000). The first two sub-lists account for approximately half of the total coverage of the AWL. According to Coxhead (2000), the AWL accounts for roughly 10% of the total words (tokens) in a selection of academic texts, whilst merely covering 1.4% of the total words of a similarly large collection of fiction. This provides evidence in favour of the AWL consisting of predominantly academic words and therewith underscores the specialized nature of this word list. Ultimately, Coxhead (2000) made the AWL available for instructors so that they could incorporate the word list in their students' vocabulary study (Bennett, 2010).

One of the limitations of the AWL, which Coxhead (2000) herself considered to be controversial, is the fact that she excluded the 2,000 words in West's (1953) GSL from this word lists when compiling the AWL. As has been argued, the GSL has been considered to be limited in range (Engels, 1968) and has been around for nearly 70 years (Richards, 1974). The question, then, is whether the AWL suffices to represent academic vocabulary today. According to Martinez et al. (2009), another weakness of the AWL is the uneven distribution of the headwords in the word list across the variety of academic texts. In addition, they point out that the words in the AWL have vastly different meaning across different disciplines. Hyland and Tse (2007) also point to the distribution of coverage of the AWL across the disciplines concerned with a 16% coverage in the field of Computer Science, for example, compared to 6.2% coverage for Biology. Therefore, it is highly recommended that ESP researchers design discipline-specific word lists to meet the needs of non-native speakers who have to use English in their jobs (Khani & Tazik, 2013). Engineering English is but one example of one of these disciplines that requires a specific word list. Another limitation of the academic corpus from which the AWL has been developed, which is especially relevant considering the scope of the present study, is the fact that it does not actually contain an engineering section (Ward, 2009). Whilst Coxhead (2000) did indeed construct a science sub-corpus, for which the AWL provides 9.1% coverage, the AWL covers 11.1% of Hyland and Tse's (2007) academic corpus and 11.3% of Ward's (2009) engineering corpus. It can therefore be argued that the Coxhead's (2000) corpus appears to do better with engineering material when compared to the generic science section of that corpus. However, the criticism that persists is the fact that the corpus underpinning the AWL is not wholly representative of the range of academic disciplines she aims to illustrate (Hyland & Tse, 2007; Watson Todd, 2017).

Despite the criticism of the AWL, Coxhead's (2000) study is impressive with regards to its consistency and comprehensiveness. The AWL was developed from a large corpus of a wide variety of academic texts, therefore adhering to principles of corpus compilation such as representativeness and genre variety as postulated by Bennett (2010). Nation (2001) has even gone as far as to describe the AWL as the "best list" (p. 12). Indeed, the AWL has been used in a plethora of domains, ranging from vocabulary testing (Nation, 2001), to the development of word lists for a variety of specific purposes (Ward, 2009), and developing corpus analysis software (Anthony, 2014). Lastly, Cobb and Horst (2004) showed that both knowledge of the AWL as well as knowledge of the GSL are crucial to understand academic texts in English. The AWL has a similarly important role in academic texts for specific purposes, as is exemplified by studies in Engineering English.

2.6.3 Engineering English word lists

Several studies have compiled word lists specifically for the field of Engineering English. However, provided that most researchers in this field are by definition active within academia, these word lists generally reside within the realm of English for Academic Purposes rather than on English for Occupational Purposes. This does not devalue the relevance of these word lists, however. The following section will examine two oft-cited Engineering English word lists in more detail.

Mudraya (2006) constructed a word list, the Student Engineering Word List (henceforth SEWL), of 1,260 word families out of a student engineering English corpus of nearly 2,000,000 running words. Her corpus consisted of compulsory textbooks for engineering students, disregarding the fields of specialisation and arguing for a more integrated approach. Furthermore, Mudraya (2006) points to the benefit of smaller corpora in language learning and teaching, since smaller corpora may tend to the specific needs of a language learner. The corpus she constructed aimed, among other things, to provide teachers with a word list they could work into a lexical syllabus foundation for Engineering English. For the construction of the SEWL, Mudraya (2006) selected the most frequent word families with a threshold of a total of 100 occurrences, or 0.005%. In her analysis, Mudraya (2006) explains that the SEWL is generally sub-technical, saying that they are "words with non-technical as well as technical senses, common in most kinds of technical writing, which are identified by their frequency and distribution as well as their collocational behaviour" (p. 242). Her study therefore shows that sub-technical vocabulary ought to receive more attention in the ESP classroom for this target group. Furthermore, Mudraya (2006) found that some verbs in her word list also occurred with high frequency in the AWL. This led her to suggest that engineering students may do well to focus on academic words in the AWL as well. Finally, Mudraya (2006) provides some suggestions for data-driven instructional activities which are in line with the lexical approach to cater for the need to focus on sub-technical words. A fundamental limitation this study, at least for the purpose of the current study, is that it did not include the full word list. Instead, only an appendix of the 100 most frequent word families of the SEWL is available.

Similarly, Ward (2009) created a basic engineering word list (BEL) with 299 non-technical word types from a corpus of English engineering textbooks. The 271,000-word corpus that Ward (2009) constructed was smaller and less comprehensive than Mudraya's (2006) corpus, but it was nonetheless claimed to be representative, reasonably comprehensive, balanced, genre-specific, and relevant to student needs. His corpus aimed to represent five major strands of engineering: chemical, civil, electrical, industrial, and mechanical. When compared to the ten most frequent words in Mudraya's (2006) corpus, which were also frequent in Ward's (2009) corpus, Ward (2009) found that these words were hardly relevant for three out of five engineering disciplines. This is to show that the engineering corpus in both studies contain several examples of uneven distributions as a result of including so many engineering disciplines. The BEL that was developed out of this corpus is aimed at Thai undergraduate students in science and engineering, who struggle with specialized usages of English in lexis. The BEL is relatively short and concentrates on word types instead of lemmas or word families so that they are learned not individually but within the context they occur. Given the non-technical nature of this word list, the word list especially lends itself to ESP teachers who may not be entirely familiar with the field of engineering itself (Ward, 2009).

Both studies have aided in providing engineering students with the means to tackle engineering texts and learn relevant vocabulary. The word lists that have been derived from these ESP corpora are a crucial source in addition to engineering dictionaries, as the latter do not contain all terms found in their textbooks (Shamsudin et al., 2012). As Nation and Waring (2000) point out, word lists and frequency information provide a sound rational basis for learners in terms of giving the best returns in their vocabulary learning efforts. These vocabulary word lists can, in their turn, feature prominently in the language curriculum of any target group.

CHAPTER 3: METHODOLOGY

3.1 Introduction

This chapter will outline and explain the considerations and challenges that have arisen with the creation of the Professional Engineering English Corpus (henceforth PEEC). The procedures and decisions in the compilation of the corpus will be laid out firstly. Then, a case is made to justify the choice of materials and design. The corpus annotation will be addressed next, specifically in light of the procedures that are carried out to edit and process the texts in the corpus. This chapter will also highlight the approach that is undertaken to analyse the PEEC to create the Professional Engineering Word List (henceforth: PEWL) which is the core focus of inquiry of the present study. These methods and procedures are justified with reference to previous studies and recommendations in the literature. Lastly, any ethical considerations that have arisen will be discussed.

3.2 Corpus compilation

This thesis will take a similar approach to Coxhead (2000) when designing the corpus to serve as a framework for a word list. Firstly, all of the relevant documents were collected and converted in the appropriate electronic form. As Nation and Waring (2000) recommend, any corpus should consist of a wide range of language that is representative for the research purposes. To meet this recommendation, the FSEs Technical Support Manager at Marel Boxmeer was consulted to report on the various target language use situations that are especially relevant for the target group. From this consultation, it became apparent that user manuals and customer reports were both the most frequent as well as the most important target language use situations for FSEs at Marel. This was also confirmed by one FSE. It was therefore decided to include these two types of sources in the PEEC. It should be noted that this corpus was compiled on the basis of written texts only, given that this was found to be both the largest component as well as the largest obstacle for FSEs' daily tasks as reported in the needs analysis (Humphrey, 2019). Similar to Mudraya (2006) and Ward (2009), the PEEC consists of whole texts instead of text extracts. This way, issues concerning the validity of sampling techniques were avoided as much as possible. Moreover, ensuring that the PEEC comprised of whole documents allows for more comprehensive linguistic analysis, given that concordance lines can be consulted to provide context. The general specifications of the PEEC are tabulated in Table 3.

Table 3. Specifications of the Professional Engineering English Corpus (PEEC).

Text type	Number of tokens	Average text length in words	Standard deviation in words
User manuals	59,813	6,222	1531.8
Customer reports	10,723	1,217	1393.2
PEEC (total)	70,536		

The present corpus consists of more than 70,000 running words. It was not decided to set a specific target in collecting a specific number of words, but instead achieve a sample of texts that are specifically relevant for the target group. The principle of relevance therefore overruled that of absolute size. The following sections will explain the nature of the corpus and illustrate the content of the user manuals and customer reports in more detail. In addition, a justification for the choice of these materials will be provided.

3.2.1 User manuals

Every customer who purchases machinery by Marel will receive a hard copy of the machine's user manual. These manuals contain blueprints, exploded views, and drawings of all machine units, as well as safety and hygiene regulations. The largest part of the manual is aimed at assembly instructions. FSEs who travel to customers always carry copies of the spare parts book as well these manuals. The reason user manuals were included in the compilation of the PEEC is because the needs analysis revealed that user manuals are the most consistent example of reading materials that FSEs encounter virtually every day (Humphrey, 2019). For the purpose of this study, a selection of manuals was made based on relevance for starting FSEs in consultation with Marel's Technical Support Manager. One FSE also confirmed the relevance of these documents. Ultimately, ten user manuals were included in the PEEC.

3.2.2 Customer reports

After visiting a customer abroad, FSEs at Marel evaluate their visit and document the tasks they performed and any issues that remain in so-called customer reports for future reference. The choice was made to include these customer reports in the PEEC for two main reasons. Firstly, in the needs analysis, writing customer reports was found to be the communicative event that occurred second-most frequently, with writing e-mails being the most frequent event

(Humphrey, 2019). Another reason for including customer reports is because these are the dominant modality in which FSEs write in formal English. Customer reports also serve as a meticulous and vivid illustration of FSEs' productive use of English, making it an apt addition to the corpus. The customer reports that were included in the PEEC were retrieved via one Field Service Engineer and a Technical Service Coordinator at Marel, who both have access to the database containing all customer reports. A selection was made on the basis of those customer reports that contained a lot of written text, as some FSEs tend to substitute written text with self-drawn figures, blueprints, and photos. Ultimately, ten customer reports written by ten different FSEs for nine customers were included. It should be noted that the amount of text varies markedly per customer report, since the length of the customer reports are dependent on the length of the FSEs' visit and the amount of tasks that needed to be performed.

3.3 Corpus annotation

As McEnery et al. (2006) argue, the degree of corpus mark-up and annotation, though essential for corpus construction, depends on the research questions. In case of specialised corpora, McEnery et al. (2006) suggest that only paragraphs, sentences, and other basic textual information are marked up. In the case of the present study, neither textual nor contextual information were relevant to mark-up or annotate in the PEEC, given that its focus is on building a word frequency list out of the data. Instead, a Python script was developed specifically for the PEEC to annotate the plain text. The full code of this script can be found in Appendix I. To analyse all texts and combine the separate documents into one larger corpus, several procedures were undertaken. First, all texts which were retrieved as *.pdf* documents were converted into plain text (*.txt*) format to match the input requirements in the software programmes. Next, all of the texts were slightly modified in the sense that any repetitive information such as contact details in headers and footers were left out to ensure that this did not skew the definitive frequency figures. The table of content and bibliography were also removed, if applicable. Any relevant text that was still in tables was manually post-edited and included in the corpus. For the customer reports written by the FSEs, all the words from the standard template were removed from the documents, so that only the texts written by the FSEs themselves remained. Ultimately, only plain, clean text remained.

Most conversion issues were fixed by virtue of the script, but any formatting issues or irrelevant content that remained was manually post-edited in the .txt documents. Given that the user manuals had already been professionally edited and proof-read by Marel, it was deemed unnecessary to review all user manuals for language errors. A random sample spell check across several user manuals confirmed this presumption. In the initial .txt output document, some issues with punctuation marks were found. The Python code was therefore altered to exclude all punctuation marks. However, this gave rise to the issue of hyphenated compound words, as these were treated as separate words in the output document. To ensure that such issues did not persist in the final output document for both the user manuals and customer reports, the output was manually checked by the researcher. All proper nouns that occurred throughout the corpus which were obviously context-irrelevant were also removed. An overview of all of these words which were ultimately excluded can be found in the Python code in Appendix I (lines 31-37). The decision to exclude these proper nouns has multiple rationales. Firstly, as Hajiyeva (2015) argues, excluding such words ensures that the overall size of the vocabulary is not overstated. Next to this, text coverage is less likely to shrink as a consequence of including proper nouns. Lastly, these proper nouns are often categorised as off-list in vocabulary profile analyses, thus not contributing much to the corpus analysis. Aside from proper nouns, abbreviations (such as *etc*) were also removed from the list, similar to Coxhead and Hirsch's (2007) pilot study. To avoid all such bias and distortion, the choice was therefore made to exclude a specific set of proper nouns.

3.4 Corpus analysis

In terms of analysis, corpora depend on both quantitative and qualitative techniques. Quantitative results generated from the corpus are consequently analysed qualitatively. This involves understanding the contexts in which the words and phrases are used (Bennett, 2010). Several software tools and webtools were used for the analysis of the PEEC. To provide an answer to the first research question concerning the frequency of specialised vocabulary in the PEEC, a vocabulary profile of the PEEC was developed. Initially, the attempt was made to use CompleatLexical Tutor (available via www.lextutor.ca) to create a vocabulary profile. More specifically, the Web VP Classic mode of CompleatLexical Tutor can provide a vocabulary profile of the corpus according to the GSL and the AWL. Provided that the size of the PEEC is larger than 35,000 words, however, the corpus could not be processed in the Web VP Classic mode. The alternative tool, VP-Compleat Input mode B (for files up to 1,000,000 words), provides less rich output with a framework based on word families in Nation's (2017) BNC/COCA word frequency lists, which fall beyond the scope of this thesis.

Instead, the PEEC was loaded into AntWordProfiler, a freeware programme developed by Anthony (2014). As reference word lists, the first and second 1,000 words of the GSL and the AWL were chosen to demonstrate how the vocabulary in the PEEC relates and provide an estimate of how specialised the vocabulary in the corpus texts are. AntWordProfiler categorises the words in the corpus into four levels: 1, 2, 3, and 0. In this study, level 1 and 2 refer to the first 1,000 and second 1,000 words in the GSL respectively, whereas level 3 refers to the words in the AWL. Level 0 words are off-list words. Broadly speaking, texts containing a large percentage of level 3 and 0 words indicate that these texts are especially difficult for its readers.

AntConc, also developed by Anthony (2019), was used create a keyword list which calculates the keyness values for the words in the PEWL. Since absolute frequency numbers do not provide enough information on word coverage, the keyness values for each of the words were calculated via the keyword creation tool in AntConc. A keyword list gives a measure of statistical saliency, based on chi-square or log-likelihood measures (Anthony, 2019), whereas a simple word list only provides frequency figures (Baker, 2006, p. 125). Positive keyness refers to words being used more frequently in the PEWL than in the reference word lists, whereas negative keyness indicates that words occur more frequently in the reference word lists than in the PEWL. For the purpose of this study, only positive keywords, which are those words that are overrepresented in the PEEC when compared to the GSL and AWL, were considered. The AntConc settings for the keyness values are shown in Table 4. The keyword list tool can be used to generate positive and negative keyness, based on chi-square or log-likelihood measures (Anthony, 2019). AntConc was also used to generate concordance lines. These concordance lines can reveal more context and information on the nature of words in the PEWL and potentially uncover compounds, idioms, and multi-word units.

Table 4. AntConc settings for keyness values.

Keyword Statistic	Log-Likelihood (2-term)
Keyword Statistic threshold	$p < 0.05$ (+ Bonferroni)
Keyword Effect Size Measure	Dice coefficient
Keyword Effect Size Threshold	All values

3.5 Creating the PEWL

In order to develop a frequency-based word list out of the Professional Engineering English Corpus and provide an answer to the first research question, the corpus was uploaded in AntWordProfiler (Anthony, 2014). In AntWordProfiler, the settings for the tool preferences for the word list were changed to treat all data as lower case, given that this study is not concerned with placement of words or use of capitals across the corpus. The first version of the frequency list that was processed by AntWordProfiler was filtered to remove any irrelevant of inappropriate words because of faulty post-editing. Examples of this include non-words and abbreviations such as *plc, fig,* and *eng.* Next, word families that were not automatically filtered out by the GSL and AWL (e.g. *adjust* and *adjustable*) were manually lemmatised by the researcher based on the alphabetised raw word frequency list. To develop a truly discipline-specific word list, the AWL was ultimately excluded in the level lists. In other words, the definitive version of the PEWL consists of words that are also found in the AWL. Whilst Mudraya (2006) argued that the AWL ought to be included as an additional resource to her SEWL, since a lot of items from the AWL also occurred in the corpus the SEWL was based on, it was decided to circumvent this by including items from the AWL in the PEWL itself instead. Next to that, provided that the target group in this study has not received academic training, they cannot be expected to be familiar with all items in the AWL.

3.5.1 Word list recommendations

Nation and Waring (2000) point to several factors that ought to be considered when developing a resource list of high frequency words. Similarly, Coxhead (2000) describes the three criteria which she uses for word selection for the AWL, which partly overlap with Nation and Waring's (2000) recommendations. This study will refer to these recommendations and elaborate on the decisions that were made when developing the PEWL out of the corpus. The authors' recommendations are summarised in Table 5.

Table 5. Summary of Nation & Waring's (2000) recommendations and Coxhead's (2000) criteria for selecting words for a word list.

Recommendation	Explanation
Representativeness	The corpora that the list is based on should adequately represent the wide range of used of language.
Frequency and range	A word should be included in a word list on the basis of occurring frequently across a wide range of texts.
Word families	The word list needs to establish a sensible set of criteria regarding what forms and uses count as being members of the same word family in the context of the purposes of the list and learners for which it is intended.

Specialised occurrence	It should be specified how specialised the occurrence of word families should be, which can be related to other word lists.
Idioms and set expressions	It should be considered whether words occurring frequently as multi-word units or idioms ought to be included in the word list.
Range of information	A word list needs to include a wide range of information of its words to be of full use in course design.
Other criteria	Several other criteria may be considered in the final stages of developing a word list, including ease of learning, necessity, cover, stylistic level and emotional words.

Several decisions were made in consideration of these recommendations. In terms of representativeness, this study employed a similar approach to that of Shams, Elsayed and Akter (2012). As has been outlined by McEnery et al. (2006), the degree of saturation is a more comprehensive measure to approach corpus size and representativeness. Such a saturation test was also performed for the present study to justify the size and representativeness of the PEEC, like Shams et al.'s (2012) study. Firstly, the corpus was divided in five equal sections of 14,107 words on average via a self-developed Swift script, which can be found in Appendix II. The script was developed in such a way that each section included an equal proportion of words from the user manuals and the customer reports to guarantee an even distribution of corpus texts and both document types. For each of these sections, AntConc created a word list that excluded word families that already occur in the GSL. The normalised frequency numbers of each of these word lists are ultimately compared to assess whether the figures differed per section or remained stable. If these frequencies appear to become more stationary for each section, a corpus is said to be lexically saturated (Shams et al., 2012).

When considering the second criterion, frequency and range, it is important to determine when a word is included in the word list. This decision may be practical and theoretical in nature, provided that a word list should be both challenging and realistic in terms of achievability. The cut-off point for deciding whether to include a word in the PEWL was a minimum of ten occurrences across the entirety of the corpus. Whilst this cut-off point may appear arbitrary, the decision was made on the basis of findings in prior studies. Ward (2009), for instance, illustrates that his BEL, which consists of 299 word types, is a "modest, plausible lexical target for a year's work" (p. 177). Other studies have cited the number of words that Japanese and EFL students learn in a year in the low to mid-hundreds (Milton & Meara, 1995; Schmitt & Meara, 1997, cited in Ward, 2009). Thornbury (2002) argues that words occur at least seven times in any text for the vocabulary to be retained. Given that the aim of this word list is to function as resource in a lexical syllabus of an ESP course which aims at vocabulary retention, Thornbury's (2002) recommendation seems suitable to follow. However, when adjusting the threshold to words occurring at least ten times, the total length of the word list becomes more manageable to incorporate in an ESP course for the target group. As a result, the words in the PEWL are expected to be retained even better by FSEs.

In terms of range, it was decided to include a word was if it occurred in either the user manuals or the customer reports, since the corpus was approached as a whole without establishing sub-corpora. In practice, however, the majority of words (53.7% out of the 270 words included in the ultimate word list) occurred in both the user manuals and the customer reports. Essentially, then, the criterion of range was applied nonetheless.

For specialised occurrence, this thesis employs the same definition as Coxhead (2000) in saying that the word families that were ultimately included in the PEWL had to be outside the first 2,000 items in the GSL. In other words, the most frequently occurring words of English were expected to be common knowledge in the FSEs language repertoire. The word families that occurred in Coxhead's (2000) AWL were not excluded, however, because the target group is not expected to be familiar with academic vocabulary given their vocational education (Humphrey, 2019). Idioms and set expressions were ultimately excluded from the word list due to the technical nature of certain compounds and the great variability of compounds and set expressions.

3.5.2 Unit of analysis

One of the recommendations by Nation and Waring (2000) and Coxhead (2000) for creating a frequency-based word list is to specify what counts as a word. This was also one of the main criticisms on the GSL, since it is inconsistent in establishing what constitutes a word (Gilner, 2011). In terms of receptive vocabulary knowledge, Nation and Webb (2011) argue that the word family is the most sensible unit of analysis. A word family represents a group of related words which can be inferred as soon as the base form of this word family is known to the language learner (Hajiyeva, 2015). However, Ward (2009) shows that less proficient learners of English from a non-Latinate background, such as the Thai students in his study, are at a disadvantage due to the fact that their first language does not have inflection. Since 10% of his corpus consists of words of Latinate origin, he therefore opts for word types as the counting unit of his word list. Another reason for this choice is because word families, with their inflected and derived forms, have an additional learning load, which is not desired for novice language learners. Despite this, as Coxhead (2000) points out, word families are shown as an important unit in the mental lexicon (p. 218). Prior research also shows that learners do not experience inflected or derived forms as more effortful as soon as they possess essential word-building skills (Bauer & Nation, 1993, cited in Coxhead, 2000). The target group in the present study should therefore be able to undertake the morphological analysis that is implicit in the notion of the word family due to their moderate-to-high proficiency levels in English. Should this prove to be false, FSEs can be coached to observe a number of affixes and may even already possess the skill of recognising different members of the same word family, similar to the English majors in Hajiyeva's (2015) study. It was therefore decided to organise the PEWL by word families. Table 6 presents several sample word families that occur in the PEWL. Appendix III provides an alphabetised overview of all word families in the PEWL.

Table 6. Sample word families from the Professional Engineering Word List.[1]

adjust	**activate**	***gear***	**synchronize**
adjustment	activated	gearbelt	synchronization
adjustments	activates	gearbox	synchronized
adjusted	activating	gearboxes	*synchronizer*
adjusting	activation	geared	synchronizing
adjustable	*deactivated*	gearing	synchronous
adjuster		gearwheels	

3.6 Qualitative analysis

Following Chung and Nation's (2004) method of combining approaches to identify the degree of technicalness in the PEWL's items, this study triangulated all four of the recommended methods. Additionally, a final check with two experts in the field was conducted.

In the first place, The computer-based approach proposed by Chung and Nation (2004) matches the vocabulary profile, as the off-list vocabulary items which are not found in the reference word lists constitute those words that are not covered by the GSL or the AWL. Next, all vocabulary items in the PEWL were coded by the researcher based on Nation's (2001) categorisation of technical vocabulary. Thirdly, as recommended by Chung and Nation (2004), the definitions of all words in the word list were retrieved via the Cambridge Dictionary and

[1] Words in italics refer to the most frequent form in that word family occurring in the Professional Engineering Word List

TecDic (TecDic, 2020). TecDic is a two-way English-Dutch technical dictionary which FSEs at Marel often frequent. In the present study, TecDic was used to affirm and monitor the Cambridge Dictionary definitions to verify or falsify whether any words in the PEWL had a specialised occurrence. The third step consisted of comparing the concordance lines that were generated via AntConc. The last stage of this qualitative analysis consisted of a final check of any definitions that were still obscure to the author. Any word with apparent technical definitions that were not understood by the author were clarified by a senior FSE. In doubtful cases, the coding was checked by one FSE who provided explanation on the technical nature of some vocabulary items.

3.7 Ethical considerations

As is argued by McEnery et al. (2006), it should not be overlooked that certain texts in the corpus may be copyrighted and therefore cannot be distributed or published in any form. This is also the case for the texts in the PEEC, which were retrieved via Marel's Technical Support Manager and a FSE. For this reason, the corpus cannot be disclosed. Depending on the context of the corpus, it may be necessary to obtain participant consent, anonymise data or explain the way in which the data was stored in compliance with the GDPR[2]. Regarding the PEEC, the largest part of the data need not be accounted for in terms of ethical considerations, given that the user manuals do not contain any personal details or participant data. The customer reports, however, are written by FSEs themselves. To ensure the FSEs' privacy, all personal details were excluded from the corpus. This includes FSEs' names, customer names and contact details, and any other data that can be traced back to the FSEs or Marel's customers. Furthermore, all data was obtained via Marel Boxmeer's Technical Support Manager, who informed all FSEs of the purpose of this study and provided the opportunity for the engineers to make mention of any possible objections with regards to the distribution of their written work. No objections were risen, provided that the data collection and analysis occurred under the terms and conditions of the GDPR.

[2] General Data Protection Regulation: regulations pertaining to privacy and data as postulated by the European Union.

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Introduction

This chapter will reveal the quantitative and qualitative results of the analysis of the Professional Engineering English Corpus (PEEC). It sets out to describe the vocabulary profile of the corpus and addresses the issue of corpus representativeness in the form of a saturation test. This is followed by an explanation on which theoretical and practical decisions were made in the development of the Professional Engineering Word List (PEWL). Next, the qualitative analysis of the nature of words in the PEWL will be addressed. The PEWL will then be compared to other word lists, both general and academic as well as discipline-specific. The present chapter will discuss the results of the corpus analysis by referring to the research questions that were presented in Chapter 1. The results will be interpreted by means of findings from prior studies in the field of ESP corpus research and (engineering) word lists. Provided that the PEWL itself, as well as its data on frequency and keyness, is the most crucial result of this study, the results will be coupled with a discussion.

4.2 Vocabulary profile

A vocabulary profile of the PEEC was developed to provide an answer to the first research question which focused on the frequency and distribution of specialised vocabulary in the PEEC. The vocabulary profile of the PEEC can be consulted in Table 7.

Table 7. The vocabulary profile of the Professional Engineering English Corpus.

Level	**PEEC**		
	Words	**%**	**Cum %**
1 (1st 1,000 GSL)	46,243	65.6	65.6
2 (2nd 1,000 GSL)	7,722	11.0	76.6
3 (AWL)	6,326	9.0	87.6
0 (off-list)	10,245	14.5	100.0
Total	**70,536**	**100.0**	**100.0**

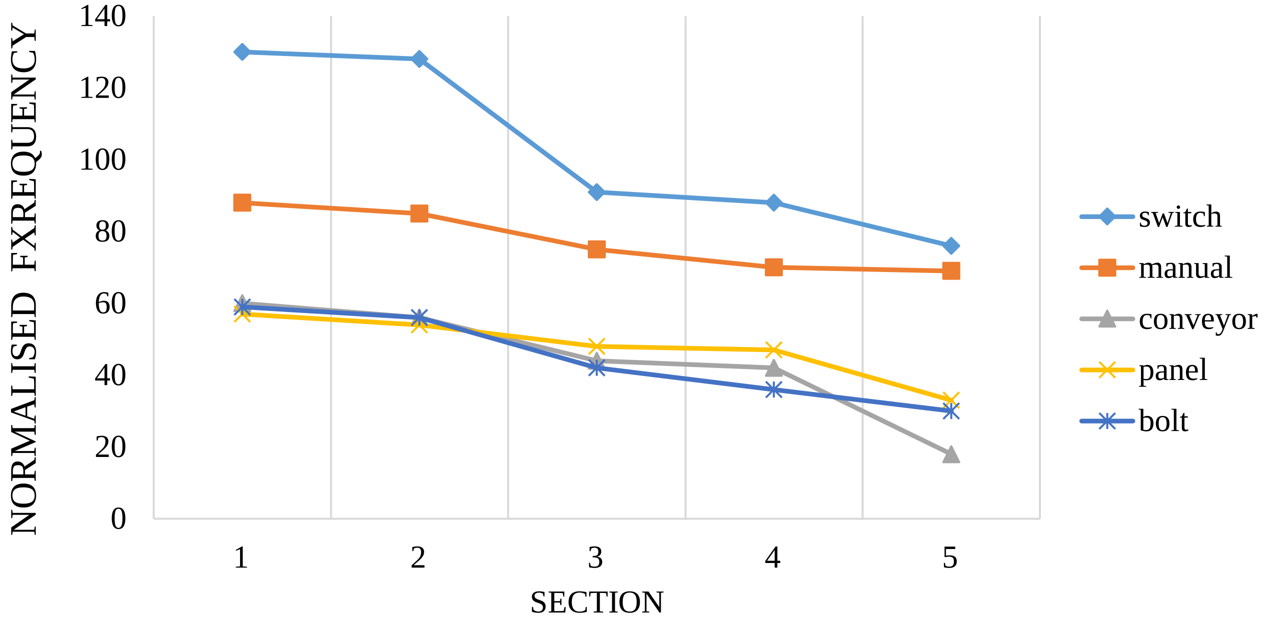
From this table, it becomes evident that 76.6% of the corpus contains words from the GSL, whereas the AWL words make up 9% of the corpus. The remaining 14.5% are off-list words that are specific to the current genre. Vice versa, to determine whether the PEWL is truly representative for the PEEC, AntWordProfiler was used to run the frequency list over the PEEC, similarly to what was done with the GSL, AWL, and other engineering-specific word lists. The results reveal that the PEWL covers 21.6% of the PEEC. Provided that a large amount of vocabulary in the PEEC, 76.6% to be precise, is included in the 2,000 most frequent words in the English language in the form of the GSL, the 21.5% figure is a respectable coverage figure. The comparison figures in Table 7 confirm that the PEWL is not a further general service list, nor an academic word list or a composite list to add to existing engineering word lists.

When comparing the vocabulary profile to that of Ward's (2009) BEL, the results are largely consistent. 188 words in the BEL, which consists of 299 word types, occur in the first 1,000 words of the GSL. This translates to a figure of nearly 63%, which is roughly similar to the 66% coverage of the GSL over the PEEC. The AWL gives 11.3% coverage of Ward's (2009) corpus, whereas the AWL gives 9% coverage of the PEEC. It must however be noted that the AWL consists of nearly twenty times as many word types as the BEL. The PEEC contains more off-list, and therewith technical, vocabulary, however. Although Ward (2009) does not provide any numbers for off-list vocabulary, it can be deduced from his frequency figures that the PEEC consists of more off-list, and therefore highly technical, vocabulary than the BEL. In sum, to provide an answer to the first research question, the percentages in Table 7 display the frequency of specialised vocabulary in PEEC.

4.3 Saturation test

One of the two core objectives of this study was to compile a corpus which is as representative of technical writing in the target domain. To justify the representativeness of the PEEC, a saturation test was executed conform Shams et al.'s (2012) study. Figure 1 shows a plot of the normalised frequency of a random sample of five of the most frequent words in the PEEC.

Figure 1. Normalised frequency figures of the PEEC for a sample of five of its most frequent words across 5 proportionate sections.



A random sample of the five most frequent words in the PEEC shows that the occurrence of these technical terms becomes stationary after a few sections, notably so after section 3. In other words, no matter how many words are added to the corpus, the normalised frequency remained stable. As is pointed out by Shams et al. (2012), this is one of the criteria of saturation which confirm the representativeness of the corpus.

Although the name of the corpus would suggest otherwise, this corpus cannot be representative for professional Engineering English as a whole, however. The PEEC does not consider all subdisciplines of engineering nor does it include all instances of occupational Engineering English. The PEEC is, however, representative of what engineers at Marel can expect to encounter in terms of Engineering English. As became evident from the needs analysis, nearly 70 per cent of all FSEs reported having to write customer reports either 2 to 3 times per week or even daily (Humphrey, 2019). Therefore, these customer reports are expected to be a good predictor of language use by FSEs. Roughly 63% of FSEs also indicated that they provide instructions, explanations, and demonstrations, usually via user manuals, either 2 to 3 times per week or daily (Humphrey, 2019). This naturally implies that they first read these manuals before being able to relay information in these manuals. Taken together, it can be argued that both the user manuals and customer reports are representative samples of language which FSEs are expected to come across. The representation of the texts in the corpus was also validated by Marel's Technical Support Manager and FSEs themselves.

4.4 The Professional Engineering Word List

After the editing stages, the definitive word list was analysed quantitatively for word frequency to answer the second research question and attain the overarching research objective of developing a word list for the target group at focus. The following sections will synthesise the findings of these analyses.

The final word frequency list in AntWordProfiler revealed several striking, yet anticipated, findings. Whilst the final version of the PEWL excludes items from the AWL and the GSL, this means that semi-technical and sub-technical words are also absent in the word list. Since this thesis aims to ascertain, among other things, what the nature of the vocabulary in the PEWL actually is, this is especially troublesome. To circumvent this, the frequency word list output file which *does not* exclude items from the AWL and GSL was analysed together with one FSE to filter out semi-technical and sub-technical words which take on both a general and specialised meaning.

To provide an example, the concordance lines for the word *roller*, a word that was initially covered by the GSL but ultimately included in the PEWL, is tabulated in Table 8. In the first 1,000 words of the GSL, the word *roller* is included in the word family *roll*. To *roll*, in its general meaning, means "to cause (something to) move somewhere by turning over and over or from side to side" (Cambridge Dictionary, n.d.-a). In its specialised usage in an engineering context, however, a *roller* is "a tube-shaped object in a machine that turns over and over in order to carry things along or press them down or together" (Cambridge Dictionary, n.d.-b). Table 8 shows several concordance lines of *rollers* in the PEEC. Additional concordance lines of a selection of semi-technical and sub-technical words in the PEWL that also carry a general meaning and are covered by the GSL can be found in Appendix IV.

Table 8. Concordance lines for the word *roller*.

1	a technical defect, for example,	**rollers**	1 and detection plate 2 go up.
2	guiding block due to seized	**rollers**	and washers used on the
3	urgent replacing, also the bearing	**rollers**	in Duplex block for transport
4	rollers (wear). Replace the cam	**rollers**	if necessary. Full machine: Clean
5	that the position of the	**roller**	complies with symbol 4 in adjacent

From this screening round, a total of 31 additional words were added to the PEWL, which ultimately consists of 293 words. To provide an insight into the composition of the PEWL, the top 20 items of the word list are presented in Table 9. The full PEWL can be consulted in Appendix V. The results of the keyness analysis for the first twenty items are represented in the third column in Table 9. From these keyness values, it becomes apparent that all the words in the PEEC are overrepresented when compared to the GSL and AWL, which can be expected due to the technical nature of the PEWL. It should be noted that the keyness figures could not be computer for all items in the PEWL since the words were manually lemmatised and could not be grouped together on the basis of these word families in AntConc.

Table 9. The top 20 most frequent words in the Professional Engineering Word List.

RANK	WORD	FREQUENCY	KEYNESS[3]
1	adjust	570	+69.07
2	switch	508	+77.12
3	line	386	+94.46
4	manual	384	+93.89
5	blade	375	+37.87
6	setting	294	+37.59
7	maintenance	292	+74.69
8	module	260	+71.93
9	height	258	+74.23
10	install	254	+26.44
11	panel	239	+45.61
12	process	235	+22.08
13	instruct	224	+23.69
14	bolt	223	+32.19
15	chain	223	+48.66
16	remove	220	+37.87
17	conveyor	218	+62.13
18	shackle	218	+50.32
19	chapter	196	+47.83
20	carrier	194	+46.01

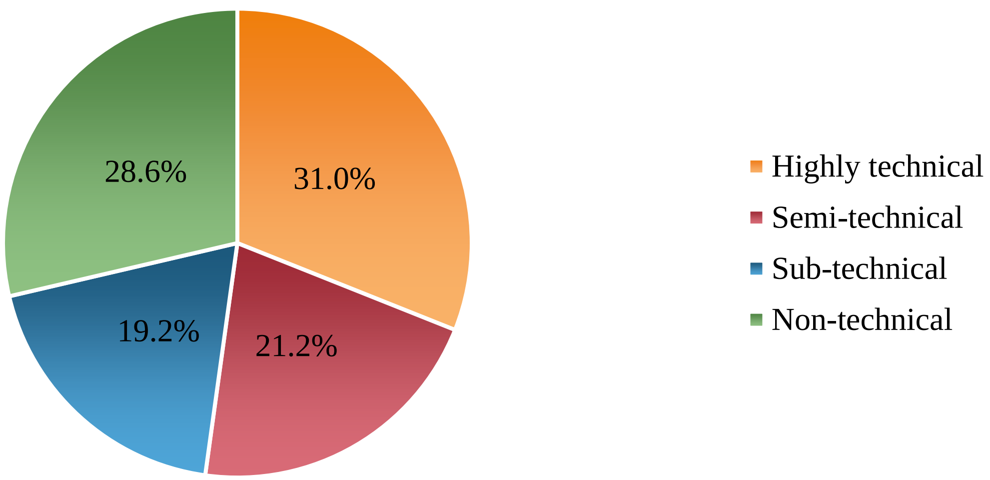
[3] Keyness based on headword only, not word family.

4.5 Word list analysis

A qualitative analysis was conducted on the definitive PEWL to determine the nature of the vocabulary in terms of Nation's (2001) classification of technical vocabulary. This analysis, then, provides an answer to the second research question which intended to uncover whether items in the PEWL are highly technical, semi-technical, sub-technical or non-technical. This analysis was conducted in three stages.

To determine what the nature of the most frequent words in the PEWL is, it is firstly important to establish whether the PEWL is truly an *engineering* word list. It is therefore especially relevant to consider the comparative statistics of the PEWL and other word lists. As is shown in Table 6, 14.5% of the words in the PEEC do not occur in the GSL or AWL. 76.6% of the words in the PEEC occur in the GSL and 9% of the words in the PEEC occur in the AWL. Since the items in the AWL were ultimately included in the PEWL, this means that overall, 23.4% of the words in the PEEC are either technical, sub-technical, or semi-technical in nature. Secondly, when considering the nature of the PEWL itself, the word list was analysed qualitatively according to a triangulation of Chung and Nation's (2004) methodology. Thirdly, the results of the manual coding were checked with one FSE. A complete overview of the distributions of the word classifications according to Nation's (2001) division can be found in Appendix VI. The percentages are tabulated in Figure 2.

Figure 2. Distribution of technical words in the PEWL according to Nation's (2001) classification of technical vocabulary across the PEWL



As becomes evident from Figure 2, the majority of words in the PEWL (31.1%) can be categorised as highly technical in nature. Non-technical words make up the second-largest category of words, with 28.6%, followed by sub-technical (21.2%) and semi-technical (19.2%) words. Altogether, 71.5% of the words in the PEWL are technical in nature, albeit in varying degrees. These findings contrast those of Mudraya (2006), who illustrated that the most frequent words in the SEEC are generally sub-technical in nature. She also provided evidence for the use of those sub-technical words being used in a non-technical sense more often than in a technical sense. The findings are also largely inconsistent with Ward's (2009) study, whose BEL is predominantly non-technical in nature. His word list shows that highly technical words are markedly underrepresented. It should be noted that both Mudraya (2006) and Ward (2009) based their word list on a corpus of undergraduate engineering textbooks. Ward (2009) also explains that the BEL is aimed at students who are not yet familiar with the engineering discipline and therefore challenged by technical terminology. The present study resides in an EOP context, which may explain the fact that more than 70 per cent of the words are technical (i.e. highly technical, sub-technical or semi-technical) in nature. Since the FSEs have had vocational engineering training, they are expected to be familiar with basic engineering terminology. To provide an answer to the second research question, it can be concluded that the PEWL is (highly) technical in nature and altogether a truly discipline-specific word list.

4.6 Word list coverage

To provide an answer to the third research question concerning how vocabulary items in the PEEC relate to those in other word lists, the PEEC was cross-checked with multiple reference word lists, both general, academic, and engineering-specific, to determine the word lists' coverage of the user manuals and customer reports across these word lists. More specifically, by comparing the PEEC to the GSL and the AWL as reference word lists, possible insights into the characteristics of this type of technical vocabulary can be generated. The results of this analysis are tabulated in Table 10.

Table 10. Coverage of West's (1953) GSL, Coxhead's (2000) AWL, Mudraya's (2006) SEWL, and Ward's (2009) BEL over the Professional Engineering English Corpus (%).

Word list	Coverage of the PEEC
West's (1953) GSL	76.5
Coxhead's (2000) AWL	9.0
Mudraya's (2006) SEWL[4]	8.4
Ward's (2009) BEL	11.6
PEWL	21.6

The coverage figures of the GSL and AWL have been discussed in light of the vocabulary profile of the PEEC and can also be reviewed in Table 6. A comparison of the PEWL to other Engineering English word lists reveals that, in terms of coverage, the PEWL is unlike other discipline-specific word lists. The word list coverage analysis shows that the PEWL indeed caters for a specific target group in a specific discipline. Interestingly, the three word lists which relate to Engineering English in an EAP context (Mudraya, 2006; Ward, 2009) provide higher coverage figures than the study which was conducted in a vocational setting and can incorporated more professional and occupational corpus texts (Ng et al., 2013). It can therefore be concluded that the PEWL is a stand-alone resource that is unique to its discipline and has little similarities with other word lists in the broader domain of Engineering English. However, it must be pointed out that comparing word lists which each use different units of analysis is by default less transparent and valid (Schmitt et al., 2017). Ward (2009) used word types as counting units and therefore has a larger word list than if word families had been used as a counting unit. Any conclusions from the coverage analysis on these word lists must therefore be drawn with caution. In light of the second objective of this study, which is to construct a word list that is sufficiently distinct from other Engineering English word list, it can nevertheless be concluded that the PEWL is indeed is unique to its discipline and bears little similarities with other word lists. It can therefore be concluded, in an answer to the third research question, that other word lists yield only little coverage over the PEWL. The GSL, however, does yield a large coverage figure, as is commonplace with the coverage figures of the GSL over any given text (Nation & Waring, 2000).

[4] Only the first 100 most frequent word families in the SEWL were available to check.

CHAPTER 5: CONCLUSION

5.1 Conclusion

For a non-native speaker of English working in an international environment, engineering is a complex field to communicate in professionally (Ng et al., 2013). This has also proven to be true for the target group in the present study: starting Field Service Engineers (FSEs) at Marel Boxmeer. As has become clear from the needs analysis, starting FSEs often do not have engineering-specific vocabulary at their disposal, having only had general English language courses in their higher vocational studies (Humphrey, 2019). Intending to remedy this in a potential future ESP course, the objective of this study was to construct an Engineering word list that is representative of technical writing in the target domain as well as sufficiently distinct from existing general service, academic, and discipline-specific word lists. A review of relevant prior literature has revealed that the lexical approach of vocabulary learning and teaching in an ESP context has become more dominant (Nation & Waring, 2000; Mudraya, 2006; Ward, 2009; Khani & Tazik, 2013). Whilst a plethora of studies have focused on Engineering English in an EAP context, there is a lacuna in the field of studies which approach Engineering English in an EOP context. With the PEWL, this study has provided the initiative and means for developing an ESP language course which can be referenced in future endeavours to cater for the target group in this context, as engineering dictionaries alone do not cover all the terms that are found and used in specific disciplines (Shamsudin et al., 2013). This word list, the Professional Engineering Word List (PEWL) consisting of 293 word families was developed from the Professional Engineering English Corpus (PEEC) of 70,536 tokens. The corpus consisted of user manuals and customer reports, which are two of the most common sources of Engineering English in FSEs' daily tasks according to two professionals in the field. The resulting word list is intended to function as the foundation for a lexical syllabus for use in an Engineering English course at the Marel Academy.

In pursuit of achieving this overarching research purpose, three research questions were addressed. To provide an answer to the first research question on the frequency of specialised vocabulary in the PEEC, a vocabulary profile of the corpus was compiled. The vocabulary profile of the PEEC revealed that 14.5% of the tokens did not occur in the GSL or AWL. Furthermore, the 293-item PEWL covers 21.5% of the PEEC. When considering the second research question, which refers to the nature of the most frequent vocabulary items in the PEEC, several things can be concluded. Firstly, the majority of the words in the PEWL are highly technical in nature. The second-most frequent category of words is that of non-technical words, despite more than 70 per cent of the words in the PEWL being technical in nature, albeit in varying degrees. These findings were shown to be inconsistent with prior studies in Engineering English (Mudraya, 2006; Ward, 2009), which shows that the PEWL is indeed sufficiently distinct from other disciplines. To answer the third research question, the PEEC was related to other general, academic, and domain-specific word lists. A comparison of the coverage figures of these word lists (the AWL, GSL, SEWL, and BEL) over of the PEEC reveals that 76.5% of the words in the corpus occur in the GSL and 9.0% occurs in the AWL. However, a qualitative analysis of concordance lines revealed that several of the words which are in the GSL had specialist usages in an engineering context. This was also confirmed by two specialists in the field. Consequently, several items that initially occurred in the GSL were included in the PEWL due to their specialised meanings. When comparing the coverage figures of other Engineering English word lists over the PEEC, it becomes evident that the PEWL functions as a unique stand-alone resource that caters for a specific target group in a specific domain of Engineering English. In that sense, it can be argued that the variety of Engineering English used in the occupational domain is vastly different from the variety of Engineering English used in an EAP context, which most studies in the field have focused on (Mudraya, 2006; Ward, 2009; Ng et al., 2013, Watson Todd, 2017).

In sum, the present study has shown that the research objective, which was to construct an EOP Engineering word list, has been realised. As has been pointed out, the PEWL is both representative of technical writing in the target domain as well as sufficiently distinct from existing word lists. Whilst it cannot be established whether the word list is indeed beneficial in terms of learning gains, yet two specialists in the field have confirmed the relevance of the PEWL and consider it to be a valuable addition to a future Engineering English course in the Marel Academy. Despite the trend of catering for heterogeneous groups of language learners and courses becoming less specific and more generalised and diffuse which is considered to be undesirable according to Hyland (2002). The present study has shown that a large percentage of words is (highly) technical in the PEWL. As a homogenous target group, FSEs benefit from what Basturkmen (2003) calls narrow-angled, specific language courses with materials targeted at authentic and relevant language use in their field. A potential ESP course should bear this in mind to make sure the course is not only perceived as more relatable and motivating yet also yield higher learning returns. Altogether, as Hyland (2002) urges, “putting the S back in ESP” is especially crucial to strive for in light of the target group in this study (p. 391).

5.2 Limitations

As is commonplace in corpus studies (but also in academic research in general) several compromises were made in the development of the corpus. Firstly, in terms of absolute size, the PEEC can be considered as small (Flowerdew, 2002). Whilst a larger corpus would increase the validity of the findings of this study, the size of the PEEC allows for qualitative analysis of the technicalness of vocabulary, which is especially relevant for the purpose of this thesis. After all, specialised corpora which have been developed with a pedagogical purpose in mind are likely to provide returns that are relevant in teaching in an ESP context (Flowerdew, 2002; Tribble; 2002). From a practical point of view, a corpus such as the PEEC which is developed by a future ESP teacher will necessarily be small and specialised by virtue of the limited availability of data (Koester, 2010). Lastly, the PEEC, though modest, has been shown to be robust and saturated by the saturation test (cf. Shams et al, 2012). Whilst it would have been possible to collect more user manuals and customer reports, both Marel's Technical Support Manager as well as an FSE who were consulted doubted whether this would yield any additional insights for this specific target group. As Bowker and Pearson (2002) argue, a small yet well-designed corpus may ultimately provide far more useful information than a large corpus which is ill-accustomed to the researcher's needs. When considering the ultimate goal of the PEEC and subsequently the PEWL, the practical considerations of the corpus outweigh its generalizability given that its focus is on being of use for a very specific target group.

Another point of criticism of the PEEC constitutes the decision of merely including two text types: customer reports and user manuals. This extends to the limitation that only written work was ultimately included in the corpus. Other corpus texts that could have been added to the PEEC are those events that, in the needs analysis, FSEs reported requiring substantial English skills for. Examples of such events constitute transcribed meetings, teleconference calls and presentations (Humphrey, 2019). If more text types that frequently feature in the target group's activities had been included, the corpus would potentially be more representative (Bennett, 2010; Koester, 2010). Although the saturation test revealed that the PEEC was saturated, this finding cannot be extended to other modalities (i.e. spoken text) and text types that were ultimately not included. Despite this, two experts in the field confirmed that the document types were largely representative of the type of language that FSE ought to be familiar with.

When considering the PEWL itself, several limitations can be distinguished. Firstly, the size of the PEWL may be contested. By virtue of setting the frequency threshold at a minimum of ten occurrences across the entire the corpus, the PEWL consists of 293 items. If this threshold had been set higher, the word list would have been considerably smaller. Nonetheless, the size of the word list itself is relative and for the purpose of the lexical syllabus, all vocabulary gain is of added value for the target group. When comparing the size of the PEWL to existing Engineering English word lists (Mudraya, 2006; Ward, 2009), the number of items is largely consistent. Secondly, it was decided not to consider multi-word units or idioms in the scope of this study, whilst this could have revealed additional interesting results and insights for the lexical syllabus. Lastly, like Coxhead (2000), the PEWL does not consider items that were already in the GSL, except those that carry specialised meaning and can therefore be characterised as sub-technical or semi-technical. Arguably, some FSEs may not be familiar with all the items in the GSL. They have had higher vocational general English, however, and can be expected to perform at a B2 level at minimum. It should also be noted that specialised knowledge is not static. Due to new developments and insights, new specialised terminology is coined all the time. It may therefore be the case that the PEWL does not prove to stand the test of time.

Despite these limitations, this study has reported on a self-compiled corpus of respectable size and painted a comprehensive picture of a discipline-specific word list for use in a lexical syllabus for an Engineering English ESP course.

5.3 Implications

There are several implications of the present study. In terms of pedagogical implications, as Ng et al. (2013) point out, ESP teachers may base themselves on intuitions and refrain from developing materials holistically and thoroughly if they do not have a discipline-specific word list at their disposal. This, then, can lead to an ESP course with words that may either be too general or at least ill-accustomed to the target group's needs. When bearing the specific target group in mind, it can safely be concluded that the PEWL consists of relevant and domain-specific words that will be of use to FSEs in their daily tasks. Although the specific interpretation is yet to be determined, the PEWL is ultimately intended for pedagogical purposes: in an ESP course for FSEs in their training time. Similar to the BEL in Ward's (2009) study, the PEWL can in principle be taught by teachers regardless if they are familiar with the field of engineering. As a component of the Marel Academy, the PEWL may function as a resource in an ESP course. Crucially, the PEWL ought to be included in the course in various manners and contexts, as prior studies demonstrate that vocabulary retention is aided by multiple exposures (Ward, 2009; Thornbury, 2002). In line with the lexical approach, integrating the PEWL in a lexical syllabus holistically may yield significant learning returns for the present target group. In terms of research implications, the present study adds to the growing body of research on corpus linguistics and ESP in general. When combining the present findings with studies on vocabulary acquisition and learning, additional insights on effective vocabulary teaching may be revealed. As Ng et al. (2013) show, specialized word lists guide learners in their vocabulary acquisition. Creative vocabulary-focused activities may aid in encouraging vocabulary acquisition of these learners (Thornbury, 2002). All in all, the PEWL should ensure that valuable time in an eventual ESP course spent on vocabulary teaching reaps the greatest benefit and saves both financial resources and effort in terms of vocabulary selection or teaching.

5.4 Suggestions for further research

There is a plethora of ESP areas that are explored only to a certain extent or not explored at all. Whilst Engineering English in and of itself is not one of those underrepresented domains, there is a relative paucity of EOP studies due to the fact that most researchers are involved in an academic institution. Future studies may do well to fill in gaps in existing ESP research and apply more focus on the EOP domain in general, but more specifically in Engineering English, to take up the call for more local and discipline-specific ESP research (Hyland & Tse, 2007). As the present study has shown, the PEWL differs vastly from other Engineering English lists in terms of frequency, distribution, and nature. It can therefore be concluded, at least on the basis of this study, that Engineering English in an EOP context cannot be equated to Engineering English in an EAP context. Future research may also consider conducting a meta-analysis of ESP, and more specifically EOP, studies to offer an insightful overview of which domains still require additional attention in research. A similar synthesis may be conducted to establish a systematic framework for the compilation and evaluation of word lists in general. This framework can serve as a reference for future studies on word lists.

Lastly, when considering the scope of this thesis, it would be interesting to provide further explanatory power to the present study and assess the effectiveness of the word list in an intervention study. This study may be conducted within the Marel Academy, since multiple groups of FSEs start per year. It would therefore be possible to conduct a longitudinal study on these target groups to assess vocabulary retention and determine the effectiveness of teaching an ESP course with a lexical syllabus based on a word list. When multiple parallel groups of comparably experienced and linguistically proficient FSEs are being trained in Engineering English, it would also be interesting to consider which vocabulary training works best (i.e. focus on form, or meaning). In sum, the Marel Academy may function as a fruitful basis for experimental studies on vocabulary learning and ESP research.

REFERENCES

Anthony, L. (2014). *AntWordProfiler* (Version 1.4.10 [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Anthony, L. (2015). The changing role and importance of ESP in Asia. *English as a Global Language Educational Journal*, *1*(1), 1–21. https://doi.org/10.6294/EaGLE.2015.0101.01

Anthony, L. (2019). *AntConc* (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.

Basturkmen, H. (2003). Specificity and ESP course design. *RELC Journal*, *34*, 48-63. https://doi.org/10.1177%2F003368820303400104

Belcher, D. (2009). What ESP is and can be: An introduction. In D. Belcher (Ed.), *English for specific purposes in theory and practice* (pp. 1–20). Ann Arbor, MI: University of Michigan Press. https://doi.org/10.3998/mpub.770237

Bennett, G. R. (2010). An introduction to corpus linguistics. In *Using corpora in the language learning classroom* (pp. 1–22). Ann Arbor, MI: University of Michigan Press. https://doi.org/10.3998/mpub.371534

Bowker, L., & Pearson, J. (2002). *Working with specialized language: A practical guide to using corpora*. London: Routledge. https://doi.org/10.4324/9780203469255

Browne, C. (2013). The New General Service List: Celebrating 60 years of Vocabulary Learning. *JALT Publications: The Language Teacher*, 13-16.

Chung, T. M., & Nation, I. S. P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language*, *15*(2), 103–116. Retrieved September 9, 2020, from http://nflrc.hawaii.edu/rfl/

Chung, T. M., & Nation, I. S. P. (2004). Identifying technical vocabulary. *System*, *32*, 251-263. https://doi:10.1016/j.system.2003.11.008

Cobb, T., & Horst, M. (2004). Is there room for an AWL in French? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 15–38). Amsterdam: John Benjamins.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238. https://doi.org/10.2307/3587951

Coxhead, A., & Hirsch, D. (2007). A pilot science-specific word list. *Revue Française de Linguistique Appliquée*, *XII*(2), 65–78. https://doi.org/10.3917/rfla.122.0065

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, *28*(3), 157–169. https://doi.org/10.1016/j.esp.2009.02.002

Engels, L. K. (1986). The fallacy of word counts. *International Review of Applied Linguistics in Language Teaching*, *6*, 213–231.https://doi.or/https://doi.org/10.1515/iral.1968.6.1-4.213

Flowerdew, J., & Peacock, M. (2001). Issues in EAP: A preliminary perspective. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 8–24). Cambridge: Cambridge University Press.

Flowerdew, L. (2002). Corpus-based analyses in EAP. In J. Flowerdew (Ed.), *Academic Discourse* (pp. 95–114). London: Longman.

Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes*, *24*(3), 321–332. https://doi.org/10.1016/j.esp.2004.09.002

Gilner, L. (2011). A primer on the general service list. *Reading in a Foreign Language*, *23*(1), 65–83. Retrieved July 10, 2020, from http://nflrc.hawaii.edu/rfl/April2011/articles/gilner.pdf

Hajiyeva, K. (2015). A corpus-based lexical analysis of subject-specific university textbooks for English majors. *Ampersand*, *2*, 136–144. https://doi.org/10.1016/j.amper.2015.10.001

Humphrey, A. J. (2019). *Engineering English for field service engineers at a poultry processing company in the Netherlands: A needs analysis* [Unpublished needs analysis]. Faculty of Arts, Radboud University Nijmegen.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

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

Ng, Y.J., Lee, Y.L., Chong, S.T., Sahiddan, N., Philip, A., Azmi, N., & Tarmizi, M. (2013). Development of the engineering technology word list for vocational schools in Malaysia. *International Education Research*, *1*(1), 43–59. https://doi.org/10.12735/ier.v1i1p43

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

Kırkgöz, Y., & Dikilitaş, K. (2018). Recent developments in ESP/EAP/EMI contexts. In Y. Kırkgöz & K. Dikilitas (Eds.), *Key issues in English for specific purposes in higher* *education* (English Language Education Vol. 11, pp. 1–10). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-70214-8

Koester, A. (2010). Building small specialised corpora. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 66–79). London: Routledge. https://doi.org/10.4324/9780203856949.ch6

Lewis, M. (1993). *The lexical approach*. Hove: Language Teaching Publications.

Marel. (n.d.). About Marel. Retrieved November 1, 2019, from https://marel.com/about/

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

McEnery, T., & Xiao, R. (2011). What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 364–380). New York: Routledge. https://doi.org/https://doi.org/10.4324/9780203836507

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. New York: Routledge. Retrieved on June 17, 2020, from https://www.lancaster.ac.uk/fass/projects/corpus/ZJU/xCBLS/CBLS.htm

Mudraya, O. (2001). Lexical approach to second language teaching: A new role for lexis. *ERIC Digest.* 1-8. Retrieved June 3, 2020, from https://files.eric.ed.gov/fulltext/ED455698.pdf

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes*, *25*(2), 235–256. https://doi.org/10.1016/j.esp.2005.05.002

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language* Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9781139858656.008

Nation, I. S. P. (2017). The BNC/COCA level 6 word family lists (Version 1.0.0) [Data file]. Victoria University of Wellington. Available from http://www.victoria.ac.nz/lals/staff/paul-nation.aspx

Nation, I. S. P. (2018). Keeping it practical and keeping it simple. *Language Teaching*, *51*(1), 138–146. https://doi.org/10.1017/S0261444817000349

Nation, I. S. P., & Waring, R. (2000). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: description, acquisition and pedagogy* (pp. 6–19). Cambridge: Cambridge University Press.

Nation, I. S. P., & Webb, S. A. (2011). *Researching and analyzing vocabulary*. Heinle: Cengage Learning.

Nelson, M. (2015). Building a written corpus. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 53-65). London: Routledge. https://doi.org/10.4324/9780203856949.ch5

Nesi, H. (2013). ESP and corpus studies. In B. Paltridge & S. Starfield (Eds.), *The handbook for English for specific purposes* (pp. 407–426). Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781118339855

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.

Paltridge, B., & Starfield, S. (2013). *The Handbook of English for Specific Purposes* (Blackwell Handbooks in Linguistics). Chichester, UK: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118339855

Richards, J. (1974). Word lists: problems and prospects. *RELC Journal*, *5*(2), 69–84.

Roll. (n.d.-a). In *Cambridge Dictionary* Retrieved September 1, 2020, from https://dictionary.cambridge.org/dictionary/english/roll

Roller. (n.d.-b). In *Cambridge Dictionary*. Retrieved September 1, 2020, from https://dictionary.cambridge.org/dictionary/english/roller

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

Shams, R., Elsayed, A., & Akter, Q. M.-Z. (2010). A corpus-based evaluation of a domain-specific text to knowledge mapping prototype. *Journal of Computers*, *5*(1), 69–80. https://doi.org/10.4304/jcp.5.1.69-80

Shamsudin, S., Husin, N., & Manan, A. A. (2013). Exploring fundamental engineering word list for engineering students: A literature review. *Procedia - Social and Behavioral Sciences*, *70*, 1275–1281. https://doi.org/10.1016/j.sbspro.2013.01.187

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

Sinclair, J. (2005). Corpus and text: Basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 22–36). University of Oxford: AHDS Literature, Languages and Linguistics.

Spence, P., & Liu, G. (2013). Engineering English and the high-tech industry: A case study of an English needs analysis of process integration engineers at a semiconductor manufacturing company in Taiwan. *English for Specific Purposes* *32*(2), 97-109. https://doi.org/10.1016/j.esp.2012.11.003

Starfield, S. (2014). Current and future directions in English for Specific Purposes. *Revue Française de Linguistique Appliquée*, *19*(1), 9–14. https://doi.org/10.3917/rfla.191.0009

Thornbury, S. (2002). *How to teach vocabulary*. Essex: Pearson Education Limited.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. (pp. 84-100). Amsterdam: John Benjamins Publishing Co. https://doi.org/10.1075/scl.6

Tribble, C. (2002). Corpora and corpus analysis: New windows on academic writing. In J. Flowerdew (Ed.), *Academic Discourse* (pp. 131–149). London: Longman.

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

Watson Todd, R. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes*, *45*, 31–39. https://doi.org/10.1016/j.esp.2016.08.003

West, M. (1953). *A general service list of English words*. London: Longman, Green & Co.

Widdowson, H. G. (1998). Context, community, and authentic language. *TESOL Quarterly*, *32*(4), 705–716. https://doi.org/10.2307/3586764

APPENDIX I: PYTHON CODE (CORPUS COMPILATION)

```
import PyPDF2
import textract
import os
import re
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords

text = ""
directory = input("Typ hier de map waar de bestanden staan: ")
for filename in os.listdir(directory):
    if filename.endswith(".pdf"):
        print('Bezig met verzamelen... - ', filename)
        pdfFileObj = open(os.path.join(directory, filename),'rb')
        pdfReader = PyPDF2.PdfFileReader(pdfFileObj)
        num_pages = pdfReader.numPages
        count = 0
        while count < num_pages:
            pageObj = pdfReader.getPage(count)
            count +=1
            text += pageObj.extractText().lower()
    else:
        continue

if text != "":
   text = text
else:
   text = textract.process(fileurl, method='tesseract',
language='eng')

tokens = word_tokenize(text)
punctuations = ['(',')',';',':','[',']',',']
exclusions = ['marel', 'stork', 'warning', 'mortal', 'danger',
'handelstraat', 'av', 'boxmeer', 'netherlands', 'preface',
'appendix', 'contents', 'eng', 'figure', 'table', 'remark']
keywords = [re.sub(r'\d+', '', word) for word in tokens if
word.isalnum() and not word in punctuations and not word in
exclusions and not word.isdigit() and not(word.isalpha() and
len(word) == 1)]
outputfilename = input("Typ hier hoe het corpus bestand moet heten:
")
outputfilename += ".txt"
with open(outputfilename, 'w') as f:
    for keyword in keywords:
        f.write("%s\n" % keyword)
```

APPENDIX II: SWIFT CODE (SATURATION)

```swift
import UIKit

let crCombinedFileURL = Bundle.main.url(forResource: "CR_Combined",
withExtension: "txt")
var crCombined = try String(contentsOf: crCombinedFileURL!,
encoding: .utf8)

let umCombinedFileURL = Bundle.main.url(forResource: "UM_Combined",
withExtension: "txt")
var umCombined = try String(contentsOf: umCombinedFileURL!,
encoding: .utf8)

var combined = crCombined + umCombined

crCombined.removingRegexMatches(pattern: "[0-9].")
var crSplit = crCombined.split { $0.isWhitespace }.compactMap{
String($0) }
var crChunked = crSplit.chunked(into: Int(ceil(Float(crSplit.count /
5))))

umCombined.removingRegexMatches(pattern: "[0-9].")
var umSplit = umCombined.split { $0.isWhitespace }.compactMap{
String($0) }
var umChunked = umSplit.chunked(into: Int(ceil(Float(umSplit.count /
5))))

var chunkedCombined: [[String]] = []

for i in 0...4 {
    chunkedCombined.append(crChunked[i] + umChunked[i])
    for word in chunkedCombined[i] {
        print(word)
    }
    print("====================")
}

extension String {
    mutating func removingRegexMatches(pattern: String, replaceWith:
String = "") {
        do {
            let regex = try NSRegularExpression(pattern: pattern,
options: NSRegularExpression.Options.caseInsensitive)
            let range = NSMakeRange(0, self.count)
            self = regex.stringByReplacingMatches(in: self, options:
[], range: range, withTemplate: replaceWith)
        } catch {
            return
        }
    }
}
extension Array {
    func chunked(into size: Int) -> [[Element]] {
        var output = stride(from: 0, to: count, by: size).map {
```

```
            Array(self[$0 ..< Swift.min($0 + size, count)])
        }

        if output.count > 5 {
            output[4] = output[4] + output[5]
            output.remove(at: 5)
        }
        return output
    }
```

APPENDIX III: ALPHABETHICAL PEWL

abdominal
access
accurate
accurately
activate
activated
activates
activating
activation
deactivated
adaptor
adapter
adjacent
adjust
adjusted
adjuster
adjusting
adjustment
adjustments
adjustable
alarm
alarms
anatomical
anatomic
anatomically
appendix
area
authorize
authorize
authorized
authorizations
authorise
authorise
automatic
available
battery
batteries
beam
bear
bearing
bearings
behalf
belt
bin
blade
blower
bolt
bolts
bracket
brackets
breast
breastbone
breastcap
breastcaps
breasts
bridge
broiler
broilers
buffer
buffers
bushing
bushings
bypass
cable
cabled
cables
cabling
calibrate
calibrates
calibrating
calibration
cam
cap
carcass
carcasses
carrier
carriers
cartilage
cartilages
cell
cells
chain
chains
chapter
chemical
chill
chilled
chiller
chilling
chute
chutes
circuit
circumstances
clamp
clamping
clamps
clockwise
counterclockwise
clutch
clutching
code
competent
component
components
compress
compressed
compresses
compression
compressor
concept
configuration
confirmation
consist
consists
consisting
consult
contact
converter
convey
conveys
conveyed
conveyor
conveyance

- **cord**
  - cords
- **corrective**
- **corresponding**
- **counter**
- **coupling**
- **crate**
  - crates
- **curve**
  - curves
- **cutter**
  - cutters
- **cylinder**
  - cylinders
- **data**
- **debone**
  - deboner
  - deboning
- **defect**
  - defective
  - defects
- **design**
- **detect**
  - detector
  - detection
- **device**
- **diagram**
  - diagrams
- **dimension**
  - dimensions
- **disc**
  - disk
  - discs
- **discharge**
  - discharged
- **disinfect**
  - disinfecting
  - disinfection
  - disinfectant
- **display**
- **document**
  - documentation
- **drive**
  - drives
- **drumstick**
  - drumsticks
  - drum
- **eject**
  - ejects
  - ejected
  - ejection
  - ejector
  - ejectors
- **emergency**
- **ensure**
- **entrapment**
- **entry**
- **equipment**
- **error**
- **eviscerate**
  - eviscerated
  - eviscerating
  - evisceration
  - eviscerator
- **exhaust**
- **exit**
- **familiarize**
- **fillet**
  - filleter
  - filleting
- **filter**
  - filters
- **flammable**
- **flange**
- **force**
- **friction**
- **function**
  - functioning
- **fuse**
- **gauge**
- **gear**
  - gearbelt
  - gearbox
  - gearboxes
  - geared
  - gearing
  - gearwheels
- **giblet**
  - giblets
- **gizzard**
- **grade**
- **grindstone**
- **groin**
- **hazard**
  - hazards
  - hazardous
- **height**
- **hoist**
  - hosting
- **holder**
  - holders
- **hood**
- **horizontal**
  - horizontally
- **humidity**
- **hydraulic**
  - hydraulics
- **hygiene**
- **identify**
  - identification
- **impairment**
- **incision**
- **index**
- **indicate**
  - indicates
  - indicator
- **infeed**
- **injury**
  - injuries
- **input**
- **insert**
- **inspect**
  - inspection
  - inspections
- **install**
  - installation
  - installed
  - installing
- **instruct**
  - instructions
- **intermediate**
- **ionizing**

issue
issues
joint
journal
kit
kits
kneecap
kneecaps
knob
label
labels
layout
legislate
legislation
limiter
line
lines
loadcell
loadcells
loader
logbook
lubricate
lubricants
lubricating
lubrication
lubricator
maintenance
manual
manuals
mechanism
membrane
membranes
minimum
mobile
modify
modifications
modular
module
modules
mortal
mount
mounted
mounting
nut
nuts
offal
option
optional
outfeed
output
overhaul
overhauls
overhead
overload
overview
pack
package
packaging
padlock
panel
panels
paragraph
piston
pitch
pitches
plastic
plate
plates
plating
plug
plugs
pneumatic
pneumatics
pneumatical
pollution
portioning
position
positioner
poultry
preface
presser
preventive
previous
prior
procedure
procedures
proceed
process
processing
professional
protrude
protrudes
protruding
proximity
puller
pullers
pulse
purchase
purchaser
pusher
pushers
race
rack
range
rear
recover
recovery
recycle
recycling
reducer
reducer
reference
references
regulate
regulations
regulator
relay
relays
release
released
relevant
remove
removed
removing
require
required
requirements
reset
resistor
revise
revision
rod

rods
roller
rollers
rotate
rotates
rotating
rotation
rotational
safeguard
safeguarded
safeguards
scald
scalder
scalding
schedule
scraper
scrapers
section
selection
sensor
sensors
separator
serial
setting
settings
shackle
shackles
shaft
shafts
skin
skinner
deskin
deskinned
deskinner
deskinning
slack
slacken
socket
sockets
software
spacer
spacers
specific
specifications
spine
spines
splitter
squeeze
squeezed
static
statically
status
storage
strain
strainer
straining
suction
sufficient
suspension
switch
switched
switches
switching
switchbox
symbol
symbols
synchronize
synchronized
synchronizer
synchronizing
synchronous
synchronization
synchronisation
tank
technical
technological
tendon
tendons
tension
tensioned
tensioning
tensioner
tensioners
terminal
terminate
text
thermistor
thermo
thermometer
thermostat
thermostatic
thermocouple
thermal
thigh
thighs
thighbone
tilt
tilts
tilted
tilting
timing
torque
track
transfer
transferred
transmit
transmission
transport
trolley
trolleys
troubleshooting
trough
truck
ultrasonic
unauthorized
unauthorised
undo
unintentional
unintendedly
vacuum
valve
valves
vent
vertical
vertically
via
vibrate
vibrating
vibration
vibrations
visible
voltage

wagon
wagons
washer
washers
wear
worn

weighbridge
weigher
weld
welds
welded
welding

wishbone
workspace
wrench
zone
zones

APPENDIX IV: CONCORDANCE LINES OF GSL WORDS INCLUDED IN THE PEWL

1	Check the functioning. Check for	**wear**	, broken parts, smooth running of
2	guiding is showing signs of	**wear**	due to shackles in by-
3	module. Shackle guiding shows some	**wear**	from by-passing shackles, still
4	general there could be less	**wear**	if it was correctly adjusted.
5	the carrier drops due to	**wear**	on friction blocks 4. The minimum
6	bit to high causing unnecessary	**wear**	on shackles. No oil visible
7	connecting plates. Tiny bit of	**wear**	on top of the plastic

1	clutch of the loader was	**worn**	and needs revision ASAP. The
2	knives are in stock, replace	**worn**	and damaged press on / blade
3	bearings on the wagon are	**worn**	and need replacing 3. Bracket had
4	Shackle guiding getting quite	**worn**	on the top. Not urgent,
5	spring loaded rubber flap. Bracket	**worn**	out on the right side
6	if one or several are	**worn**	. Proceed as follows: 1. Remove
7	see the User’s manual 90803)	**worn**	out or broken lugs. Check

1	the tendons around the hip	**joint**	and the tendon in the
2	the units of the (Knee	**joint**	incision) have been replaced by
3	have the coror the drum	**joint**	rectify blade height. Adjust the
4	much more work. Position 9: Wing	**joint**	cutting module. No issues noticed
5	correct position for the knee	**joint**	cutter. This wheel is installed

1	in a bad state, the	**bearing**	are maintenance free (sealed).
2	Play in rod ends and	**bearings**	, both cylinders are leaking air
3	locking on the width adjustment	**bearing**	bush broken 3x width adjustment.
4	Lever shaft is broken (Welded) -	**bearing**	is worn (Still the case)
5	branch is needed (See Summary),	**bearing**	has collapsed. The Clevis fork

1	a technical defect, for example,	**rollers**	1 and detection plate 2 go up.
2	guiding block due to seized	**rollers**	and washers used on the
3	urgent replacing, also the bearing	**rollers**	in Duplex block for transport
4	rollers (wear). Replace the cam	**rollers**	if necessary. Full machine: Clean
5	that the position of the	**roller**	complies with symbol 4 in adjacent

1	the holder. Leg skin cutter	**blade**	. Adjust drive unit gear wheels
2	that you turn off the	**blade**	after you turn off the
3	the distance between the static	**blade**	and both product holders is
4	he filleting system. The rotating	**blade**	comes to a stop within
5	explained how to raise the	**blades**	in the tendon cutter.

1	spring tension of the carrier	**chain**	and tension it if necessary.
2	Crop skin removing module	**chain**	a bit loose. I tensioned
3	When you install a new	**chain**	at the rehang section, loosening

4	been replaced or repaired, the	**chain**	guides must be reset.
5	are driven by the drive	**chain**	of the machine.

1	The eye meat knife clamping	**plate**	at the bottom is missing.
2	the data of the machine	**plate**	below in correspondence about the
3	*the data on the type*	***plate***	*of the main drive. Follow*
4	od ends (Still The case) 5.	**plate**	Control: No findings
5	5 on the inside. 7. Place the	**plate**	back in the machine. 8. Tighten

1	bypass guide as follows: Slacken	**nuts**	a number of turns. Set
2	Proceed as follows: 1. Loosen lock	**nut**	6 a few turns. 2. Set a 1
3	of the vertical springs with	**nut**	4 and 5 at value respectively.
4	of line 1 (Only shackles) Use	**nuts**	to adjust the transmission wheel.
5	Vertically Set with the	**nuts**	on the sensors the distance

1	is composed of a conveyor	**belt**	system and processing units. This
2	by knife. Tension the toothed	**belt**	1 as follows: Remove the blade
3	a collecting bin or conveyor	**belt**	. The drumstick stays in the
4	be lowered due to conveyor	**belt**	. 3x spoon bended
5	No issues noticed. Replace	**belt**	. 4103812 1 drive 1 bearing

1	been replaced but the bearing	**rod**	ends are severely worn, these
2	The case) - The tension special	**rod**	extended by welding
3	nut 6 and the rotating piston	**rod**	for this. 3. Put the static
4	the rod is. When the	**rod**	has been shortened, check whether
5	be switched on if signaling	**rod**	4 is pressed upwards by carrier 7.

1	in line with the adjacent	**track**	sections. Install reinforcement
2	are clean and dry. Cleaning	**track**	weigher: If the gap is
3	two trolleys touching the weigh	**track**	piece. This part of the
4	damaged ones - please replace T-	**track**	2. Units: - cam rollers are seized;
5	conveyor Switchbox. Examine the	**track**	height. See chapter 5: Adjust the

1	down sigma used instead of	**beam**	plan. Crop skin removing module
2	leg skin cutters across the	**beam**	on which they have been
3	does not protrude the back	**beam**	The incorrect support is at
4	is fixed on the frame	**beam**	via the load cell. The
5	soon as possible. Modify the	**beam**	in breast cap cutter. This

1	on the width adjustment. Bearing	**bush**	broken on the R side.
2	stop swivel head missing. Bronze	**bush**	is pushed out of the
3	coupling in the opening of	**bush**	4. The connection between the
4	Second flipper L	**bushes**	4308215 2 unloading station
5	broken 3x width adjustment. Plastic	**bushing**	width adjustment pushed out.

1	during production whether the breast	**caps**	are cut off evenly.
2	Technical issues on unit branch	**cap**	loader Line 3 were discussed.
3	and the performance on smaller	**caps**	. 2. Presser block Norm scales
4	pictures of the bad quality	**caps**	they have been receiving lately.
5	or pieces of the knee	**cap**	remain in the thigh meat

APPENDIX V: FULL PEWL

abdominal	10
access	16
accurately	32
activate	22
adaptor	10
adjacent	19
adjust	570
alarm	20
anatomical	35
appendix	36
area	35
authorize	25
automatic	71
available	30
battery	13
beam	30
bear	158
behalf	12
belt	60
bin	15
blade	375
blower	17
bolt	223
bracket	33
breast	153
bridge	18
broiler	15
buffer	11
bushing	26
bypass	42
cable	51
calibrate	14
cam	35
cap	53
carcass	12
carrier	194
cartilage	27
cell	18
chain	223
chapter	196
chemical	21
chill	24
chute	14
circuit	32
circumstances	21
clamp	18
clockwise	45
clutch	21
code	11
competent	46
component	82
compress	56
concept	24
configuration	13
confirmation	20
consist	30
consult	26
contact	44
converter	13
convey	254
cord	34
corrective	15
corresponding	14
counter	13
coupling	17
crate	10
curve	32
cutter	114
cylinder	74
data	191
debone	14
defect	17
design	15
detect	92
device	17
diagram	43
dimension	20
disc	64
discharge	27
disinfect	24
display	15
document	69
drive	168
drumstick	40
eject	54
emergency	168
ensure	20
entrapment	11
entry	16
equipment	64
error	11
eviscerate	12
exhaust	11

exit	10
familiarize	12
fillet	68
filter	37
flammable	13
flange	12
force	12
friction	48
function	29
fuse	10
gauge	10
gear	59
giblet	10
gizzard	11
grade	10
grindstone	16
groin	22
hazard	36
height	258
hoist	12
holder	37
hood	14
horizontal	38
humidity	10
hydraulic	15
hygiene	18
identify	34
impairment	13
incision	63
index	25
indicate	38
infeed	65
injury	189
input	12
insert	11
inspect	72
install	254
instruct	224
intermediate	11
ionizing	11
issue	42
joint	55
journal	10
kit	26
kneecap	20
knob	35
label	64
layout	33
legislate	41
limiter	57
line	386
loadcell	17
loader	13
logbook	39
lubricate	35
maintenance	292
manual	384
mechanism	21
membrane	18
minimum	11
mobile	12
modify	92
modular	57
module	260
mortal	112
mount	75
nut	113
offal	11
option	36
outfeed	10
output	20
overhaul	32
overhead	123
overload	39
overview	23
pack	42
padlock	45
panel	239
paragraph	85
piston	18
pitch	17
plastic	10
plate	167
plug	59
pneumatic	67
pollution	43
portioning	94
position	167
poultry	160
preface	31
presser	41
preventive	24
previous	31
prior	44
procedure	42
proceed	47
process	235
professional	46
protrude	13

proximity	147
puller	16
pulse	10
purchase	54
pusher	32
race	16
rack	11
range	12
rear	23
recover	28
recycle	18
reducer	26
reference	38
regulate	119
regulator	16
relay	31
release	40
relevant	19
remove	220
require	69
reset	35
resistor	10
revise	53
rod	43
roller	53
rotate	142
safeguard	42
scald	14
schedule	56
scraper	36
section	68
selection	33
sensor	59
separator	74
serial	19
setting	294
shackle	218
shaft	71
skin	63
slack	12
socket	46
software	13
spacer	15
specific	70
spine	31
splitter	15
squeeze	12
static	38
status	11
storage	15
strain	17
suction	10
sufficient	17
suspension	12
switch	508
symbol	27
synchronize	69
tank	25
technical	165
tendon	37
tension	100
terminal	56
text	29
thermistor	11
thermo	13
thigh	87
tilt	12
timing	15
torque	11
track	38
transfer	104
transmit	57
transport	101
trolley	55
troubleshooting	20
trough	15
truck	13
ultrasonic	11
unauthorized	10
undo	20
unintentional	30
vacuum	17
valve	75
vent	12
vertical	16
via	14
vibrate	13
visible	11
voltage	31
wagon	13
washer	38
wear	114
weighbridge	26
weigher	33
weld	18
wishbone	16
workspace	19
wrench	12
zone	26

APPENDIX VI: PEWL CLASSIFICATION

Highly technical	**Semi-technical**	**Sub-technical**	**Non-technical**
conveyor	line	switch	adjust
carrier	blade	setting	manual
poultry	module	maintenance	height
overhead	panel	remove	install
bearing	process	breast	instruct
mount	bolt	proximity	chapter
cylinder	chain	rotate	data
separator	shackle	regulate	injury
pneumatic	drive	transport	emergency
infeed	plate	modify	technical
disc	cutter	thigh	mortal
limiter	wear	component	paragraph
modular	nut	synchronize	inspect
friction	transfer	fillet	automatic
rod	tension	incision	document
bypass	portioning	terminate	require
pack	detect	compress	equip
presser	valve	schedule	specifications
overload	shaft	trolley	purchase
static	section	cable	proceed
track	label	contact	competent
washer	skin	diagram	professional
scraper	belt	pollution	clockwise
cam	gear	issue	padlock
cord	plug	safeguard	prior
weigher	sensor	release	procedure
circuit	transmit	logbook	legislation
curve	terminal	anatomical	horizontal
overhaul	joint	lubricate	indicate

pusher	eject	reset	reference
relay	cap	function	appendix
spine	roller	symbol	hazard
voltage	socket	chill	option
beam	drumstick	disinfect	identify
cartilage	filter	chemical	layout
discharge	holder	dimension	selection
bushing	tendon	output	accurately
kit	area	troubleshooting	preface
reducer	knob	serial	previous
weighbridge	bracket	cell	available
rear	zone	membrane	consist
groin	tank	defect	unintentional
clutch	concept	device	text
kneecap	preventive	display	recover
bridge	mechanism	calibrate	consult
clamp	entry	debone	authorize
piston	race	battery	index
weld	corrective	impairment	overview
blower	via	familiarize	activate
coupling	configuration	input	revision
loadcell	position	tilt	circumstances
pitch	protrude	entrapment	alarm
strain	thermo	ultrasonic	confirmation
vacuum	force	abdominal	ensure
grindstone	range	humidity	undo
puller	suspension	suction	adjacent
regulator	insert		relevant
wishbone	rack		workspace
bin	gauge		hygiene
broiler	grade		recycle
hydraulic	journal		sufficient
spacer	pulse		access

splitter			vertical
trough			design
chute			specific
hood			storage
scald			timing
converter			corresponding
counter			software
flammable			truck
loader			vibrate
wagon			behalf
carcass			mobile
eviscerate			squeeze
flange			code
hoist			error
slack			intermediate
vent			minimum
wrench			status
buffer			crate
exhaust			exit
gizzard			plastic
ionizing			visible
offal			unauthorized
thermistor			
torque			
adaptor			
fuse			
giblet			
outfeed			
resistor			

APPENDIX VII: COVER SHEET

Verklaring geen fraude en plagiaat

Ondergetekende
[voornaam, achternaam en studentnummer],

Allysha Humphrey, s4532473

Bachelorstudent aan de Letterenfaculteit van de Radboud Universiteit Nijmegen,

verklaart dat de beoordeelde scriptie volledig oorspronkelijk is en uitsluitend door hem/haarzelf geschreven is. Bij alle informatie en ideeën ontleend aan andere bronnen, heeft ondergetekende expliciet en in detail verwezen naar de vindplaatsen. De erin gepresenteerde onderzoeksgegevens zijn door ondergetekende zelf verzameld op de in de scriptie beschreven wijze.

Plaats en datum:

Nijmegen, 3 Oktober 2020

Handtekening: